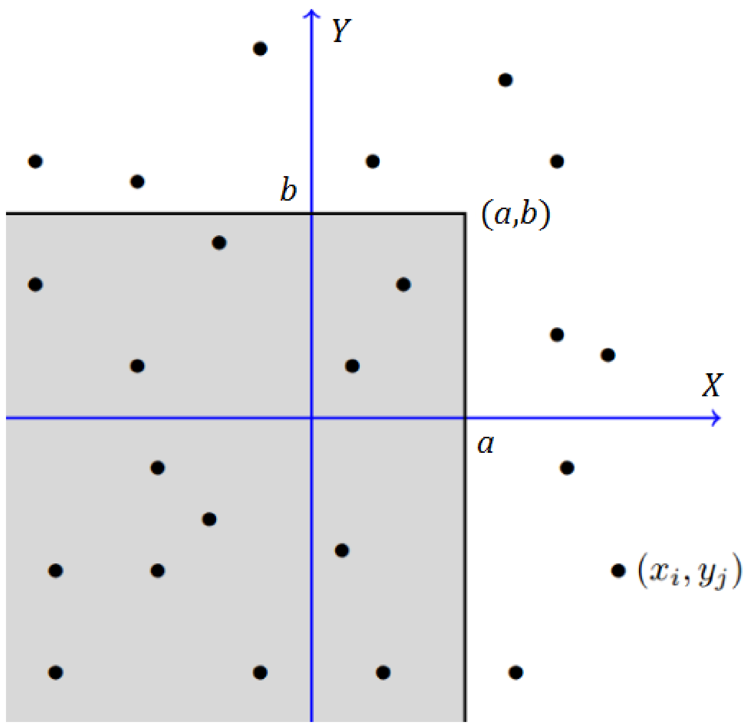
Chapter 5 Jointly distributed random variables

Definition 5.1

Define, for any two random variables X and Y, the **joint probability distribution function of X and Y** by

$$F_{X,Y}(a,b) = P(X \le a, Y \le b) \quad -\infty < a,b < \infty$$
$$= P\big((X \le a) \cap (Y \le b)\big)$$



We have

$$P(a_1 < X \le a_2, b_1 < Y \le b_2) = F_{X,Y}(a_2,b_2) - F_{X,Y}(a_2,b_1) - F_{X,Y}(a_1,b_2) + F_{X,Y}(a_1,b_1)$$

whenever $a_1 < a_2$ and $b_1 < b_2$.

Definition 5.2

In the case when X and Y are both discrete random variables, it is convenient to define the **joint probability mass function of X and Y** by

$$p_{X,Y}(x,y) = P(X = x, Y = y).$$

Definition 5.3

The **marginal probability mass function of a discrete random variable** X can be obtained from $p_{X,Y}(x,y)$ by

$$p_X(x) = P(X = x) = \sum_y p(x,y).$$

Similarly, the **marginal probability mass function of a discrete random variable** Y is

$$p_Y(y) = P(Y = y) = \sum_x p(x,y).$$

The **marginal distribution of** X can be obtained from the joint distribution of X and Y as follows:

$$\begin{aligned}
F_X(a) &= P(X \le a) \\
&= P(X \le a, Y < \infty) \\
&= P\left(\lim_{b\to\infty}(X \le a, Y \le b)\right) \\
&= \lim_{b\to\infty} P(X \le a, Y \le b) \\
&= \lim_{b\to\infty} F_{X,Y}(a,b) \\
&= F_{X,Y}(a,\infty)
\end{aligned}$$

Similarly, the **marginal distribution of** Y is given by

$$F_Y(b) = F_{X,Y}(\infty, b).$$

Also, note that we must have

$$\begin{aligned}
&F_{X,Y}(\infty,\infty) = \lim_{b\to\infty} F_{X,Y}(\infty,b) = \lim_{b\to\infty} F_Y(b) = 1, \\
&F_{X,Y}(-\infty,b) = \lim_{a\to-\infty} F_{X,Y}(a,b) \quad \text{for any } b, \\
&F_{X,Y}(a,-\infty) = \lim_{b\to-\infty} F_{X,Y}(a,b) \quad \text{for any } a.
\end{aligned}$$

Example 5.4

Suppose that 3 balls are randomly selected from an urn containing 3 red, 4 white, and 5 blue balls. If we let X and Y denote, respectively, the number of red and white balls chosen, then the joint probability mass function of X and Y, $p_{X,Y}(i,j) = P(X = i, Y = j)$, is given by

$P_X(i)$	$\frac{84}{220}$	$\frac{108}{220}$	$\frac{27}{220}$	$\frac{1}{220}$	
3	$p(0,3)=\frac{\binom{4}{3}}{\binom{12}{3}}=\frac{4}{220}$	0	0	0	$\frac{4}{220}$
2	$p(0,2)=\frac{\binom{4}{2}\binom{5}{1}}{\binom{12}{3}}=\frac{30}{220}$	$p(1,2)=\frac{\binom{3}{1}\binom{4}{2}}{\binom{12}{3}}=\frac{18}{220}$	0	0	$\frac{48}{220}$
1	$p(0,1)=\frac{\binom{4}{1}\binom{5}{2}}{\binom{12}{3}}=\frac{40}{220}$	$p(1,1)=\frac{\binom{3}{1}\binom{4}{1}\binom{5}{1}}{\binom{12}{3}}=\frac{60}{220}$	$p(2,1)=\frac{\binom{3}{2}\binom{4}{1}}{\binom{12}{3}}=\frac{12}{220}$	0	$\frac{112}{220}$
0	$p(0,0)=\frac{\binom{5}{3}}{\binom{12}{3}}=\frac{10}{220}$	$p(1,0)=\frac{\binom{3}{1}\binom{5}{2}}{\binom{12}{3}}=\frac{30}{220}$	$p(2,0)=\frac{\binom{3}{2}\binom{5}{1}}{\binom{12}{3}}=\frac{15}{220}$	$p(3,0)=\frac{\binom{3}{3}}{\binom{12}{3}}=\frac{1}{220}$	$\frac{56}{220}$
$j \;\; \cdot^{\cdot^{\cdot}}$ $\cdot^{\cdot^{\cdot}} \;\; i$	0	1	2	3	$P_Y(j)$

Definition 5.5

We say that X and Y are **jointly continuous** if there exists a function $f_{X,Y}(x,y)$, defined for all real x and y, having the property that for every set Ω in the two-dimensional plane,

$$P\big((X,Y)\in\Omega\big)=\iint_{(x,y)\in\Omega} f_{X,Y}(x,y)dxdy$$

If $\Omega=\boldsymbol{R}^2$,

$$\iint_{(x,y)\in R^2} f_{X,Y}(x,y)dxdy=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)dydx=F_{X,Y}(\infty,\infty)=1.$$

Example 5.6

A device runs until either of two components fails, at which point the device stops running. The joint density function of the lifetimes of the two components, both measured in hours, is

$$f_{X,Y}(x,y)=C(x+y) \quad \text{for} \quad 0<x,y<3.$$

Calculate the probability that the device fails during its first hour of operation.

Solution

$$1 = \iint_{(x,y)\in R^2} f_{X,Y}(x,y)\,dx\,dy = \int_0^3\int_0^3 C(x+y)\,dy\,dx$$

$$= C\int_0^3 \left[xy + \frac{y^2}{2} \right]_0^3 dx$$

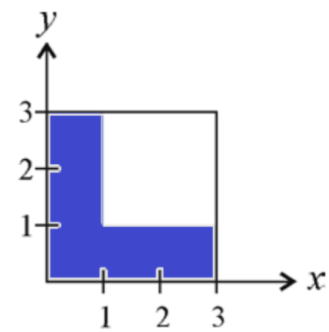
$$= C\int_0^3 3x + \frac{9}{2}\,dx$$

$$= C\left[\frac{3}{2}x^2 + \frac{9}{2}x \right]_0^3$$

$$= 27C$$

$C = \frac{1}{27}$.

That the device fails within the first hour means the joint density function must be integrated over the shaded region shown below.



$$P\big((X<1)\cup(Y<1)\big) = 1 - \int_1^3\int_1^3 \frac{1}{27}(x+y)\,dy\,dx$$

$$= 1 - \frac{1}{27}\int_1^3 \left[xy + \frac{y^2}{2} \right]_1^3 dx$$

$$= 1 - \frac{1}{27}\int_1^3 \left(3x + \frac{9}{2}\right) - \left(x + \frac{1}{2}\right) dx$$

$$= 1 - \frac{1}{27}\int_1^3 (2x+4)\,dx$$

$$= 1 - \frac{1}{27}\left[x^2 + 4x \right]_1^3$$

$$= 1 - \frac{1}{27}\left[21 - 5 \right]$$

$$= \frac{11}{27}$$

Because

$$F_{X,Y}(a,b) = P(X \le a, Y \le b) = \int_{-\infty}^{a}\int_{-\infty}^{b} f_{X,Y}(x,y)dydx,$$

by fundamental theorem of calculus,

$$\begin{aligned}
\frac{\partial^2}{\partial a \partial b} F_{X,Y}(a,b) &= \frac{\partial^2}{\partial a \partial b}\int_{-\infty}^{a}\int_{-\infty}^{b} f_{X,Y}(x,y)dydx \\
&= \frac{\partial}{\partial a}\int_{-\infty}^{a}\left(\frac{\partial}{\partial b}\int_{-\infty}^{b} f_{X,Y}(x,y)dy\right)dx \\
&= \frac{\partial}{\partial a}\int_{-\infty}^{a} f_{X,Y}(x,b)dx \\
&= f_{X,Y}(a,b)
\end{aligned}$$

Definition 5.7

If X and Y are jointly continuous, the **marginal probability density function of** X can be obtained from $f_{X,Y}(x,y)$, by

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy.$$

Similarly, the **marginal probability density function of** Y is

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx.$$

Furthermore, we define

$$P(X \in A) = \int_A \int_{-\infty}^{\infty} f_{X,Y}(x,y)dydx = \int_A f_X(x)dx, \quad P(Y \in B) = \int_B \int_{-\infty}^{\infty} f_{X,Y}(x,y)dxdy = \int_B f_Y(y)dy$$

Definition 5.8

If X and Y are discrete random variables, define the **conditional probability mass function of** X **given that** $Y = y$, by

$$\begin{aligned}
p_{X|Y}(x \mid y) &= P(X = x \mid Y = y) \\
&= \frac{P(X = x, Y = y)}{P(Y = y)} \\
&= \frac{p(x,y)}{p_Y(y)}
\end{aligned}$$

for all values of y such that $p_Y(y) > 0$. Similarly, the **conditional probability distribution function of** X **given that** $Y = b$ is defined, for all b such that $p_Y(b) > 0$, by

$$\begin{aligned} F_{X|Y}(a \mid b) &= P(X \le a \mid Y = b) \\ &= \sum_{x \le a} p_{X|Y}(x \mid b) \end{aligned}$$

In other words, the definitions are exactly the same as in the unconditional case, except that everything is now conditional on the event that $Y = b$.

Definition 5.9

If X and Y have a joint probability density function $f_{X,Y}(x, y)$, then the **conditional probability density function of** X **given that** $Y = y$ is defined, for all values of y such that $f_Y(y) > 0$, by

$$f_{X|Y}(x \mid y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}.$$

To motivate this definition, multiply the left-hand side by dx and the right-hand side by $(dxdy)/dy$ to obtain

$$\begin{aligned} f_{X|Y}(x \mid y)dx &= \frac{f(x, y)dxdy}{f_Y(y)dy} \\ &\approx \frac{P(x \le X \le x + dx, y \le Y \le y + dy)}{P(y \le Y \le y + dy)} \\ &= P(x \le X \le x + dx \mid y \le Y \le y + dy) \end{aligned}$$

In other words, for small values of dx and dy, $f_{X|Y}(x \mid y)dx$ represents the conditional probability that X is between x and $x + dx$ given that Y is between y and $y + dy$.

