

Multiple Random Variables

Chapter Intended Learning Outcomes:

- (i) Understand distribution functions of two random variables
- (ii) Able to generate jointly distributed random variables
- (iii) Able to compute expected values of jointly distributed random variables
- (iv) Know generalization to Gaussian random variable vectors
- (v) Able to apply multiple random variables to solve problems

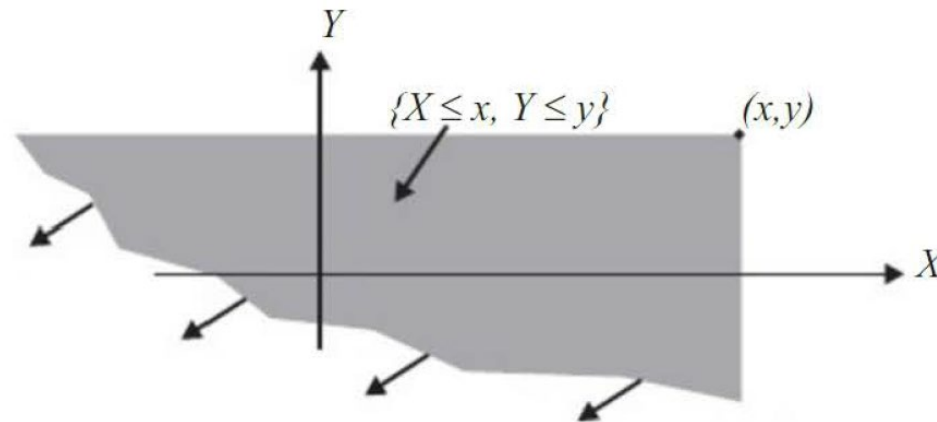
Two Random Variables

To handle multiple random variables (RVs), **joint** distributions are needed.

For two RVs X and Y , the **joint cumulative distribution function (CDF)** of (2.3) or (2.11), which describes the probability for a RV pair, is extended to:

$$F(x, y) = P(X \leq x, Y \leq y) = P_{XY}(X \leq x, Y \leq y) \quad (3.1)$$

For **continuous** RVs, it refers to the shaded region:



Is $F(x, y)$ 2-D or 3-D?

We also have:

$$0 \leq F(x, y) \leq 1$$

$$F(x, -\infty) = F(-\infty, y) = 0$$

$$F(\infty, \infty) = 1$$

The CDFs of $F(x)$ and $F(y)$ can also be determined from (3.1):

$$F(x) = P(X \leq x) = P(X \leq x, Y < \infty) = F(x, \infty) \quad (3.2)$$

$$F(y) = P(Y \leq y) = P(X < \infty, Y \leq y) = F(\infty, y) \quad (3.3)$$

For **discrete** RVs, the **joint probability mass function (PMF)** of X and Y is:

$$p(x, y) = P(X = x, Y = y) \geq 0 \quad (3.4)$$

The **marginal** PMFs of X and Y are obtained from (3.4) as:

$$p(x) = \sum_y p(x, y) \quad (3.5)$$

$$p(y) = \sum_x p(x, y) \quad (3.6)$$

Joint probability density function (PDF) of **continuous** X and Y is:

$$p(x, y) = \frac{\partial^2}{\partial x \partial y} F(x, y) = \frac{\partial^2}{\partial x \partial y} P(X \leq x, Y \leq y) \quad (3.7)$$

The **marginal** PDFs of X and Y can be obtained from (3.7) as:

$$p(x) = \int_{-\infty}^{\infty} p(x, y) dy \quad (3.8)$$

$$p(y) = \int_{-\infty}^{\infty} p(x, y) dx \quad (3.9)$$

We also have:

$$P(X \in A, Y \in B) = \int_B \int_A p(x, y) dx dy \quad (3.10)$$

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

$$P(Y \in B) = P(X \in (-\infty, \infty), Y \in B) = \int_B \int_{-\infty}^{\infty} p(x, y) dx dy = \int_B p(y) dy$$

Example 3.1

We denote X as the age of a child entering the first grade of a school. Similarly, we let Y be the age of a child entering the second grade.

The age for entering first grade is between 5 and 6, while that age for entering second grade is between 6 and 7.

The joint CDF of X and Y is:

$$F(x, y) = \begin{cases} 0, & x < 5 \\ 0, & y < 6 \\ (x - 5)(y - 6), & 5 \leq x < 6, 6 \leq y < 7 \\ y - 6, & x \geq 6, 6 \leq y < 7 \\ x - 5, & 5 \leq x < 6, y \geq 7 \\ 1, & \text{otherwise} \end{cases}$$

Find $F(x)$, $F(y)$, $p(x, y)$ as well as marginal PDFs $p(x)$ and $p(y)$.

From (3.2), we see that all values of Y are included. Hence:

$$F(x) = \begin{cases} 0, & x < 5 \\ x - 5, & 5 \leq x < 6 \\ 1, & x \geq 6 \end{cases}$$

Similarly, using (3.3), we have:

$$F(y) = \begin{cases} 0, & y < 6 \\ y - 6, & 6 \leq y < 7 \\ 1, & y \geq 7 \end{cases}$$

Applying (3.7), we easily obtain:

$$p(x, y) = \frac{\partial^2}{\partial x \partial y} F(x, y) = \begin{cases} 1, & 5 \leq x < 6, 6 \leq y < 7 \\ 0, & \text{otherwise} \end{cases}$$

where $p(x, y)$ is non-zero only for the valid ranges of first and second grade students.