If X and Y are jointly continuous, then, for any set A, define

$$P(X \in A \mid Y = y) = \int_A f_{X|Y}(x \mid y)dx.$$

Then

$$P(X \in A) = \int_A f_X(x)dx = \int_{-\infty}^{\infty} P(X \in A, Y = y)dy = \int_{-\infty}^{\infty} P(X \in A \mid Y = y) f_Y(y)dy.$$

In particular, by letting $A = (-\infty, a]$ we can define the **conditional cumulative distribution function of** X **given that** $Y = y$ by

$$F_{X|Y}(a|y) \equiv P(X \le a|Y = y) = \int_{-\infty}^{a} f_{X|Y}(x|y)dx.$$

Definition 5.10

The random variables X and Y are said to be **independent** if, for any two sets of real numbers A and B,

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B).$$

In other words, X and Y are independent if, for all A and B, the events $\{X \in A\}$ and $\{Y \in B\}$ are independent. It can be shown that X and Y are independent if and only if

$$F_{X,Y}(a,b) = F_X(a)F_Y(b) \text{ for all } a, b.$$

Two discrete random variables X and Y are independent if and only if

$$p_{X,Y}(x, y) = p_X(x)p_Y(y) \text{ for all } x, y.$$

The conditional mass function and the distribution function are the same as the respective unconditional ones

$$p_{X|Y}(x\,|\,y) = \frac{p(x, y)}{p_Y(y)} = \frac{p_X(x)p_Y(y)}{p_Y(y)} = p_X(x).$$

Two continuous random variables X and Y are said to be independent if and only if

$$f_{X,Y}(x, y) = f_X(x)f_Y(y) \text{ for all } x, y.$$

If X and Y are independent continuous random variables, the conditional density of X given that $Y = y$ is just the unconditional density of X. This is so because, in the independent case,

$$f_{X|Y}(x|y) = \frac{f(x, y)}{f_Y(y)} = \frac{f_X(x)f_Y(y)}{f_Y(y)} = f_X(x).$$

Thus, loosely speaking, X and Y are independent if knowing the value of one does not change the distribution of the other. Random variables that are not independent are said to be dependent.

Example 5.11

(a) If the joint density function of X and Y is

$$f_{X,Y}(x,y) = 6e^{-2x}e^{-3y} \quad 0 < x, y < \infty$$

and is equal to 0 outside this region, are the random variables independent?

(b) What if the joint density function is

$$f_{X,Y}(x,y) = 24xy \quad 0 < x, y < 1, \quad 0 < x + y < 1$$

and is equal to 0 otherwise?

Solution

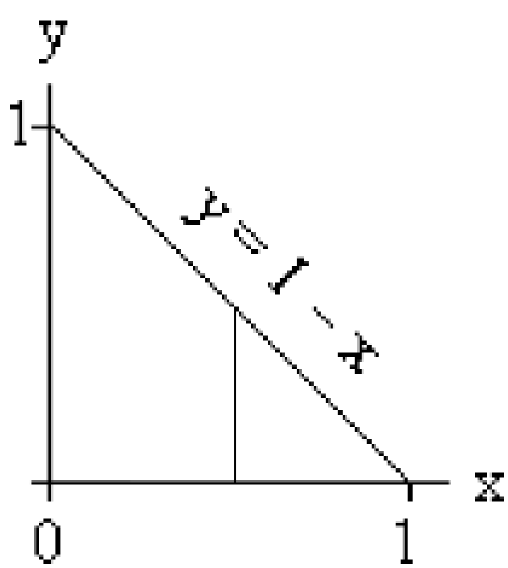
(a) Clearly, $f_{X,Y}(x,y) = f_X(x) f_Y(y)$ where

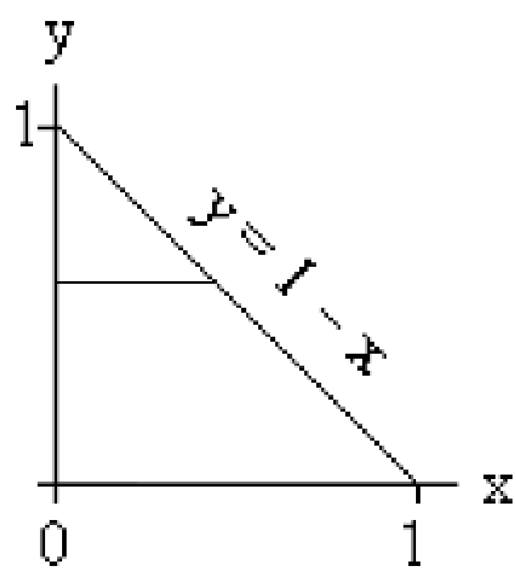
$$f_X(x) = \begin{cases} 2e^{-2x} & x > 0 \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad f_Y(y) = \begin{cases} 3e^{-3y} & y > 0 \\ 0 & \text{otherwise} \end{cases}$$

are probability density functions of exponential distribution with parameter 2 and 3 respectively.

(b) $$\begin{aligned} f_X(x) &= \int_0^{1-x} f_{X,Y}(x,y)dy \\ &= \int_0^{1-x} 24xydy \\ &= 12x(1-x)^2 \end{aligned}$$

Similarly, $f_Y(y) = 12y(1-y)^2$. Therefore $f_{X,Y}(x,y) \neq f_X(x) f_Y(y)$, and X and Y are dependent.





Theorem 5.12

Let X and Y be 2 independent random variables. Let g and h be real-valued functions of X and Y respectively. Then $g(X)$ and $h(Y)$ are independent random variables.

Proof

$$\begin{aligned} P(g(X)\in A, h(Y)\in B) &= P(X \in g^{-1}(A), Y\in h^{-1}(B)) \\ &= P(X \in g^{-1}(A))P(Y\in h^{-1}(B)) \\ &= P(g(X)\in A)P(h(Y)\in B) \end{aligned}$$

Let X and Y be nonnegative independent discrete random variables. Then the probability mass function of $X+Y$ is

$$P(X+Y=n)=\sum_{k=0}^{n} P(X=k, Y=n-k)=\sum_{k=0}^{n} P(X=k)P(Y=n-k).$$

In terms of probability generating functions, we have $p_{X+Y}(z)=p_X(z)p_Y(z).$

Theorem 5.13

If X and Y are independent Poisson random variables with respective parameters λ_1 and λ_2, then $X+Y$ has a Poisson distribution with parameter $\lambda_1+\lambda_2$.

Proof

Because the event $\{X+Y=n\}$ can be written as the union of the disjoint events $\{X=k, Y=n-k\},\ 0\le k\le n,$ we have

$$\begin{aligned} P(X+Y=n) &= \sum_{k=0}^{n} P(X=k)P(Y=n-k) \\ &= \sum_{k=0}^{n} e^{-\lambda_1}\frac{\lambda_1^k}{k!}e^{-\lambda_2}\frac{\lambda_2^{n-k}}{(n-k)!} \\ &= e^{-(\lambda_1+\lambda_2)}\sum_{k=0}^{n}\frac{\lambda_1^k\lambda_2^{n-k}}{k!(n-k)!} \\ &= \frac{e^{-(\lambda_1+\lambda_2)}}{n!}\sum_{k=0}^{n}\frac{n!}{k!(n-k)!}\lambda_1^k\lambda_2^{n-k} \\ &= \frac{e^{-(\lambda_1+\lambda_2)}}{n!}(\lambda_1+\lambda_2)^n \end{aligned}$$

or

$$p_X(z)=\sum_{k=0}^{\infty} e^{-\lambda_1}\frac{\lambda_1^k}{k!}z^k = e^{\lambda_1(z-1)}$$

$$p_Y(z)=\sum_{k=0}^{\infty} e^{-\lambda_2}\frac{\lambda_2^k}{k!}z^k = e^{\lambda_2(z-1)}$$

$$\begin{aligned} p_{X+Y}(z) &= p_X(z)p_Y(z) \\ &= e^{(\lambda_1+\lambda_2)(z-1)} \\ &= \sum_{k=0}^{\infty} e^{-(\lambda_1+\lambda_2)}\frac{(\lambda_1+\lambda_2)^k}{k!}z^k \end{aligned}$$

$$P(X+Y=n)=e^{-(\lambda_1+\lambda_2)}\frac{(\lambda_1+\lambda_2)^n}{n!}$$

Thus, $X + Y$ has a Poisson distribution with parameter $\lambda_1 + \lambda_2$.

Example 5.14

If X and Y are independent Poisson random variables with respective parameters λ_1 and λ_2, calculate the conditional distribution of X given that $X + Y = n$.

Solution

Since $X + Y$ has a Poisson distribution with parameter $\lambda_1 + \lambda_2$ (Theorem 4.13),

$$
\begin{aligned}
P(X = k \mid X + Y = n) &= \frac{P(X = k, X + Y = n)}{P(X + Y = n)} \\
&= \frac{P(X = k, Y = n - k)}{P(X + Y = n)} \\
&= \frac{P(X = k)P(Y = n - k)}{P(X + Y = n)} \\
&= \frac{e^{-\lambda_1}\lambda_1^k}{k!}\frac{e^{-\lambda_2}\lambda_2^{n-k}}{(n-k)!} \bigg/ \frac{e^{-(\lambda_1+\lambda_2)}(\lambda_1 + \lambda_2)^n}{n!} \\
&= \frac{n!}{k!(n-k)!}\frac{\lambda_1^k \lambda_2^{n-k}}{(\lambda_1 + \lambda_2)^n} \\
&= \binom{n}{k}\left(\frac{\lambda_1}{\lambda_1 + \lambda_2}\right)^k \left(\frac{\lambda_2}{\lambda_1 + \lambda_2}\right)^{n-k}
\end{aligned}
$$

In other words, the conditional distribution of X given that $X + Y = n$ is the binomial distribution with parameters n and $\lambda_1/(\lambda_1+\lambda_2)$.

Suppose that X and Y are independent, continuous random variables having probability density functions f_X and f_Y. The cumulative distribution function and probability density function of $X + Y$ are obtained as follows:

$$
\begin{aligned}
F_{X+Y}(a) &= P(X + Y \le a) \\
&= P(X \le a - Y) \\
&= \int_{-\infty}^{\infty} P(X \le a - Y \mid Y = y) f_Y(y)dy \\
&= \int_{-\infty}^{\infty} P(X \le a - y) f_Y(y)dy \qquad [X \text{ and } Y \text{ are independent}] \\
&= \int_{-\infty}^{\infty} F_X(a - y) f_Y(y)dy
\end{aligned}
$$

$$f_{X+Y}(a) = \frac{d}{da}\int_{-\infty}^{\infty} F_X(a-y)\, f_Y(y)dy$$
$$= \int_{-\infty}^{\infty} \frac{d}{da} F_X(a-y)\, f_Y(y)dy$$
$$= \int_{-\infty}^{\infty} f_X(a-y)\, f_Y(y)dy$$

Example 5.15

If X and Y are independent random variables, both uniformly distributed on (0, 1), calculate the probability density function of $X + Y$.

Solution

Since $f_X(a-y) = \begin{cases} 1 & 0 < a-y < 1 \\ 0 & \text{otherwise} \end{cases}$ and $f_Y(y) = \begin{cases} 1 & 0 < y < 1 \\ 0 & \text{otherwise,} \end{cases}$ $f_X(a-y)\, f_Y(y) = 1$ if and only if $0 < a-y < 1$ and $0 < y < 1$. That means $a-1 < y < a$ and $0 < y < 1$.