We use (3.8) and (3.9) to compute $p(x)$ and $p(y)$ at their nonzero regions $5 \leq x < 6$ and $6 \leq y < 7$:

$$p(x) = \int_{-\infty}^{\infty} p(x, y) dy = \int_6^7 1 dy = 1 \Rightarrow p(x) = \begin{cases} 1, & 5 \leq x < 6 \\ 0, & \text{otherwise} \end{cases}$$

$$p(y) = \int_{-\infty}^{\infty} p(x, y) dx = \int_5^6 1 dx = 1 \Rightarrow p(y) = \begin{cases} 1, & 6 \leq y < 7 \\ 0, & \text{otherwise} \end{cases}$$

Using $p(x,y)$, can we get $F(x,y)$?

Using $F(x)$ and/or $F(y)$, can we get $F(x,y)$?

Using $p(x)$ and/or $p(y)$, can we get $p(x,y)$?

Analogous to the single-variable case, we have:

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(x, y) dx dy = 1 \quad (3.11)$$

$$\sum_x \sum_y p(x, y) = 1 \quad (3.12)$$

Similarly, we have:

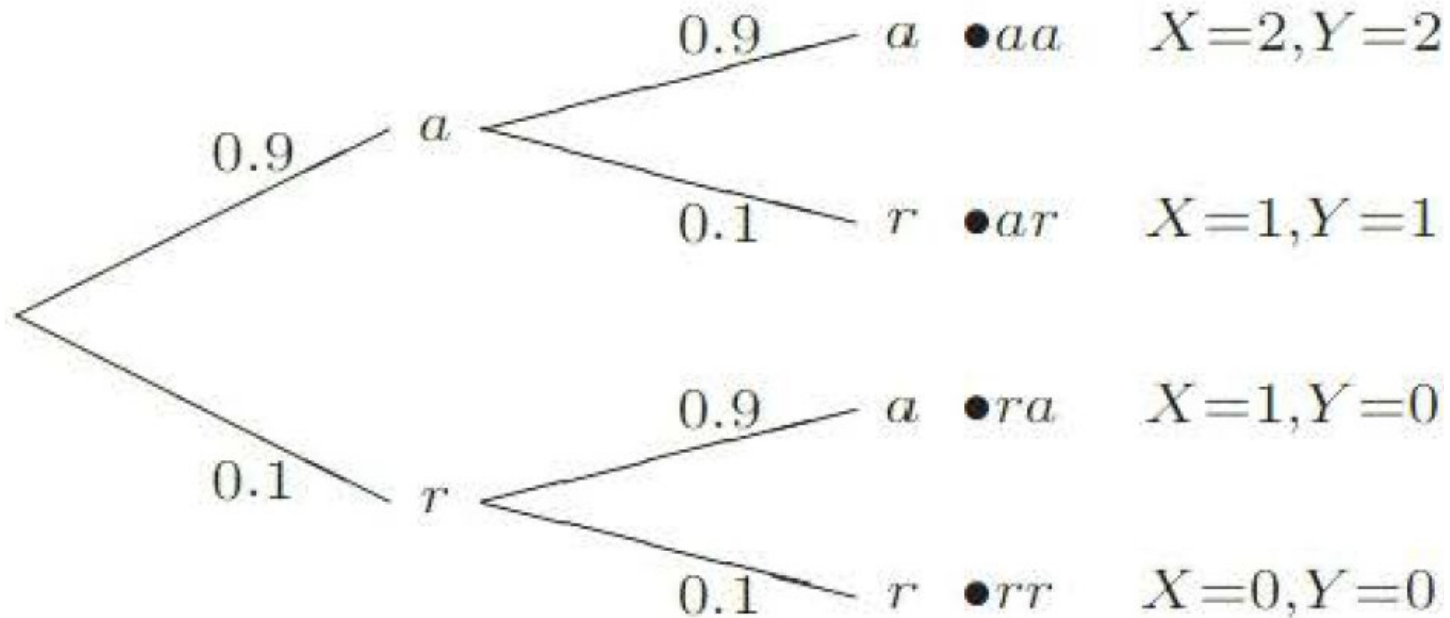
$$F(x, y) = \int_{-\infty}^y \int_{-\infty}^x p(u, v) du dv \quad (3.13)$$

$$F(x, y) = \sum_u^x \sum_v^y p(u, v) \quad (3.14)$$

Example 3.2

Consider testing 2 circuits one after the other. On each test, there are two possible outcomes, accept or reject, denoted by a and r , respectively. Assume the acceptance probability is 0.9 and the outcomes of successive tests are independent. Denote RVs X and Y as the number of accepted circuit(s) and the number of successful test(s) before observing the first reject. If both tests are successful, then $Y = 2$. Determine the joint PMF $p(x, y) = P(X = x, Y = y)$.