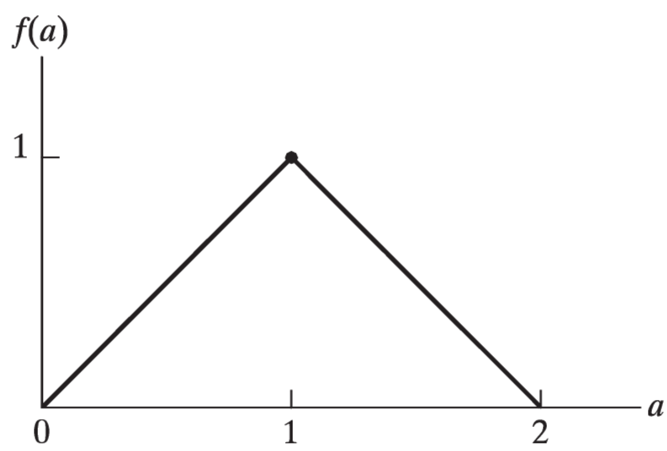
For $0 \le a < 1,$ this yields $0 < y < a$. We obtain

$$f_{X+Y}(a) = \int_0^a 1dy = a.$$

For $1 \le a < 2,$ this yields $a-1 < y < 1.$ We get

$$f_{X+Y}(a) = \int_{a-1}^1 1dy = 2-a.$$

Hence $f_{X+Y}(a) = \begin{cases} a & 0 \le a < 1 \\ 2-a & 1 \le a < 2 \\ 0 & \text{otherwise.} \end{cases}$



Example 5.16

Consider ABC bank with two clerks. Four people, Alan, Bonny, Cora, and Dr. Wong, need bank service. Alan and Bonny go directly to the first and second clerk, and Cora waits until either Alan or Bonny leaves before she begins the service. Dr. Wong is the last one who enters the bank and wait until one of the clerk is available. Let *X* and *Y* be the serving time of the first and second clerk respectively. Assume *X* and *Y* are 2 independent exponential random variables with parameters λ_1 and λ_2 respectively.

(a) Find the probability that Alan will leave before Bonny.

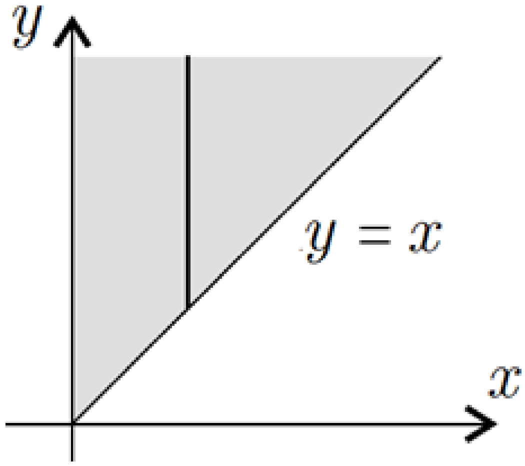
(b) What is the distribution of Dr. Wong's waiting time?

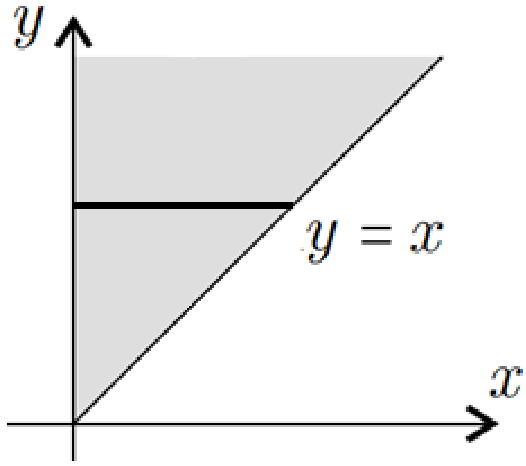
Solution

(a)
$$\begin{aligned}
P(X<Y) &= \iint_{x<y} f_{X,Y}(x,y)\,dydx \\
&= \int_0^\infty \int_x^\infty f_X(x) f_Y(y)\,dydx \\
&= \int_0^\infty \int_x^\infty \lambda_1\lambda_2 e^{-\lambda_1 x} e^{-\lambda_2 y}\,dydx \\
&= \int_0^\infty \lambda_1 e^{-\lambda_1 x}\left[-e^{-\lambda_2 y}\right]_x^\infty dx \\
&= \int_0^\infty \lambda_1 e^{-\lambda_1 x} e^{-\lambda_2 x}\,dx \\
&= \left[-\frac{\lambda_1}{\lambda_1+\lambda_2} e^{-(\lambda_1+\lambda_2)x}\right]_0^\infty \\
&= \frac{\lambda_1}{\lambda_1+\lambda_2}.
\end{aligned}$$

or

$$\begin{aligned}
P(X<Y) &= \iint_{x<y} f_{X,Y}(x,y)\,dxdy \\
&= \int_0^\infty \int_0^y f_X(x) f_Y(y)\,dxdy \\
&= \int_0^\infty \int_0^y \lambda_1\lambda_2 e^{-\lambda_1 x} e^{-\lambda_2 y}\,dxdy \\
&= \int_0^\infty \lambda_2 e^{-\lambda_2 y}\left[-e^{-\lambda_1 x}\right]_0^y dy \\
&= \int_0^\infty \lambda_2 e^{-\lambda_2 y}\left(1-e^{-\lambda_1 y}\right) dy \\
&= \int_0^\infty \lambda_2 e^{-\lambda_2 y} - \lambda_2 e^{-(\lambda_1+\lambda_2)y}\,dy \\
&= \left[-e^{-\lambda_2 y}\right]_0^\infty + \left[\frac{\lambda_2}{\lambda_1+\lambda_2} e^{-(\lambda_1+\lambda_2)y}\right]_0^\infty \\
&= 1-\frac{\lambda_2}{\lambda_1+\lambda_2} = \frac{\lambda_1}{\lambda_1+\lambda_2}.
\end{aligned}$$





(b) Let Z_1 be waiting time of Cora. Then $Z_1 = \min\{X, Y\}$. Suppose $z > 0$.

$$\begin{aligned} P(Z_1 > z) &= P(\min\{X, Y\} > z) \\ &= P(X > z, Y > z) \\ &= P(X > z)P(Y > z) \\ &= e^{-\lambda_1 z} e^{-\lambda_2 z} \\ &= e^{-(\lambda_1 + \lambda_2) z} \end{aligned}$$

Then $F_{Z_1}(z) = \begin{cases} 1 - e^{-(\lambda_1+\lambda_2)z} & \text{if } z > 0 \\ 0 & \text{otherwise.} \end{cases}$ Z_1 is an exponential random variable with parameter $\lambda_1 + \lambda_2$. Since exponential distribution has memoryless property, the distribution of waiting time Z_2 after serving Cora and before serving Dr. Wong is the same as the distribution of waiting time of Cora Z_1. Then the random variable of Dr. Wong's waiting time is $Z_1 + Z_2$, which is a Gamma distribution with parameter $(n = 2,\ \lambda = \lambda_1 + \lambda_2)$.

Theorem 5.17

If X and Y have a joint probability mass function $p(x, y)$, then

$$E[g(X,Y)] = \sum_x \sum_y g(x, y) p(x, y).$$

If X and Y have a joint probability density function $f_{X,Y}(x, y)$, then

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

Theorem 5.18

Suppose $E[X]$ and $E[Y]$ are both finite and let $g(X, Y) = X + Y$. Then $E[X+Y] = E[X] + E[Y]$. If $E[X_i]$ is finite for all $i = 1, \cdots, n$, then $E[X_1 + \cdots + X_n] = E[X_1] + \cdots + E[X_n]$.

Proof

In discrete case,

$$\begin{aligned}
E[X+Y] &= \sum_x \sum_y (x+y)\,p(x,y) \\
&= \sum_x \sum_y xp(x,y) + \sum_x \sum_y yp(x,y) \\
&= \sum_x x \sum_y p(x,y) + \sum_y y \sum_x p(x,y) \\
&= \sum_x xp_X(x) + \sum_y yp_Y(y) \\
&= E[X] + E[Y].
\end{aligned}$$

In continuous case,

$$\begin{aligned}
E[X+Y] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x+y)\, f_{X,Y}(x,y)dydx \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xf_{X,Y}(x,y)dydx + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} yf_{X,Y}(x,y)dydx \\
&= \int_{-\infty}^{\infty} x\left(\int_{-\infty}^{\infty} f_{X,Y}(x,y)dy\right)dx + \int_{-\infty}^{\infty} y\left(\int_{-\infty}^{\infty} f_{X,Y}(x,y)dx\right)dy \\
&= \int_{-\infty}^{\infty} xf_X(x)dx + \int_{-\infty}^{\infty} yf_Y(y)dy \\
&= E[X] + E[Y].
\end{aligned}$$

Theorem 5.19

If X and Y are independent, then, for any functions h and g, then

$$E[g(X)h(Y)] = E[g(X)]E[h(Y)].$$

Proof

Suppose that X and Y are jointly continuous with joint density $f_{X,Y}(x,y)$. Then

$$\begin{aligned}
E[g(X)h(Y)] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x)h(y)\, f_{X,Y}(x,y)dxdy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x)h(y)\, f_X(x)\, f_Y(y)dxdy \\
&= \left(\int_{-\infty}^{\infty} g(x)\, f_X(x)dx\right)\left(\int_{-\infty}^{\infty} h(y)\, f_X(x)dy\right) \\
&= E[g(X)]E[h(Y)].
\end{aligned}$$

The proof in the discrete case is similar.

Let $E[X|Y]$ be function of the random variable Y whose value at $Y = y$ is $E[X|Y = y]$. Note that $E[X|Y]$ is itself a random variable.

Theorem 5.20 (Law of Total Expectation)

If Y is a discrete random variable, then

$$E[X] = \sum_y E[X|Y = y]P(Y = y)$$

whereas if Y is continuous, then

$$E[X] = \int_{-\infty}^{\infty} E[X|Y = y] f_Y(y)dy.$$

In notation, we write $E[X] = E\left[E[X|Y]\right]$.

Proof

When X and Y are discrete:

$$\begin{aligned}
\sum_y E[X|Y = y]P(Y = y) &= \sum_y \sum_x xP(X = x|Y = y)P(Y = y) \\
&= \sum_y \sum_x x \frac{P(X = x, Y = y)}{P(Y = y)} P(Y = y) \\
&= \sum_y \sum_x xP(X = x, Y = y) \\
&= \sum_x \sum_y xP(X = x, Y = y) \\
&= \sum_x x \sum_y P(X = x, Y = y) \\
&= \sum_x xP_X(X = x) \\
&= E[X]
\end{aligned}$$

When X and Y are continuous:

$$\begin{aligned}\int_{-\infty}^{\infty} E[X \mid Y = y] f_Y(y)dy &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f_{X|Y}(y)dx f_Y(y)dy \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x \frac{f_{X,Y}(x,y)}{f_Y(y)} f_Y(y)dxdy \\ &= \int_{-\infty}^{\infty} x \int_{-\infty}^{\infty} f_{X,Y}(x,y)dydx \\ &= \int_{-\infty}^{\infty} x f_X(x)dx \\ &= E[X].\end{aligned}$$

The other cases leave as an exercise.

Example 5.21

A miner is trapped in a mine containing 3 doors. The first door leads to a tunnel that takes him to safety after 2 hours of travel. The second door leads to a tunnel that returns him to the mine after 3 hours of travel. The third door leads to a tunnel that returns him to the mine after 5 hours. Assuming that the miner is at all times equally likely to choose any one of the doors, find $E[X]$ and $Var[X]$ for random variable X, the time when the miner reaches safety.