The admissible values of X and Y are 0, 1, and 2. We can draw a tree diagram for the probability computation:



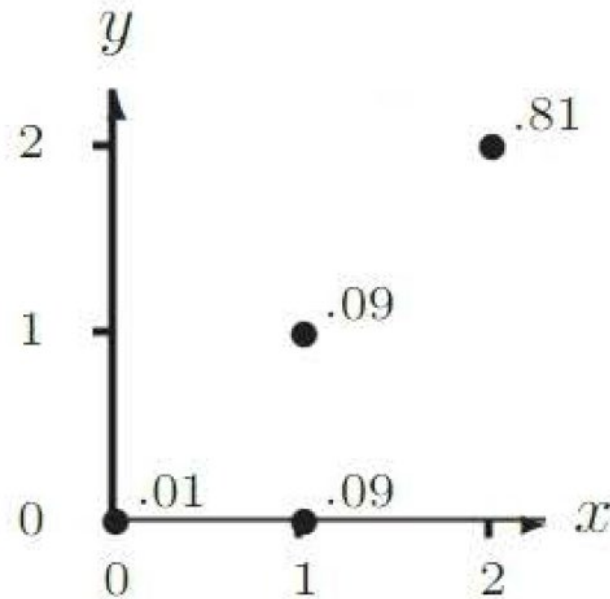
$$P(aa) = P(X = 2, Y = 2) = 0.9 \cdot 0.9 = 0.81$$

$$P(ar) = P(X = 1, Y = 1) = 0.9 \cdot 0.1 = 0.09$$

$$P(ra) = P(X = 1, Y = 0) = 0.1 \cdot 0.9 = 0.09$$

$$P(rr) = P(X = 0, Y = 0) = 0.1 \cdot 0.1 = 0.01$$

Graphically:



We can combine the results as:

$$p(x, y) = \begin{cases} 0.81, & x = 2, y = 2 \\ 0.09, & x = 1, y = 1 \\ 0.09, & x = 1, y = 0 \\ 0.01, & x = 0, y = 0 \\ 0, & \text{otherwise} \end{cases}$$

Representing in table form, we easily obtain marginal PMFs:

$P_{X,Y}(x, y)$	$y = 0$	$y = 1$	$y = 2$	$P_X(x)$
$x = 0$	0.01	0	0	0.01
$x = 1$	0.09	0.09	0	0.18
$x = 2$	0	0	0.81	0.81
$P_Y(y)$	0.10	0.09	0.81	

According to (3.5) and (3.6), summing along rows, we get the marginal PMF $p(x)$, while summing along columns, we get the marginal PMF $p(y)$:

$$p(x) = \begin{cases} 0.01, & x = 0 \\ 0.18, & x = 1 \\ 0.81, & x = 2 \\ 0, & \text{otherwise} \end{cases}$$

and

$$p(y) = \begin{cases} 0.1, & y = 0 \\ 0.09, & y = 1 \\ 0.81, & y = 2 \\ 0, & \text{otherwise} \end{cases}$$

We may see that $p(x)$ corresponds to the probability of 0, 1, and 2 accepted circuits for $x = 0, 1, 2$, respectively.

We can see:

$$\sum_x \sum_y p(x, y) = 1$$

$$\sum_x p(x) = \sum_y p(y) = 1$$

Can we get $F(x, y)$ from $p(x, y)$?

Example 3.3

Suppose two RVs X and Y have joint PDF:

$$p(x, y) = \begin{cases} c, & 0 \leq x \leq 5, 0 \leq y \leq 3 \\ 0, & \text{otherwise} \end{cases}$$

Determine the value of c . Then find $P(2 \leq X < 3, 1 \leq Y < 3)$. Also find $P(Y > X)$.

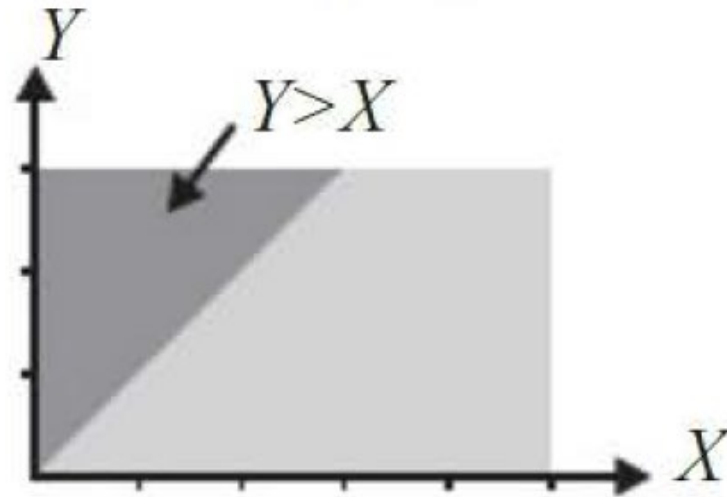
Applying (3.11), we have

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(x, y) dx dy = \int_0^3 \int_0^5 c dx dy = 15c = 1 \Rightarrow c = \frac{1}{15}$$

Using (3.10):

$$P(2 \leq X < 3, 1 \leq Y < 3) = \int_1^3 \int_2^3 \frac{1}{15} dx dy = 2 \cdot 1 \cdot \frac{1}{15} = \frac{2}{15}$$

The region of $Y > X$ can be illustrated as follows:



Hence $P(Y > X)$ is obtained based on the triangle area:

$$P(Y > X) = c \cdot 3 \cdot 3 \div 2 = \frac{3}{10}$$

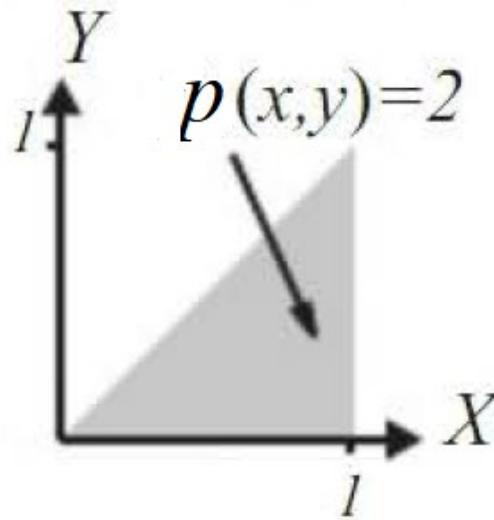
Alternatively, (3.10) can be applied:

$$P(Y > X) = \int_0^3 \left[\int_x^3 \frac{1}{15} dy \right] dx = \int_0^3 \frac{3-x}{15} dx = -\frac{(3-x)^2}{30} \Big|_0^3 = \frac{3}{10}$$

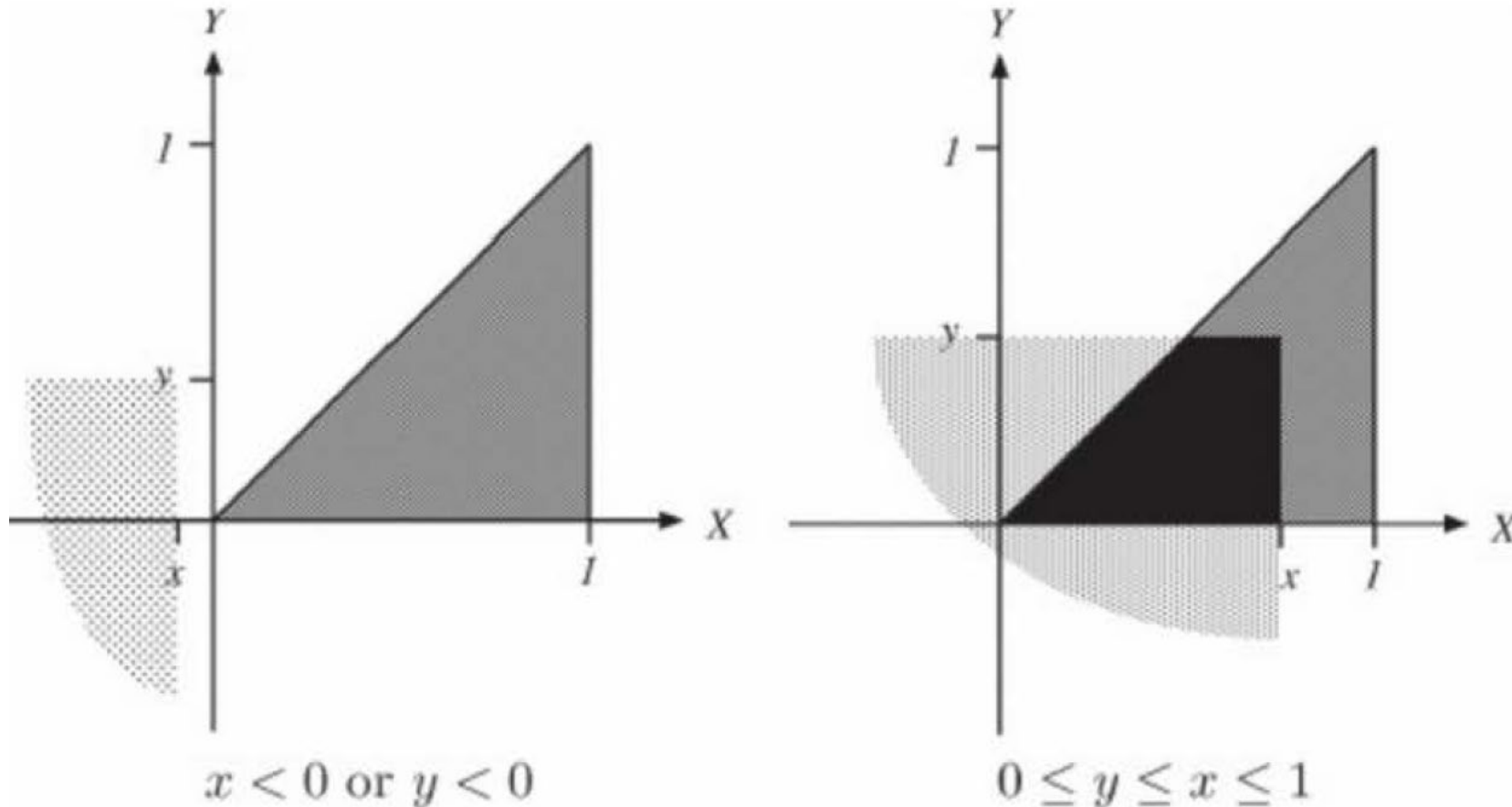
Example 3.4

Find the joint CDF $F(x, y)$ when RVs X and Y have joint PDF:

$$p(x, y) = \begin{cases} 2, & 0 \leq y \leq x \leq 1 \\ 0, & \text{otherwise} \end{cases}$$