Solution

Let Y denote the door initially chosen. Then

$$E[X] = E[X|Y=1]P(Y=1) + E[X|Y=2]P(Y=2) + E[X|Y=3]P(Y=3)$$

Now given that $Y = 1$, it follows that $X = 2$, and so

$$E[X|Y=1] = 2.$$

Also, given that $Y = 2$, it follows that $X = 3 + X'$, where X' is the number of additional hours to safety after returning to the mine. But once the miner returns to the mine the problem is exactly as before, and thus X' has the same distribution as X. Therefore,

$$E[X|Y=2] = E[3 + X'] = 3 + E[X].$$

Similarly,

$$E[X|Y=3] = 5 + E[X].$$

Then

$$E[X] = E[X|Y=1]P(Y=1) + E[X|Y=2]P(Y=2) + E[X|Y=3]P(Y=3)$$
$$= \frac{1}{3}\left(E[X|Y=1] + E[X|Y=2] + E[X|Y=3]\right)$$
$$= \frac{1}{3}\left(2 + 3 + E[X] + 5 + E[X]\right)$$
$$E[X] = 10.$$
$$E[X^2] = E[X^2|Y=1]P(Y=1) + E[X^2|Y=2]P(Y=2) + E[X^2|Y=3]P(Y=3)$$
$$= \frac{1}{3}\left(E[X^2|Y=1] + E[X^2|Y=2] + E[X^2|Y=3]\right)$$
$$= \frac{1}{3}\left(2^2 + E[(3+X')^2] + E[(5+X')^2]\right)$$
$$= \frac{1}{3}\left(2^2 + 9 + 6E[X] + E[X^2] + 25 + 10E[X] + E[X^2]\right)$$
$$E[X^2] = 198.$$
$$Var[X] = E[X^2] - E[X]^2 = 198 - 100 = 98.$$

Define the conditional variance of X to be

$$Var(X|Y) = E[(X - E[X|Y])^2 \mid Y)].$$

$Var(X|Y)$ is a random variable which depends on Y. Similar to $Var[X] = E[X^2] - E[X]^2$, we have

$$Var(X|Y) = E\left[X^2|Y\right] - E[X|Y]^2.$$

Theorem 5.22

$$Var(X) = E[Var(X \mid Y)] + Var(E[X \mid Y])$$

Proof

$$E[Var(X|Y)] = E\left[E\left[X^2|Y\right]\right] - E\left[E[X|Y]^2\right] = E[X^2] - E\left[E[X|Y]^2\right]$$
$$Var(E[X \mid Y]) = E\left[E[X|Y]^2\right] - \left(E\left[E[X|Y]\right]\right)^2 = E\left[E[X|Y]^2\right] - E[X]^2$$
$$Var[X] = E[X^2] - E[X]^2 = E[Var(X \mid Y)] + Var(E[X \mid Y])$$

Example 5.23

The intensity of a hurricane is a random variable that is uniformly distributed on the interval [0, 3]. The damage from a hurricane with a given intensity *y* is exponentially distributed with a mean equal to *y*. Calculate the variance of the damage from a random hurricane.

Solution

Let *X* be the hurricane damage and *Y* the intensity. Then $X \big| Y \sim \exp(1/Y)$ and $Y \sim U[0,3]$.

$$\begin{aligned} Var(X) &= E[Var(X \mid Y)] + Var(E[X \mid Y]) \\ &= E[Y^2] + Var(Y) \\ &= \frac{(3-0)^2}{12} + 1.5^2 + \frac{(3-0)^2}{12} \\ &= 3.75 \end{aligned}$$

Suppose there are *N* payments $X_1, X_2, \cdots, X_N$. The number of payment *N* is a discrete random variable. Each payment X_i is independent and identically distributed (i.i.d.). More formally, the independence assumption are:

1. Condition on *N* = *n*, the random variables $X_1, X_2, \cdots, X_n$ are i.i.d. to a random variable *X*.
2. Condition on *N* = *n*, the common distribution of random variables $X_1, X_2, \cdots, X_n$ does not depend on *n*.
3. The distribution of *N* does not depend on the values of $X_1, X_2, \cdots$

The total sum of payments is defined by

$$S = X_1 + X_2 + \cdots + X_N, \quad N = 0, 1, 2, \cdots,$$

where $S = 0$ when $N = 0$.

Theorem 5.24

Based on the above assumption, we have

$$E[S] = E[N]E[X] \text{ and } Var[S] = E[N]Var[X] + Var[N]E[X]^2.$$

Proof

$$\begin{aligned}E[S] &= E\left[E[S|N]\right]\\&= \sum_n E[X_1 + X_2 + \cdots + X_N | N = n]P_N(n)\\&= \sum_n E[X_1 + X_2 + \cdots + X_n]P_N(n)\\&= \sum_n nE[X]P_N(n)\\&= E[X]\sum_n nP_N(n)\\&= E[N]E[X]\end{aligned}$$

$$\begin{aligned}Var(S) &= E[Var(S \mid N)] + Var(E[S \mid N])\\&= E[Var(X_1 + X_2 + \cdots + X_N | N = n)] + Var(E[X_1 + X_2 + \cdots + X_N | N = n])\\&= E[NVar(X)] + Var(NE[X])\\&= E[N]Var[X] + Var[N]E[X]^2\end{aligned}$$

Example 5.25

The number of hurricanes that will hit a certain house in the next ten years is Poisson distributed with mean 4. Each hurricane results in a loss that is exponentially distributed with mean 1000. Losses are mutually independent and independent of the number of hurricanes. Calculate the mean and variance of the total loss due to hurricanes hitting this house in the next ten years.

Solution

Let N denote the number of hurricanes, which is Poisson distributed with mean and variance 4. Let X_i denote the loss due to the i-th hurricane, which is exponentially distributed with mean 1000 and therefore variance $(1000)^2 = 1000000$. Let S denote the total loss due to the N hurricanes.

$$\begin{aligned}E[S] &= E[N]E[X] = 4 \times 1000 = 4000\\Var(S) &= E[N]Var[X] + Var[N]E[X]^2\\&= 4 \times 1000000 + 4(1000)^2\\&= 8000000\end{aligned}$$

Definition 5.26

The **covariance between** X **and** Y, denoted by $Cov(X, Y)$, is defined by

$$Cov(X, Y) = E\left[(X - E[X])(Y - E[Y])\right]$$

We see that

$$\begin{aligned} Cov(X,Y) &= E\left[(X-E[X])(Y-E[Y])\right] \\ &= E\left[XY - E[X]Y - E[Y]X + E[X]E[Y]\right] \\ &= E[XY] - E[X]E[Y] - E[Y]E[X] + E[X]E[Y] \\ &= E[XY] - E[X]E[Y] \end{aligned}$$

Note that if X and Y are independent, then $Cov(X, Y) = 0$. However, the converse is not true.

Example 5.27

Let X be a random variable that takes on the values –1, 0, and 1 with equal probabilities. Let $Y = X^2$. Then $P(Y=0) = 1/3, P(Y=1) = 2/3.$ Clearly, $E[X] = E[X^3] = \frac{1}{3}(1+0+(-1)) = 0$ and

$$Cov(X,Y) = Cov(X,X^2) = E[X^3] - E[X]E[X^2] = 0.$$

However, they are dependent. Note that

$$\begin{aligned} &P(X=0, Y=0) = P(X=0, X^2=0) = P(X=0) = 1/3 \\ &P(X=0)P(Y=0) = (1/3)^2 = 1/9. \end{aligned}$$

Example 5.28

Let X be a uniform random variable over an interval (–1,1). Let $Y = X^2$. Note that $f_X(x)$ is an even function but $xf_X(x)$ and $x^3 f_X(x)$ are odd functions. So $E[X^3] = E[X] = 0,$ and

$$Cov(X,Y) = Cov(X,X^2) = E[X^3] - E[X]E[X^2] = 0.$$

Clearly, Y depends on X. In Example 4.23

$$f_Y(y) = F_Y'(y) = \begin{cases} \dfrac{1}{2\sqrt{y}} & 0 < y < 1 \\ 0 & \text{otherwise.} \end{cases}$$

$f_{X,Y}(\tfrac{1}{2},\tfrac{1}{2}) = f_{X,X^2}(\tfrac{1}{2},\tfrac{1}{2}) = 0$ since $(\tfrac{1}{2})^2 \neq \tfrac{1}{2}$. However, $f_X(\tfrac{1}{2}) f_Y(\tfrac{1}{2}) = \dfrac{1}{2} \times \dfrac{1}{2\sqrt{\tfrac{1}{2}}} \neq 0.$

Theorem 5.29

(i) $Cov(X,Y) = Cov(Y,X)$

(ii) $Cov(X,X) = Var(X)$

(iii) $Cov(aX,Y) = aCov(X,Y)$

(iv) $Cov(X,1) = 0$

(v) $Cov\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{m} Y_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} Cov(X_i, Y_j).$

Suppose $m = n$ and $X_i = Y_i, i = 1,\cdots,n$ in Theorem 5.29 (v). We obtain

$$
\begin{aligned}
Var\left(\sum_{i=1}^{n} X_i\right) &= Cov\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{n} X_j\right) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} Cov(X_i, X_j) \\
&= \sum_{i=1}^{n} Var(X_i) + \sum_{i \neq j} Cov(X_i, X_j) \\
&= \sum_{i=1}^{n} Var(X_i) + 2\sum_{i<j} Cov(X_i, X_j)
\end{aligned}
$$

If $X_1,\cdots,X_n$ are pairwise independent, in that X_i and X_j are independent for $i \neq j$, then

$$Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i).$$

In particular, if $X_1,\cdots,X_n$ are independent and identically distributed (i.i.d.) to a common distribution *X*, then

$$Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i) = nVar(X).$$

Question 5.30

If $X_1,\cdots,X_n$ are i.i.d. to *X*, then $X_1 = \cdots = X_n = X$. We have

$$Var\left(\sum_{i=1}^{n} X_i\right) = Var\left(nX\right) = n^2 Var(X).$$

Is it correct?

Example 5.31

Let $X_1,\cdots,X_n$ be independent and identically distributed random variables having expected value μ and variance σ^2. Define sample mean

$$\bar{X}=\frac{X_1+\cdots+X_n}{n}$$

and sample variance

$$S^2=\sum_{i=1}^{n}\frac{(X_i-\bar{X})^2}{n-1}.$$

Find $E[\bar{X}], Var[\bar{X}]$ and $E[S^2]$.

Solution

$$E[\bar{X}]=E\left[\frac{X_1+\cdots+X_n}{n}\right]=\frac{E[X_1]+\cdots+E[X_n]}{n}=\frac{n\mu}{n}=\mu$$

$$Var(\bar{X})=Var\left(\frac{X_1+\cdots+X_n}{n}\right)=\frac{Var(X_1)+\cdots+Var(X_n)}{n^2}=\frac{n\sigma^2}{n^2}=\frac{\sigma^2}{n}$$

$$\begin{aligned}(n-1)S^2&=\sum_{i=1}^{n}(X_i-\bar{X})^2\\&=\sum_{i=1}^{n}(X_i-\mu+\mu-\bar{X})^2\\&=\sum_{i=1}^{n}\left[(X_i-\mu)^2+2(X_i-\mu)(\mu-\bar{X})+(\mu-\bar{X})^2\right]\\&=\sum_{i=1}^{n}(X_i-\mu)^2+2(\mu-\bar{X})\sum_{i=1}^{n}(X_i-\mu)+\sum_{i=1}^{n}(\mu-\bar{X})^2\\&=\sum_{i=1}^{n}(X_i-\mu)^2+2(\mu-\bar{X})n(\bar{X}-\mu)+n(\mu-\bar{X})^2\\&=\left[\sum_{i=1}^{n}(X_i-\mu)^2\right]-n(\bar{X}-\mu)^2\end{aligned}$$

$$(n-1)E[S^2]=E\left[\sum_{i=1}^{n}(X_i-\mu)^2\right]-nE\left[(\bar{X}-\mu)^2\right]=n\sigma^2-nVar(\bar{X})=n\sigma^2-\sigma^2$$

$$E[S^2]=\sigma^2$$

Example 5.32

Let $X_1, \cdots, X_n$ be independent and identical Bernoulli random variables such that

$$X_i = \begin{cases} 1 & \text{if the } i\text{-th trial is a success with probability } p \\ 0 & \text{failure with probability } 1-p \end{cases}$$

Let $X = X_1 + \cdots + X_n$. Such a random variable is a binomial random variable with parameter *n*, *p* which represents the number of successes in *n* independent trials.

$$E[X] = E[X_1] + \cdots + E[X_n] = np \quad \text{and} \quad Var(X) = Var(X_1) + \cdots + Var(X_n) = np(1-p).$$

Definition 5.33

The **correlation of two random variables** *X* **and** *Y*, denoted by $\rho_{X,Y}$, is defined, as long as $\sigma_X \sigma_Y > 0$, by

$$\rho_{X,Y} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}.$$

Theorem 5.34

$$-1 \le \rho_{X,Y} \le 1.$$

If $\rho_{X,Y} = 1$, then $\sigma_Y X - \sigma_X Y = \sigma_Y \mu_X - \sigma_X \mu_Y$. If $\rho_{X,Y} = -1$, then $\sigma_Y X - \sigma_X Y = \sigma_Y \mu_X - \sigma_X \mu_Y$.

Proof

$0 \le Var(\sigma_Y X + t\sigma_X Y) = t^2 \sigma_X^2 \sigma_Y^2 + 2t\sigma_X \sigma_Y Cov(X,Y) + \sigma_X^2 \sigma_Y^2 = \sigma_X^2 \sigma_Y^2 \left(t^2 + 2t\rho_{X,Y} + 1\right)$ for all *t*. We have $0 \ge \Delta = (2\rho_{X,Y})^2 - 4$. Hence $-1 \le \rho_{X,Y} \le 1$.

If $\rho_{X,Y} = 1$, choose $t = -1$. Then $Var(\sigma_Y X - \sigma_X Y) = 0$ and $\sigma_Y X - \sigma_X Y = \sigma_Y \mu_X - \sigma_X \mu_Y$ constant.

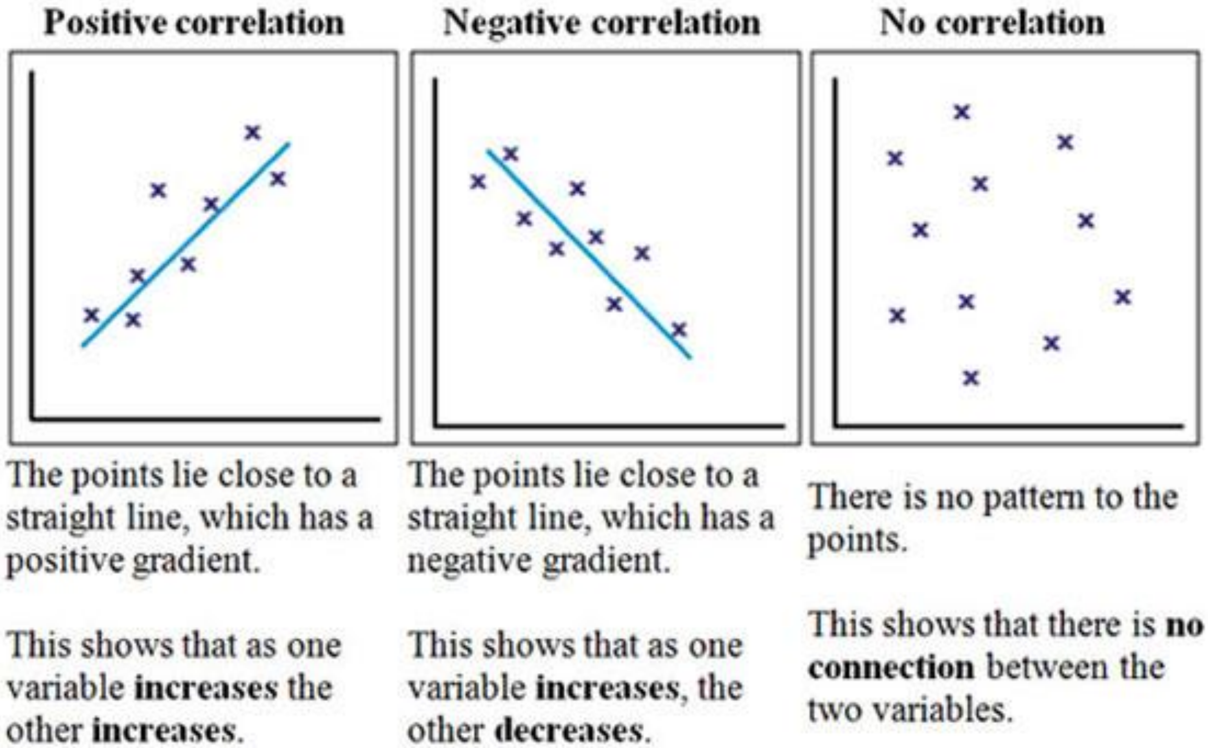
If $\rho_{X,Y} = -1$, choose $t = 1$. Then $Var(\sigma_Y X + \sigma_X Y) = 0$ and $\sigma_Y X + \sigma_X Y = \sigma_Y \mu_X + \sigma_X \mu_Y$ constant.

Consider two random variables *X* and *Y*:

(i) If $\rho_{X,Y} > 0$, we say that *X* and *Y* are positively correlated.

(ii) If $\rho_{X,Y} < 0$, we say that *X* and *Y* are negatively correlated.

(iii) If $\rho_{X,Y} = 0$, we say that X and Y are uncorrelated.



Theorem 5.35

Let X_1 and X_2 be two jointly continuous random variables. Let $(Y_1, Y_2) = g(X_1, X_2) = (g_1(X_1, X_2), g_2(X_1, X_2))$, where $g : \boldsymbol{R}^2 \to \boldsymbol{R}^2$ is a continuous one-to-one (invertible) function with continuous partial derivatives. Let $h = g^{-1}$, i.e., $(X_1, X_2) = h(Y_1, Y_2) = (h_1(Y_1, Y_2), h_2(Y_1, Y_2))$. Then Y_1 and Y_2 are jointly continuous and their joint PDF, $f_{Y_1,Y_2}(y_1, y_2)$, for (y_1, y_2) in the range, is given by

$$f_{Y_1,Y_2}(y_1, y_2) = f_{X_1,X_2}\left(h_1(y_1, y_2), h_2(y_1, y_2)\right)|J|,$$

where J is the Jacobian of h defined by

$$J = \det\begin{pmatrix} \dfrac{\partial h_1}{\partial y_1} & \dfrac{\partial h_1}{\partial y_2} \\ \dfrac{\partial h_2}{\partial y_1} & \dfrac{\partial h_2}{\partial y_2} \end{pmatrix} = \frac{\partial h_1}{\partial y_1}\frac{\partial h_2}{\partial y_2} - \frac{\partial h_1}{\partial y_2}\frac{\partial h_2}{\partial y_1}.$$

Example 5.36

Let X_1 and X_2 be jointly continuous random variables with probability density function f_{X_1,X_2}. Let $Y_1 = X_1 + X_2, Y_2 = X_1 - X_2$. Find the joint density function of Y_1 and Y_2 in terms of f_{X_1,X_2}.

Solution

Let $g_1(X_1, X_2) = X_1 + X_2$ and $g_2(X_1, X_2) = X_1 - X_2$. Then

$$X_1 = h_1(Y_1,Y_2) = \frac{Y_1+Y_2}{2} \text{ and } X_2 = h_2(Y_1,Y_2) = \frac{Y_1-Y_2}{2}.$$

Hence

$$J = \det\begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{pmatrix} = -\frac{1}{2}.$$

It follows from Theorem 4.50 that

$$f_{Y_1,Y_2}(y_1, y_2) = \frac{1}{2} f_{X_1,X_2}\left(\frac{y_1+y_2}{2}, \frac{y_1-y_2}{2}\right).$$

For instance, if X_1 and X_2 are independent uniform (0, 1) random variables, then

$$f_{Y_1,Y_2}(y_1, y_2) = \begin{cases} \frac{1}{2} & 0 < y_1 + y_2, y_1 - y_2 < 2 \\ 0 & \text{otherwise} \end{cases}$$

or if X_1 and X_2 are independent exponential random variables with respective parameters λ_1 and λ_2, then

$$f_{Y_1,Y_2}(y_1, y_2) = \begin{cases} \frac{\lambda_1\lambda_2}{2} e^{-\lambda_1\left(\frac{y_1+y_2}{2}\right)-\lambda_2\left(\frac{y_1-y_2}{2}\right)} & y_1 + y_2, y_1 - y_2 \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Finally, if X_1 and X_2 are independent standard normal random variables, then

$$f_{Y_1,Y_2}(y_1, y_2) = \frac{1}{4\pi} e^{-\frac{1}{2}\left[\left(\frac{y_1+y_2}{2}\right)^2+\left(\frac{y_1-y_2}{2}\right)^2\right]} = \frac{1}{4\pi} e^{-\frac{\left(y_1^2+y_2^2\right)}{4}} = \frac{1}{\sqrt{4\pi}} e^{-\frac{y_1^2}{4}} \frac{1}{\sqrt{4\pi}} e^{-\frac{y_2^2}{4}}.$$

Y_1 and Y_2 are also independent normal random variables, each with parameters $\mu = 0$ and $\sigma^2 = 2$.

Theorem 5.37

Let X and Y be two random variables with joint PDF $f_{X,Y}(x,y)$. Let $Z = X + Y$. Then

$$f_Z(z) = \int_{-\infty}^{\infty} f_{X,Y}(w, z-w)dw.$$

In particular, if X and Y are independent, then $f_Z(z) = \int_{-\infty}^{\infty} f_X(w) f_Y(z-w)dw$

Proof

Define $W = X$ such that $(Z,W) = g(X,Y) = (g_1(X,Y), g_2(X,Y)) = (X+Y, X)$. Then

$$(X,Y) = h(Z,W) = (h_1(Z,W), h_2(Z,W)) = (W, Z-W).$$

Hence

$$J = \begin{pmatrix} 0 & 1 \\ 1 & -1 \end{pmatrix} \quad \text{and} \quad f_{Z,W}(z,w) = f_{X,Y}\left(h_1(z,w), h_2(z,w)\right)\left|J\right| = f_{X,Y}\left(w, z-w\right).$$

But since we are interested in the marginal PDF of Z, $f_Z(z)$, we have

$$f_Z(z) = \int_{-\infty}^{\infty} f_{Z,W}(z,w)dw = \int_{-\infty}^{\infty} f_{X,Y}(w, z-w)dw.$$

Note that, if X and Y are independent, then $f_{X,Y}(x,y) = f_X(x) f_Y(y)$ and we conclude that

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(w) f_Y(z-w)dw.$$

An n-dimensional vector $\boldsymbol{X}$ is called **random vector** if

$$\boldsymbol{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix},$$

where $X_1, \cdots, X_n$ are random variables.