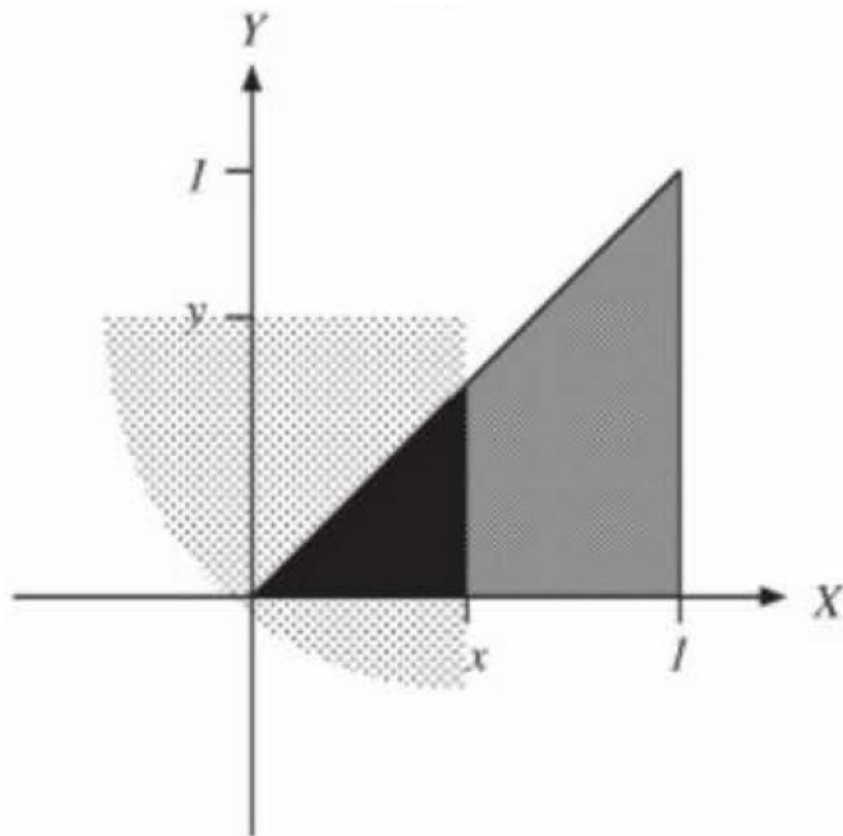
To find the CDF, we apply (3.13) with considering different regions of x and y .



For $x < 0$ or $y < 0$, it is clear that $F(x, y) = 0$.

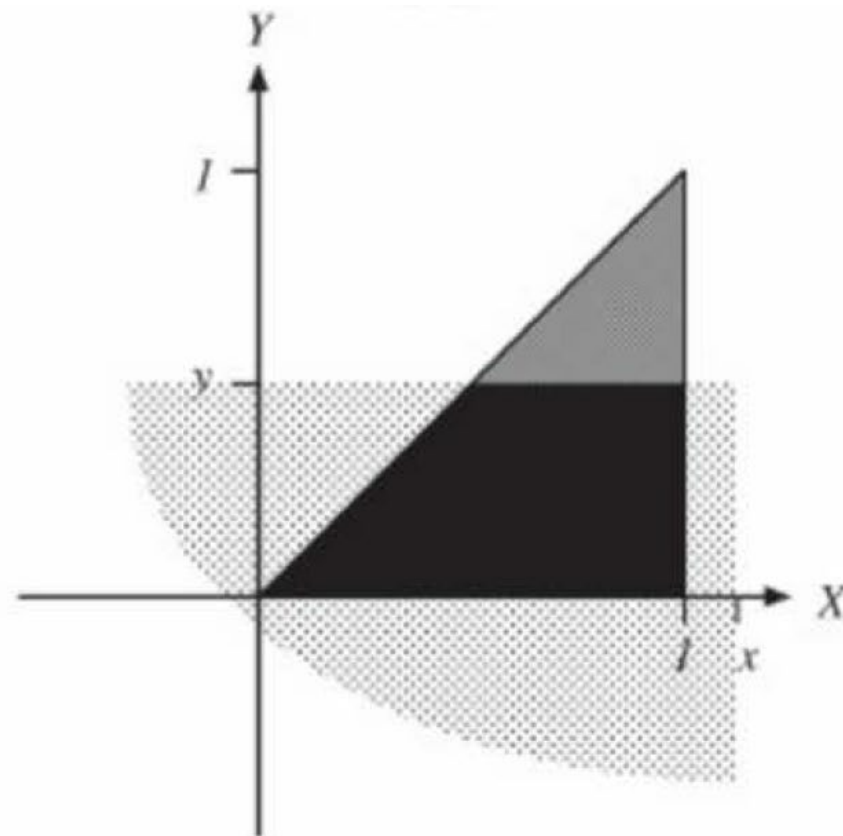
For $0 \leq y \leq x \leq 1$, applying (3.13) yields:

$$F(x, y) = \int_0^y \int_v^x 2dudv = \int_0^y 2(x - v)dv = 2xv - v^2 \Big|_0^y = 2xy - y^2$$



$$0 \leq x < y$$

$$0 \leq x \leq 1$$



$$0 \leq y \leq 1$$

$$x > 1$$

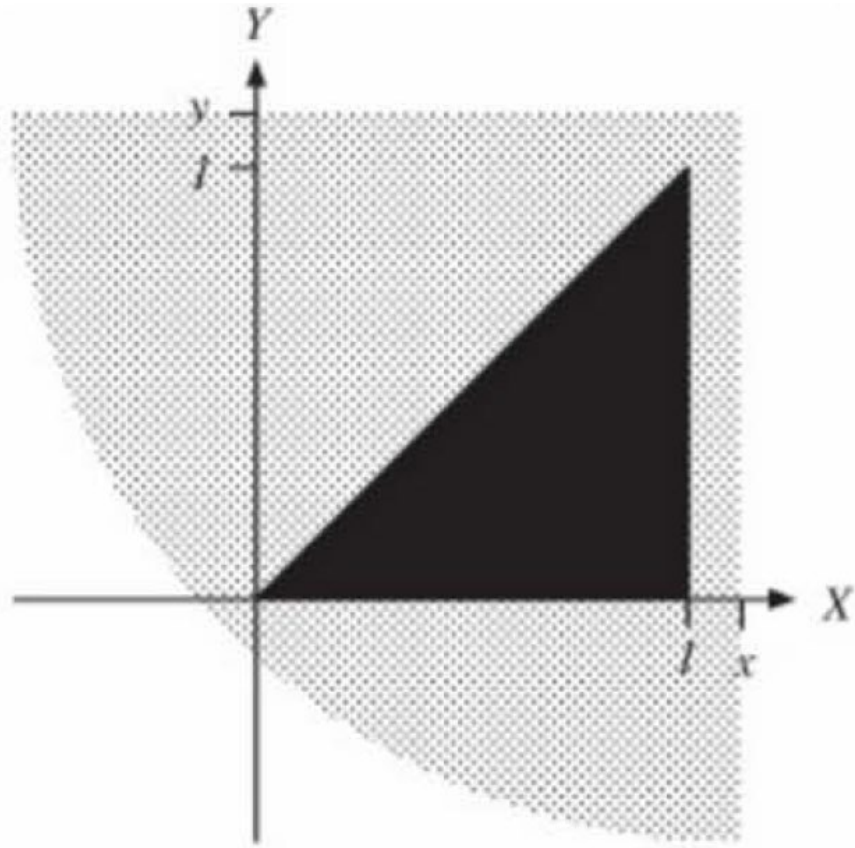
For $0 \leq x < y$ and $0 \leq x \leq 1$, applying (3.13) yields:

$$F(x, y) = \int_0^x \int_v^x 2 \, du \, dv = \int_0^x 2(x - v) \, dv = 2xv - v^2 \Big|_0^x = x^2$$

For $0 \leq y \leq 1$ and $x > 1$, we have:

$$F(x, y) = \int_0^y \int_v^1 2dudv = \int_0^y 2(1 - v)dv = 2v - v^2 \Big|_0^y = 2y - y^2$$

The final case corresponds to $x > 1$ and $y > 1$, i.e., $F(x, y) = 1$.



Combining the results, we have:

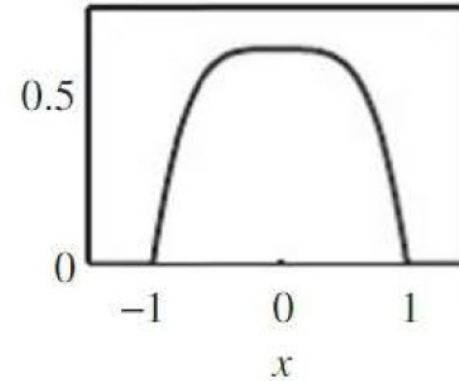
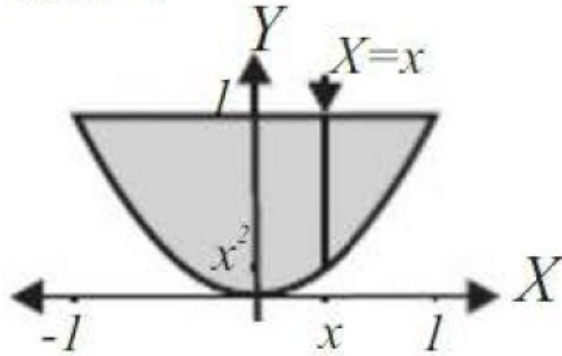
$$F(x, y) = \begin{cases} 0, & x < 0 \text{ or } y < 0 \\ 2xy - y^2, & 0 \leq y \leq x \leq 1 \\ x^2, & 0 \leq x < y, 0 \leq x \leq 1 \\ 2y - y^2, & 0 \leq y \leq 1, x > 1 \\ 1, & x > 1, y > 1 \end{cases}$$

Example 3.5

Suppose the joint PDF of two RVs X and Y is:

$$p(x, y) = \begin{cases} 5y/4, & -1 \leq x \leq 1, x^2 \leq y \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

Determine the marginal PDFs $p(x)$ and $p(y)$.



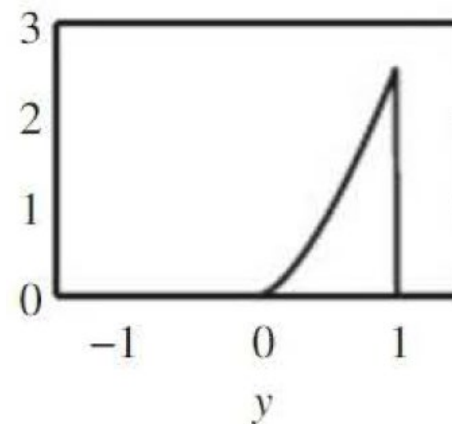
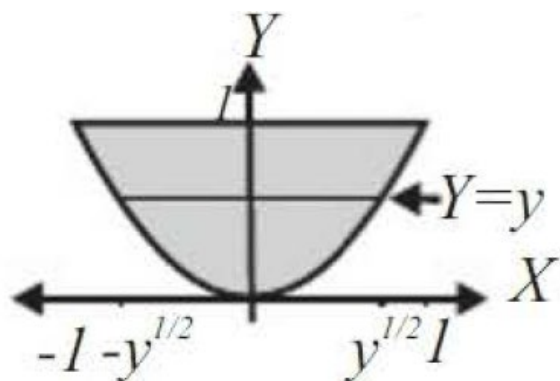
We know that $p(x) = 0$ for $x < -1$ or $x > 1$.