The **expected value vector** or the **mean vector of the random vector $\boldsymbol{X}$** is defined as

$$E[\boldsymbol{X}] = \begin{pmatrix} E[X_1] \\ \vdots \\ E[X_n] \end{pmatrix}.$$

Similarly, an $m \times n$ **random matrix *M*** is a matrix whose elements are random variables.

$$\boldsymbol{M} = \begin{pmatrix} X_{1,1} & \cdots & X_{1,n} \\ \vdots & \ddots & \vdots \\ X_{m,1} & \cdots & X_{m,n} \end{pmatrix}.$$

The **mean matrix of *M*** is given by

$$E[\boldsymbol{M}] = \begin{pmatrix} E[X_{1,1}] & \cdots & E[X_{1,n}] \\ \vdots & \ddots & \vdots \\ E[X_{m,1}] & \cdots & E[X_{m,n}] \end{pmatrix}.$$

Linearity of expectation is also valid for random vectors and matrices. In particular, let $\boldsymbol{X}$ be an n-dimensional random vector and the m-dimensional random vector $\boldsymbol{Y}$ be defined as

$$\boldsymbol{Y} = A\boldsymbol{X} + \boldsymbol{b}$$

where A is a fixed $m \times n$ matrix and $\boldsymbol{b}$ is a fixed m-dimensional vector. Then we have

$$E[\boldsymbol{Y}] = AE[\boldsymbol{X}] + \boldsymbol{b}.$$

The covariance matrix of $\boldsymbol{X}$ is

$$\begin{aligned} \boldsymbol{C}_{\boldsymbol{X}} &= \begin{pmatrix} Var(X_1) & \cdots & Cov(X_1, X_n) \\ \vdots & \ddots & \vdots \\ Cov(X_n, X_1) & \cdots & Var(X_n) \end{pmatrix} \\ &= E\begin{pmatrix} (X_1 - E[X_1])(X_1 - E[X_1]) & \cdots & (X_1 - E[X_1])(X_n - E[X_n]) \\ \vdots & \ddots & \vdots \\ (X_n - E[X_n])(X_1 - E[X_1]) & \cdots & (X_n - E[X_n])(X_n - E[X_n]) \end{pmatrix} \\ &= E\left(\begin{pmatrix} X_1 - E[X_1] \\ \vdots \\ X_n - E[X_n] \end{pmatrix} \begin{pmatrix} X_1 - E[X_1] & \cdots & X_n - E[X_n] \end{pmatrix} \right) \\ &= E\left((\boldsymbol{X} - E[\boldsymbol{X}])(\boldsymbol{X} - E[\boldsymbol{X}])^T \right) \end{aligned}$$

Theorem 5.38

Let $\boldsymbol{X}$ be an $\boldsymbol{R}^n$ vector-valued random variable with covariance matrix $\boldsymbol{C}_{\boldsymbol{X}}$. Let $\boldsymbol{Y}$ be an $\boldsymbol{R}^m$ vector-valued random variable such that $\boldsymbol{Y} = A\boldsymbol{X} + \boldsymbol{b}$ where A is an $m \times n$ matrix and $\boldsymbol{b}$ is a $m \times 1$ column vector. Then covariance matrix of $\boldsymbol{Y}$ is $\boldsymbol{C}_{\boldsymbol{Y}} = A\boldsymbol{C}_{\boldsymbol{X}}A^T$.

Proof

Note that by linearity of expectation, we have

$$\boldsymbol{Y}-E[\boldsymbol{Y}]=\left(A\boldsymbol{X}+\boldsymbol{b}\right)-\left(AE[\boldsymbol{X}]+\boldsymbol{b}\right)=A\left(\boldsymbol{X}-E[\boldsymbol{X}]\right).$$

Then

$$\begin{aligned}\boldsymbol{C}_{\boldsymbol{Y}} &= E\left(\left(\boldsymbol{Y}-E[\boldsymbol{Y}]\right)\left(\boldsymbol{Y}-E[\boldsymbol{Y}]\right)^T\right)\\ &= E\left(A\left(\boldsymbol{X}-E[\boldsymbol{X}]\right)\left(\boldsymbol{X}-E[\boldsymbol{X}]\right)^T A^T\right)\\ &= AE\left(\left(\boldsymbol{X}-E[\boldsymbol{X}]\right)\left(\boldsymbol{X}-E[\boldsymbol{X}]\right)^T\right)A^T\\ &= A\boldsymbol{C}_{\boldsymbol{X}}A^T\end{aligned}$$

Definition 5.39

A random vector $\boldsymbol{Z}=(Z_1,\cdots,Z_n)^T$ is called the **standard normal random vector** if $Z_1,\cdots,Z_n$ are independent standard normal random variables. Its joint probability density function is

$$\begin{aligned}f_{\boldsymbol{Z}}(\boldsymbol{z}) &= f_{Z_1,\cdots,Z_n}(z_1,\cdots,z_n)\\ &= f_{Z_1}(z_1)\cdots f_{Z_n}(z_n)\\ &= \frac{1}{(2\pi)^{n/2}}e^{-\frac{1}{2}\left(z_1^2+\cdots+z_n^2\right)}\\ &= \frac{1}{(2\pi)^{n/2}}e^{-\frac{1}{2}\left(\boldsymbol{z}^T\boldsymbol{z}\right)}\end{aligned}$$

Theorem 5.40

Let $\boldsymbol{Z}=\begin{pmatrix}Z_1\\Z_2\end{pmatrix}$ be standard normal vector. A function $g:\boldsymbol{R}^2\to\boldsymbol{R}^2$ is defined by $\boldsymbol{X}=\begin{pmatrix}X_1\\X_2\end{pmatrix}=g\begin{pmatrix}Z_1\\Z_2\end{pmatrix}=A\begin{pmatrix}Z_1\\Z_2\end{pmatrix}+\boldsymbol{b}$, where A is a 2×2 invertible matrix and $\boldsymbol{b}$ is a 2×1 column vector. Then

$$f_{X_1,X_2}(x_1,x_2)=\frac{1}{2\pi\sqrt{\det(\boldsymbol{C}_{\boldsymbol{X}})}}e^{-\frac{1}{2}\left(\boldsymbol{x}-E[\boldsymbol{X}]\right)^T\boldsymbol{C}_{\boldsymbol{X}}^{-1}\left(\boldsymbol{x}-E[\boldsymbol{X}]\right)}$$

Proof

Clearly, $E[\boldsymbol{Z}] = \boldsymbol{0}$ and $C_{\boldsymbol{Z}} = I_2$. Then

$$E[\boldsymbol{X}] = AE[\boldsymbol{Z}] + \boldsymbol{b} = \boldsymbol{b} \quad \text{and} \quad C_{\boldsymbol{X}} = AC_{\boldsymbol{Z}}A^T = AA^T.$$

Let $h = g^{-1}$. Then $\begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} = h\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = A^{-1}\left(\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} - \boldsymbol{b}\right) = A^{-1}\left(\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} - \begin{pmatrix} E[X_1] \\ E[X_2] \end{pmatrix}\right) = A^{-1}\left(\boldsymbol{X} - E[\boldsymbol{X}]\right)$.

Hence

$$\begin{aligned}
\boldsymbol{z}^T\boldsymbol{z} &= \left(A^{-1}\left(\boldsymbol{X} - E[\boldsymbol{X}]\right)\right)^T A^{-1}\left(\boldsymbol{X} - E[\boldsymbol{X}]\right) \\
&= \left(\boldsymbol{X} - E[\boldsymbol{X}]\right)^T \left(A^{-1}\right)^T A^{-1}\left(\boldsymbol{X} - E[\boldsymbol{X}]\right) \\
&= \left(\boldsymbol{X} - E[\boldsymbol{X}]\right)^T \left(AA^T\right)^{-1}\left(\boldsymbol{X} - E[\boldsymbol{X}]\right) \\
&= \left(\boldsymbol{X} - E[\boldsymbol{X}]\right)^T C_{\boldsymbol{X}}^{-1}\left(\boldsymbol{X} - E[\boldsymbol{X}]\right)
\end{aligned}$$

The Jacobian of h is

$$J = \det\begin{pmatrix} \dfrac{\partial h_1}{\partial x_1} & \dfrac{\partial h_1}{\partial x_2} \\ \dfrac{\partial h_2}{\partial x_1} & \dfrac{\partial h_2}{\partial x_2} \end{pmatrix} = \det\left(A^{-1}\right) = \frac{1}{\pm\sqrt{\det C_X}}.$$

By Theorem 5.35, the joint probability density function is

$$\begin{aligned}
f_{X_1,X_2}(x_1,x_2) &= f_{Z_1,Z_2}\left(h_1(x_1,x_2), h_2(x_1,x_2)\right)|J| \\
&= \frac{1}{(2\pi)^{2/2}\sqrt{\det(\boldsymbol{C_X})}} e^{-\frac{1}{2}\left(\boldsymbol{z}^T\boldsymbol{z}\right)} \\
&= \frac{1}{2\pi\sqrt{\det(\boldsymbol{C_X})}} e^{-\frac{1}{2}\left(\boldsymbol{x} - E[\boldsymbol{X}]\right)^T \boldsymbol{C}_{\boldsymbol{X}}^{-1}\left(\boldsymbol{x} - E[\boldsymbol{X}]\right)}
\end{aligned}$$

Definition 5.41

A random vector $\boldsymbol{X} = (X_1, \cdots, X_n)^T$ with mean $E[\boldsymbol{X}]$ and positive definite covariance matrix $\boldsymbol{C_X}$, is said to be **normal** or **Gaussian** if its probability density function is given by

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{1}{(2\pi)^{n/2}\sqrt{\det(\boldsymbol{C_X})}} e^{-\frac{1}{2}\left(\boldsymbol{x} - E[\boldsymbol{X}]\right)^T \boldsymbol{C}_{\boldsymbol{X}}^{-1}\left(\boldsymbol{x} - E[\boldsymbol{X}]\right)}$$

Example 5.42

Let $X \sim N(\mu_X, \sigma_X^2), Y \sim N(\mu_Y, \sigma_Y^2)$ and $\rho(X,Y) = \rho.$ Using definition5.41 to derive

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}e^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right)+\left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]}$$

Solution

n = 2, $\boldsymbol{X} = \begin{pmatrix} X \\ Y \end{pmatrix}, E[\boldsymbol{X}] = \begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}$ and $\boldsymbol{C_X} = \begin{pmatrix} Var(X) & Cov(X,Y) \\ Cov(X,Y) & Var(Y) \end{pmatrix} = \begin{pmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{pmatrix}.$

Then

$$\det(\boldsymbol{C_X}) = \sigma_X^2\sigma_Y^2(1-\rho^2) \text{ and } \boldsymbol{C_X}^{-1} = \frac{1}{\sigma_X^2\sigma_Y^2(1-\rho^2)}\begin{pmatrix} \sigma_Y^2 & -\rho\sigma_X\sigma_Y \\ -\rho\sigma_X\sigma_Y & \sigma_X^2 \end{pmatrix}.$$

We have

$$\begin{aligned}
\left(\boldsymbol{x} - E[\boldsymbol{X}]\right)^T \boldsymbol{C_X}^{-1}\left(\boldsymbol{x} - E[\boldsymbol{X}]\right) &= \frac{1}{\sigma_X^2\sigma_Y^2\left(1-\rho^2\right)}(x-\mu_X, y-\mu_Y)\begin{pmatrix} \sigma_Y^2 & -\rho\sigma_X\sigma_Y \\ -\rho\sigma_X\sigma_Y & \sigma_X^2 \end{pmatrix}\begin{pmatrix} x-\mu_X \\ y-\mu_Y \end{pmatrix} \\
&= \frac{1}{\sigma_X^2\sigma_Y^2\left(1-\rho^2\right)}\left[\sigma_Y^2(x-\mu_X)^2 - 2\rho\sigma_X\sigma_Y(x-\mu_X)(y-\mu_Y) + \sigma_X^2(y-\mu_Y)^2\right] \\
&= \frac{1}{\left(1-\rho^2\right)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right)+\left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right].
\end{aligned}$$

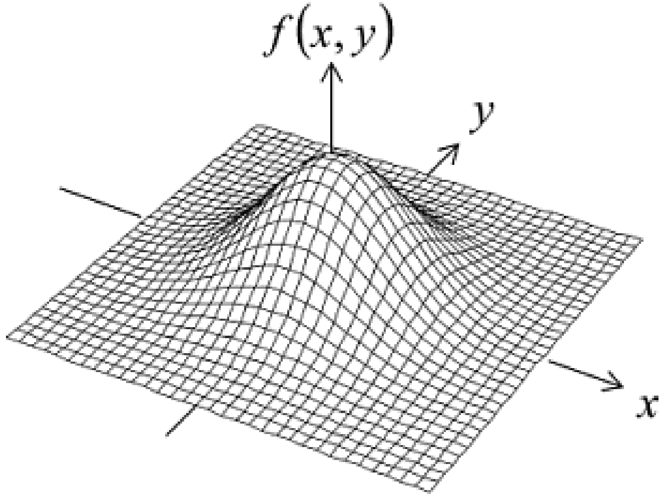
Hence

$$\begin{aligned}
f_{X,Y}(x,y) &= \frac{1}{(2\pi)^{n/2}\sqrt{\det(\boldsymbol{C_X})}}e^{-\frac{1}{2}\left(\boldsymbol{x}-E[\boldsymbol{X}]\right)^T\boldsymbol{C_X}^{-1}\left(\boldsymbol{x}-E[\boldsymbol{X}]\right)} \\
&= \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}e^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right)+\left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]}
\end{aligned}$$

Definition 5.43

Two random variables X, Y have a **bivariate normal distribution** if, for constants $\mu_X, \mu_Y; \sigma_X, \sigma_Y > 0; -1 < \rho < 1,$ their joint density function is given, for all $-\infty < x, y < \infty,$ by

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}e^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right)+\left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]}$$



Theorem 5.44

Two random variables X, Y have a bivariate normal distribution as in Definition 5.43. Then X is normal with mean μ_X and variance σ_X. Similarly, Y is normal with mean μ_Y and variance σ_Y.

Proof

With $C = \dfrac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}$, the marginal density of X can be obtained from

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy \\ &= C\int_{-\infty}^{\infty} e^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right) + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]}dy \end{aligned}$$

Making the change of variables $w = \dfrac{y-\mu_Y}{\sigma_Y}$ gives

$$\begin{aligned} f_X(x) &= C\sigma_Y e^{-\frac{1}{2(1-\rho^2)}\left(\frac{x-\mu_X}{\sigma_X}\right)^2}\int_{-\infty}^{\infty} e^{-\frac{1}{2(1-\rho^2)}\left[-2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)w + w^2\right]}dw \\ &= C\sigma_Y e^{-\frac{1}{2(1-\rho^2)}\left(\frac{x-\mu_X}{\sigma_X}\right)^2} e^{\frac{1}{2(1-\rho^2)}\rho^2\left(\frac{x-\mu_X}{\sigma_X}\right)^2}\int_{-\infty}^{\infty} e^{-\frac{1}{2(1-\rho^2)}\left[\rho^2\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)w + w^2\right]}dw \\ &= C\sigma_Y e^{-\frac{1}{2(1-\rho^2)}\left(\frac{x-\mu_X}{\sigma_X}\right)^2(1-\rho^2)}\int_{-\infty}^{\infty} e^{-\frac{1}{2(1-\rho^2)}\left[w-\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\right]^2}dw \\ &= C\sigma_Y e^{-\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2}\int_{-\infty}^{\infty} e^{-\frac{1}{2(1-\rho^2)}\left[w-\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\right]^2}dw \end{aligned}$$

Because

$$\frac{1}{\sqrt{2\pi\left(1-\rho^2\right)}}\int_{-\infty}^{\infty} e^{-\frac{1}{2\left(1-\rho^2\right)}\left[w-\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\right]^2} dw = 1,$$

we see that

$$f_X(x) = C\sigma_Y\sqrt{2\pi\left(1-\rho^2\right)}e^{-\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2} = \frac{1}{\sqrt{2\pi}\sigma_X}e^{-\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2}.$$

That is, X is normal with mean μ_X and variance σ_X. Similarly, Y is normal with mean μ_Y and variance σ_Y.

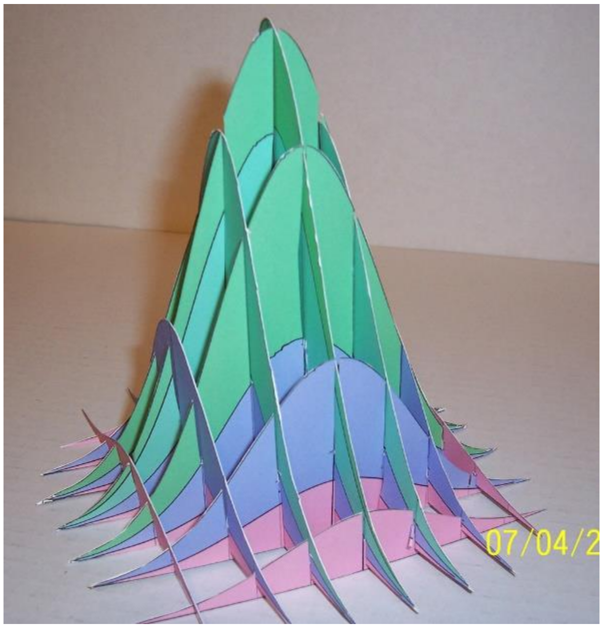
Theorem 5.45

Two random variables X, Y have a bivariate normal distribution as in Definition 5.43, then $f_{X|Y}(x|y)$ is normally distributed with mean $E[X|Y=y] = \mu_X + \rho\sigma_X\frac{y-\mu_Y}{\sigma_Y}$ and variance $Var[X|Y=y] = (1-\rho^2)\sigma_X^2$. Similarly, $f_{Y|X}(y|x)$ is normally distributed with mean $E[Y|X=x] = \mu_Y + \rho\sigma_Y\frac{x-\mu_X}{\sigma_X}$ and variance $Var[Y|X=x] = (1-\rho^2)\sigma_Y^2$.

Proof

$$\begin{aligned} f_{X|Y}(x|y) &= \frac{f_{X,Y}(x,y)}{f_Y(y)} \\ &= C_1 f_{X,Y}(x,y) \\ &= C_2 e^{-\frac{1}{2\left(1-\rho^2\right)}\left[\frac{x^2-2x\mu_X}{\sigma_X^2}-2\rho\left(\frac{x}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right)\right]} \\ &= C_3 e^{-\frac{1}{2\sigma_X^2\left(1-\rho^2\right)}\left[x^2-2x\left(\mu_X+\rho\sigma_X\frac{y-\mu_Y}{\sigma_Y}\right)\right]} \\ &= C_4 e^{-\frac{1}{2\sigma_X^2\left(1-\rho^2\right)}\left[x-\left(\mu_X+\rho\sigma_X\frac{y-\mu_Y}{\sigma_Y}\right)\right]^2} \end{aligned}$$

Recognizing the preceding equation as a normal density, we can conclude that given $Y = y$, the random variable X is normally distributed with mean $\mu_X + \rho\sigma_X\frac{y-\mu_Y}{\sigma_Y}$ and variance $(1-\rho^2)\sigma_X^2$. Similarly the conditional distribution of Y given $X = x$ is the normal distribution with mean $\mu_Y + \rho\sigma_Y\frac{x-\mu_X}{\sigma_X}$ and variance $(1-\rho^2)\sigma_Y^2$.



Theorem 5.46

Two random variables *X*, *Y* have a bivariate normal distribution as in definition 5.43. Then

$$\rho(X, Y) = \rho.$$

Theorem 5.47

The bivariate normal random variables *X* and *Y* are independent if and only if *Cov*(*X*, *Y*) = 0.

Proof

$\Rightarrow$ part is clear. Suppose *Cov*(*X*, *Y*) = 0. Then $\rho = 0$.

$$f_{X,Y}(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y} e^{-\frac{1}{2}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]} = \frac{1}{\sqrt{2\pi}\sigma_X} e^{-\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2} \frac{1}{\sqrt{2\pi}\sigma_Y} e^{-\frac{1}{2}\left(\frac{x-\mu_Y}{\sigma_Y}\right)^2} = f_X(x) f_Y(y).$$

X and *Y* are independent.

Theorem 5.48

If $\boldsymbol{X} = (X_1, \cdots, X_n)^T$ is a normal random vector, $\boldsymbol{X} \sim N(E[\boldsymbol{X}], \boldsymbol{C})$, A is an $m \times n$ fixed matrix, and $\boldsymbol{b}$ is an m-dimensional fixed vector, then the random vector $\boldsymbol{Y} = A\boldsymbol{X} + \boldsymbol{b}$ is a normal random vector with mean $E[\boldsymbol{Y}] = AE[\boldsymbol{X}] + \boldsymbol{b}$ and covariance matrix $\boldsymbol{C_Y} = A\boldsymbol{C_X}A^T$, that is

$$\boldsymbol{Y} \sim N(AE[\boldsymbol{X}] + \boldsymbol{b}, A\boldsymbol{C_X}A^T).$$

Since $\boldsymbol{C_X}$ is a symmetric matrix, it must be diagonalizable. Furthermore, there exists an orthogonal matrix P (that is $PP^T = P^TP = I_n$) such that

$$\boldsymbol{C_X} = PDP^T$$

for some diagonal matrix D.

In particular, if $m = n$ and $A = P^T$, Then $\boldsymbol{C_Y}$ is a diagonal matrix. Furthermore, if $\boldsymbol{X}$ is a normal random vector, then $Y_1, \cdots, Y_n$ are independent normal random variables.

Remark 5.49

If X_1 and X_2 (not necessary independent) are normal, then $X_1 + X_2$ is normal by choosing A = (1, 1) and $\boldsymbol{b} = \boldsymbol{0}$ in Theorem 5.48.

Example 5.50

Let $\boldsymbol{X} = (X_1, X_2, X_3)^T$ be a normal random vector such that $E[\boldsymbol{X}] = \boldsymbol{0}$ and covariance matrix of $\boldsymbol{X}$ is

$$\boldsymbol{C_X} = \begin{pmatrix} \frac{13}{49} & -\frac{12}{49} & \frac{18}{49} \\ -\frac{12}{49} & \frac{45}{49} & \frac{6}{49} \\ \frac{18}{49} & \frac{6}{49} & \frac{40}{49} \end{pmatrix}.$$

Express X_3 in terms of X_1 and X_2. Hence find two non-constant independent random variables in terms of X_1 and X_2.