For $-1 \leq x \leq 1$, we apply (3.8) with the range of $x^2 \leq y \leq 1$:

$$p(x) = \int_{-\infty}^{\infty} p(x, y) dy = \int_{x^2}^1 \frac{5y}{4} dy = \frac{5(1 - x^4)}{8}$$

Hence:

$$p(x) = \begin{cases} \frac{5(1 - x^4)}{8}, & -1 \leq x \leq 1 \\ 0, & \text{otherwise} \end{cases}$$



We also note that $p(y) = 0$ for $y < 0$ or $y > 1$, and the range boundaries for $0 \leq y \leq 1$ are $x = -\sqrt{y}$ and $x = \sqrt{y}$. Using (3.9):

$$p(y) = \int_{-\infty}^{\infty} p(x, y) dx = \int_{-\sqrt{y}}^{\sqrt{y}} \frac{5y}{4} dx = \frac{5y}{4} \int_{-\sqrt{y}}^{\sqrt{y}} dx = \frac{5y^{3/2}}{2}$$

Hence:

$$p(y) = \begin{cases} \frac{5y^{3/2}}{2}, & 0 \leq y \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

Expectation of Two Random Variables

The expectation formulas of (2.24)-(2.25) for two RVs can be generalized as:

$$\mathbb{E}\{g(X, Y)\} = \sum_y \sum_x g(x, y)p(x, y) \quad (3.15)$$

$$\mathbb{E}\{g(X, Y)\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y)p(x, y)dx dy \quad (3.16)$$

Example 3.6

Consider continuous RVs X and Y , compute $\mathbb{E}\{aX + bY\}$ for any constants a and b .

Using (3.16), we have $g(X, Y) = aX + bY$:

$$\begin{aligned}
& \mathbb{E}\{aX + bY\} \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (ax + by)p(x, y)dx dy \\
&= a \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xp(x, y)dx dy + b \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} yp(x, y)dx dy \\
&= a \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} p(x, y)dy \right] x dx + b \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} p(x, y)dx \right] y dy \\
&= a \int_{-\infty}^{\infty} p(x)x dx + b \int_{-\infty}^{\infty} p(y)y dy = a\mathbb{E}\{X\} + b\mathbb{E}\{Y\}
\end{aligned}$$

This indicates that expectation is a **linear** operator. Note that the same result can be obtained for discrete RVs X and Y .

In general, for continuous or discrete RVs X_1, \dots, X_n , we have:

$$\mathbb{E}\{a_1X_1 + \dots + a_nX_n\} = a_1\mathbb{E}\{X_1\} + \dots + a_n\mathbb{E}\{X_n\} \quad (3.17)$$

Example 3.7

Based on joint distribution, compute the expected value of the sum of the faces in the experiment of rolling 2 fair dice.

We assign RVs X_1 and X_2 as the faces of the first and second dice. It is clear that $1 \leq X_1, X_2 \leq 6$, and thus

$$\mathbb{E}\{X_1\} = \mathbb{E}\{X_2\} = \frac{1 + 2 + 3 + 4 + 5 + 6}{6} = 3.5$$

Let X be the sum of two faces, which is $X_1 + X_2$. Using (3.17), the expected value is:

$$\mathbb{E}\{X\} = \mathbb{E}\{X_1 + X_2\} = \mathbb{E}\{X_1\} + \mathbb{E}\{X_2\} = 7$$

which is identical to the result computed in Example 2.19.

It is clear that applying (3.17) is more efficient than using a single RV to represent the sum in Example 2.19.

Independent Random Variables

Two RVs X and Y are **independent** if

$$F(x, y) = P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y) = F(x)F(y) \quad (3.18)$$

That is, the **events** of $X \leq x$ and $Y \leq y$ are **independent** for all x and y .

The joint PMF or PDF becomes:

$$p(x, y) = p(x)p(y) \quad (3.19)$$

For independent X and Y , we also have:

$$\mathbb{E}\{g(X)h(Y)\} = \mathbb{E}\{g(X)\}\mathbb{E}\{h(Y)\} \quad (3.20)$$

where g and h are any functions.

Example 3.8

Investigate the children age RVs X and Y in Example 3.1 are independent or not.

We can use either (3.18) or (3.19) for investigation.

Recall the results in Example 3.1:

$$p(x, y) = \begin{cases} 1, & 5 \leq x < 6, 6 \leq y < 7 \\ 0, & \text{otherwise} \end{cases}$$

$$p(x) = \begin{cases} 1, & 5 \leq x < 6 \\ 0, & \text{otherwise} \end{cases} \quad p(y) = \begin{cases} 1, & 6 \leq y < 7 \\ 0, & \text{otherwise} \end{cases}$$

It is clear that $p(x, y) = p(x)p(y)$. According to (3.19), X and Y are independent. Note that same conclusion can be obtained by using $F(x)$, $F(y)$ and $F(x, y)$.

Example 3.9

Check whether X and Y are independent if their joint PDF is:

$$p(x, y) = \begin{cases} 5y/4, & -1 \leq x \leq 1, x^2 \leq y \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

Recall the results in Example 3.5:

$$p(x) = \begin{cases} \frac{5(1 - x^4)}{8}, & -1 \leq x \leq 1 \\ 0, & \text{otherwise} \end{cases} \quad p(y) = \begin{cases} \frac{5y^{3/2}}{2}, & 0 \leq y \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

Apparently, $p(x, y) \neq p(x)p(y)$. Hence X and Y are not independent.