Solution

Diagonalize $\boldsymbol{C_X}$, we have

$$\boldsymbol{C_X} = \begin{pmatrix} \frac{13}{49} & -\frac{12}{49} & \frac{18}{49} \\ -\frac{12}{49} & \frac{45}{49} & \frac{6}{49} \\ \frac{18}{49} & \frac{6}{49} & \frac{40}{49} \end{pmatrix} = \begin{pmatrix} \frac{3}{7} & \frac{2}{7} & \frac{6}{7} \\ -\frac{6}{7} & \frac{3}{7} & \frac{2}{7} \\ \frac{2}{7} & \frac{6}{7} & -\frac{3}{7} \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \frac{3}{7} & \frac{2}{7} & \frac{6}{7} \\ -\frac{6}{7} & \frac{3}{7} & \frac{2}{7} \\ \frac{2}{7} & \frac{6}{7} & -\frac{3}{7} \end{pmatrix}^T .$$

Let $\begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix} = \begin{pmatrix} \frac{3}{7} & \frac{2}{7} & \frac{6}{7} \\ -\frac{6}{7} & \frac{3}{7} & \frac{2}{7} \\ \frac{2}{7} & \frac{6}{7} & -\frac{3}{7} \end{pmatrix}^T \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} = \begin{pmatrix} \frac{3}{7}X_1 - \frac{6}{7}X_2 + \frac{2}{7}X_3 \\ \frac{2}{7}X_1 + \frac{3}{7}X_2 + \frac{6}{7}X_3 \\ \frac{6}{7}X_1 + \frac{2}{7}X_2 - \frac{3}{7}X_3 \end{pmatrix}$. Then $\boldsymbol{C_Y} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$. We obtain

$$0 = Var(Y_3) = Var\left(\tfrac{6}{7}X_1 + \tfrac{2}{7}X_2 - \tfrac{3}{7}X_3\right).$$

Hence $\frac{6}{7}X_1 + \frac{2}{7}X_2 - \frac{3}{7}X_3 = c.$

Since $E[\boldsymbol{X}] = \boldsymbol{0}$, that is $E[X_1] = E[X_2] = E[X_3] = 0,$ then $c = \frac{6}{7}E[X_1] + \frac{2}{7}E[X_2] - \frac{3}{7}E[X_3] = 0.$

Therefore

$$\tfrac{6}{7}X_1 + \tfrac{2}{7}X_2 - \tfrac{3}{7}X_3 = 0 \quad \text{and} \quad X_3 = 2X_1 + \tfrac{2}{3}X_2$$

By Theorem 5.47, $Cov(Y_1, Y_2) = 0$ implies that Y_1, Y_2 are independent.

$$\begin{aligned} Y_1 &= \tfrac{3}{7}X_1 - \tfrac{6}{7}X_2 + \tfrac{2}{7}X_3 = \tfrac{3}{7}X_1 - \tfrac{6}{7}X_2 + \tfrac{2}{7}(2X_1 + \tfrac{2}{3}X_2) = X_1 - \tfrac{2}{3}X_2 \\ Y_2 &= \tfrac{2}{7}X_1 + \tfrac{3}{7}X_2 + \tfrac{6}{7}X_3 = \tfrac{2}{7}X_1 + \tfrac{3}{7}X_2 + \tfrac{6}{7}(2X_1 + \tfrac{2}{3}X_2) = 2X_1 + X_2 \end{aligned}$$

Hence $X_1 - \frac{2}{3}X_2$ and $2X_1 + X_2$ are independent.

Definition 5.51

The **moment generating function** $M_X(t)$ **of random variable** X is defined for all real values of t by

$$M_X(t) = E\left[e^{tX}\right] = \begin{cases} \sum_x e^{tX} p(x) & \text{if } X \text{ is discrete with probability mass function } p(x) \\ \int_{-\infty}^{\infty} e^{tx} f(x)dx & \text{if } X \text{ is continuous with probability density function } f_X(x) \end{cases}$$

We can write

$$e^{tX} = 1 + tX + \frac{(tX)^2}{2!} + \frac{(tX)^3}{3!} + \cdots$$

Thus, we have

$$M_X(t) = E\left[e^{tX}\right] = 1 + E[X]t + E[X^2]\frac{t^2}{2!} + E[X^3]\frac{t^3}{3!} + \cdots$$

Hence

$$M_X(0) = 1 \quad \text{and} \quad E[X^k] = \frac{d^k}{dt^k} M_X(t)\Big|_{t=0}.$$

Furthermore,

$$\frac{d}{dt}\ln M_X(t)\Big|_{t=0} = \frac{M_X'(t)}{M_X(t)}\Big|_{t=0} = E[X] \text{ and}$$

$$\frac{d^2}{dt^2}\ln M_X(t)\Big|_{t=0} = \frac{d}{dt}\left(\frac{M_X'(t)}{M_X(t)}\right)\Big|_{t=0} = \frac{M_X(t)M_X''(t) - M_X'(t)^2}{M_X^2(t)}\Big|_{t=0} = Var[X]$$

If $M_X(t)$ exists and is finite in some region about t = 0, then the distribution of X is uniquely determined.

Example 5.52

Suppose that the moment generating function of a random variable X is given by $M_X(t) = e^{3(e^t - 1)}$. What is $P(X = 0)$?

Solution

Clearly, $M_X(t) = e^{3(e^t - 1)}$ is the moment generating function of a Poisson random variable with mean $\lambda = 3$. Hence, by the one-to-one correspondence between moment generating functions and distribution functions, X must be a Poisson random variable with mean 3. Thus, $P(X = 0) = e^{-3}$.

Suppose that X and Y are independent and have moment generating functions $M_X(t)$ and $M_Y(t)$, respectively. Then $M_{X+Y}(t)$ the moment generating function of $X + Y$, is given by

$$M_{X+Y}(t) = E\left[e^{t(X+Y)}\right] = E\left[e^{tX}e^{tY}\right] = E\left[e^{tX}\right]E\left[e^{tY}\right] = M_X(t)M_Y(t).$$

Theorem 5.53

If X and Y are independent normal random variables with respective parameters (μ_1, σ_1^2) and (μ_2, σ_2^2), then $X + Y$ is normal with mean $\mu_1 + \mu_2$ and variance $\sigma_1^2 + \sigma_2^2$.

Solution

$$M_{X+Y}(t) = M_X(t)M_Y(t) = e^{\mu_1 t + \frac{\sigma_1^2 t^2}{2}} e^{\mu_2 t + \frac{\sigma_2^2 t^2}{2}} = e^{(\mu_1+\mu_2)t + \frac{\left(\sigma_1^2+\sigma_2^2\right)t^2}{2}}$$

which is the moment generating function of a normal random variable with mean $\mu_1 + \mu_2$ and variance $\sigma_1^2 + \sigma_2^2$. The desired result then follows because the moment generating function uniquely determines the distribution.

In general, let $X_i, i = 1, \cdots, n,$ be independent normal random variables with respective parameters (μ_i, σ_i^2), then $X_1 + \cdots + X_n$ is normal with mean $\mu_1 + \cdots + \mu_n$ and variance $\sigma_1^2 + \cdots + \sigma_n^2$.

Theorem 5.54

Let $X_1, \cdots, X_n$ be i.i.d. random variables each having a normal distribution with mean μ and variance σ^2. Define the random variables

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{and} \quad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

Then $\bar{X}$ and S^2 are independent.

Proof

$$Cov(\bar{X}, X_i - \bar{X}) = Cov(\bar{X}, X_i) - Cov(\bar{X}, \bar{X}) = Cov(\tfrac{1}{n} X_i, X_i) - Var(\bar{X}) = \tfrac{1}{n}\sigma^2 - \tfrac{1}{n}\sigma^2 = 0.$$

Since $\bar{X}$ and $X_i - \bar{X}$ are normal for all i, by Theorem 5.54, $\bar{X}$ and $X_i - \bar{X}$ are independent for all i. Because S^2 is a function of $X_i - \bar{X}, i = 1, \cdots, n,$ S^2 is independent of $\bar{X}$.

It is also possible to define the joint moment generating function of two or more random variables. This is done as follows: For any n random variables $X_1, \cdots, X_n$, the **joint moment generating function**, $M(t_1, \cdots, t_n)$, is defined, for all real values of $t_1, \cdots, t_n$, by

$$M(t_1, \cdots, t_n) = E\left[e^{t_1 X_1 + \cdots + t_n X_n}\right].$$

The individual moment generating functions can be obtained from $M(t_1,\cdots,t_n)$ by letting all but one of the t_j 's be 0. That is,

$$M_{X_i}(t) = E\left[e^{tX_i}\right] = M(0,\cdots,0,t,0,\cdots,0)$$

where the t is in the i-th position. It can be proven that the joint moment generating function $M(t_1,\cdots,t_n)$ uniquely determines the joint distribution of $X_1,\cdots,X_n$.

Example 5.55

Let X_1 and X_2 be random variables with joint moment generating function

$$M_{X_1,X_2}(t_1,t_2) = 0.1 + 0.2e^{5t_1} + 0.3e^{6t_2} + 0.4e^{7t_1+8t_2}.$$

What is $Cov(X_1, X_2)$?

Solution

Method 1:

$$E[X_1] = \left.\frac{\partial M_{X_1,X_2}(t_1,t_2)}{\partial t_1}\right|_{t_1=t_2=0} = \left.\left(0.2(5)e^{5t_1} + 0.4(7)e^{7t_1+8t_2}\right)\right|_{t_1=t_2=0} = 3.8$$
$$E[X_2] = \left.\frac{\partial M_{X_1,X_2}(t_1,t_2)}{\partial t_2}\right|_{t_1=t_2=0} = \left.\left(0.3(6)e^{6t_2} + 0.4(8)e^{7t_1+8t_2}\right)\right|_{t_1=t_2=0} = 5$$
$$E[X_1X_2] = \left.\frac{\partial^2 M_{X_1,X_2}(t_1,t_2)}{\partial t_1 \partial t_2}\right|_{t_1=t_2=0} = \left.\left(0.4(7)(8)e^{7t_1+8t_2}\right)\right|_{t_1=t_2=0} = 22.4$$
$$Cov(X_1, X_2) = 22.4 - 3.8(5) = 3.4$$

Method 2:

The joint probability mass function of X_1 and X_2 is

$$P(X_1 = 0,\, X_2 = 0) = 0.1, \quad P(X_1 = 5,\, X_2 = 0) = 0.2,$$
$$P(X_1 = 0,\, X_2 = 6) = 0.3, \quad P(X_1 = 7,\, X_2 = 8) = 0.4.$$

Hence

$$P(X_1 = 0) = 0.4, \quad P(X_1 = 5) = 0.2, \quad P(X_1 = 7) = 0.4$$
$$P(X_2 = 0) = 0.3, \quad P(X_2 = 6) = 0.3, \quad P(X_2 = 8) = 0.4$$

We have

$$E[X_1] = 0.2(5) + 0.4(7) = 3.8, \quad E[X_2] = 0.3(6) + 0.4(8) = 5, \quad E[X_1 X_2] = 0.4(7)(8) = 22.4$$
$$Cov(X_1, X_2) = 22.4 - 3.8(5) = 3.4$$

Theorem 5.56

Random variables $X_1, \cdots, X_n$ are independent if and only if $M(t_1, \cdots, t_n) = M_{X_1}(t_1) \cdots M_{X_n}(t_n)$.

Proof

If the *n* random variables are independent, then

$$M(t_1, \cdots, t_n) = E\left[e^{t_1 X_1 + \cdots + t_n X_n} \right] \overset{\text{independence}}{=} E\left[e^{t_1 X_1} \right] \cdots E\left[e^{t_n X_n} \right] = M_{X_1}(t_1) \cdots M_{X_n}(t_n).$$

Suppose $M(t_1, \cdots, t_n) = M_{X_1}(t_1) \cdots M_{X_n}(t_n)$. Then the joint moment generating function $M(t_1, \cdots, t_n)$ is the same as the joint moment generating function of *n* independent random variables, the *i*-th of which has the same distribution as X_i. As the joint moment generating function uniquely determines the joint distribution, this must be the joint distribution; hence, the random variables are independent.

Example 5.57

Let X and Y be independent normal random variables, each with mean μ and variance σ^2. Consider the joint moment generating function of $X + Y$ and $X - Y$.

$$\begin{aligned} M(s,t) &= E\left[e^{s(X+Y)+t(X-Y)} \right] \\ &= E\left[e^{(s+t)X+(s-t)Y} \right] \\ &= E\left[e^{(s+t)X} \right] E\left[e^{(s-t)Y} \right] \\ &= e^{\mu(s+t)+\sigma^2(s+t)^2/2} e^{\mu(s-t)+\sigma^2(s-t)^2/2} \\ &= e^{2\mu s+\sigma^2 s^2} e^{\sigma^2 t^2} \\ &= M_{X+Y}(s) M_{X-Y}(t) \end{aligned}$$

It follows that $X + Y$ and $X - Y$ are independent normal random variables. This result is compatible with result in Example 5.36.