Covariance of Random Variables

The covariance of two RVs X and Y , denoted by $\text{cov}(X, Y)$ is:

$$\begin{aligned}\text{cov}(X, Y) &= \mathbb{E}\{(X - \mathbb{E}\{X\})(Y - \mathbb{E}\{Y\})\} \\ &= \mathbb{E}\{XY - Y\mathbb{E}\{X\} - X\mathbb{E}\{Y\} + \mathbb{E}\{X\}\mathbb{E}\{Y\}\} \\ &= \mathbb{E}\{XY\} - \mathbb{E}\{X\}\mathbb{E}\{Y\}\end{aligned}\quad (3.21)$$

If X and Y are **independent**, then $\text{cov}(X, Y) = 0$.

Properties of covariance include:

$$\begin{aligned}\text{cov}(X, X) &= \text{var}(X) \\ \text{cov}(X, Y) &= \text{cov}(Y, X) \\ \text{cov}(aX, Y) &= a\text{cov}(X, Y), \text{ } a \text{ is constant} \\ \text{cov}(X, Y + Z) &= \text{cov}(X, Y) + \text{cov}(X, Z), \text{ } Z \text{ is RV}\end{aligned}$$

Example 3.10

Prove

$$\text{var}(aX) = a^2\text{var}(X)$$

Using the variance definition, the proof is given as follows:

$$\begin{aligned}\text{var}(aX) &= \mathbb{E}\{(aX - \mathbb{E}\{aX\})^2\} \\ &= \mathbb{E}\{(aX - a\mathbb{E}\{X\})^2\} \\ &= \mathbb{E}\{(a(X - \mathbb{E}\{X\}))^2\} \\ &= a^2\mathbb{E}\{(X - \mathbb{E}\{X\})^2\} = a^2\text{var}(X)\end{aligned}$$

This aligns with

$$\begin{aligned}\text{cov}(X, X) &= \text{var}(X) \\ \text{cov}(aX, Y) &= a\text{cov}(Y, X), \text{ } a \text{ is constant}\end{aligned}$$

The covariance can be generalized to **sum** of RVs:

$$\text{COV} \left(\sum_{i=1}^m X_i, \sum_{j=1}^n Y_j \right) = \sum_{i=1}^m \sum_{j=1}^n \text{COV}(X_i, Y_j) \quad (3.22)$$

In particular, the variance is

$$\begin{aligned} \text{var} \left(\sum_{i=1}^n X_i \right) &= \sum_{i=1}^n \sum_{j=1}^n \text{COV}(X_i, X_j) \\ &= \sum_{i=1}^n \text{var}(X_i) + 2 \sum_{i=1}^n \sum_{j<i}^n \text{COV}(X_i, X_j) \end{aligned} \quad (3.23)$$

When X_1, \dots, X_n are independent, then (3.23) becomes:

$$\text{var} \left(\sum_{i=1}^n X_i \right) = \sum_{i=1}^n \text{var}(X_i) \quad (3.24)$$

Example 3.11

Suppose that X_1, \dots, X_n are independent and identically distributed (IID) with mean $\mathbb{E}\{X_1\} = \dots = \mathbb{E}\{X_n\} = \mu$ and variance σ^2 . Let the average or sample mean be $\bar{X} = \sum_{i=1}^n X_i/n$. Compute:

$$\mathbb{E}\{\bar{X}\}$$

$$\text{var}(\bar{X})$$

$$\text{cov}(\bar{X}, X_i - \bar{X}), \quad i = 1, \dots, n$$

Applying (3.17) and since X_1, \dots, X_n are independent, we get:

$$\mathbb{E}\{\bar{X}\} = \frac{1}{n} \mathbb{E} \left\{ \sum_{i=1}^n X_i \right\} = \frac{1}{n} \sum_{i=1}^n \mathbb{E}\{X_i\} = \mu$$

$$\text{var}(\bar{X}) = \left(\frac{1}{n} \right)^2 \text{var} \left(\sum_{i=1}^n X_i \right) = \left(\frac{1}{n} \right)^2 \sum_{i=1}^n \text{var}(X_i) = \frac{\sigma^2}{n}$$

$$\begin{aligned}
\text{cov}(\bar{X}, X_i - \bar{X}) &= \text{cov}(\bar{X}, X_i) - \text{cov}(\bar{X}, \bar{X}) \\
&= \frac{1}{n} \text{cov} \left(X_i + \sum_{j \neq i} X_j, X_i \right) - \text{var}(\bar{X}) \\
&= \frac{1}{n} \text{cov}(X_i, X_i) + \frac{1}{n} \text{cov} \left(\sum_{j \neq i} X_j, X_i \right) - \frac{\sigma^2}{n} \\
&= \frac{\sigma^2}{n} - 0 - \frac{\sigma^2}{n} = 0, \quad i = 1, \dots, n
\end{aligned}$$

Hence we see that \bar{X} has the same mean with X_1, \dots, X_n , but its variance reduces by n times.

It can also be shown that \bar{X} and $X_i - \bar{X}$ are independent.

We may realize the findings with the use of MATLAB. Consider 10 IID RVs with $X_i \sim U[0, 1]$ as an illustration.

From Example 2.21 and (2.23), it is clear that

$$\mu = \mathbb{E}\{X_i\} = 0.5$$

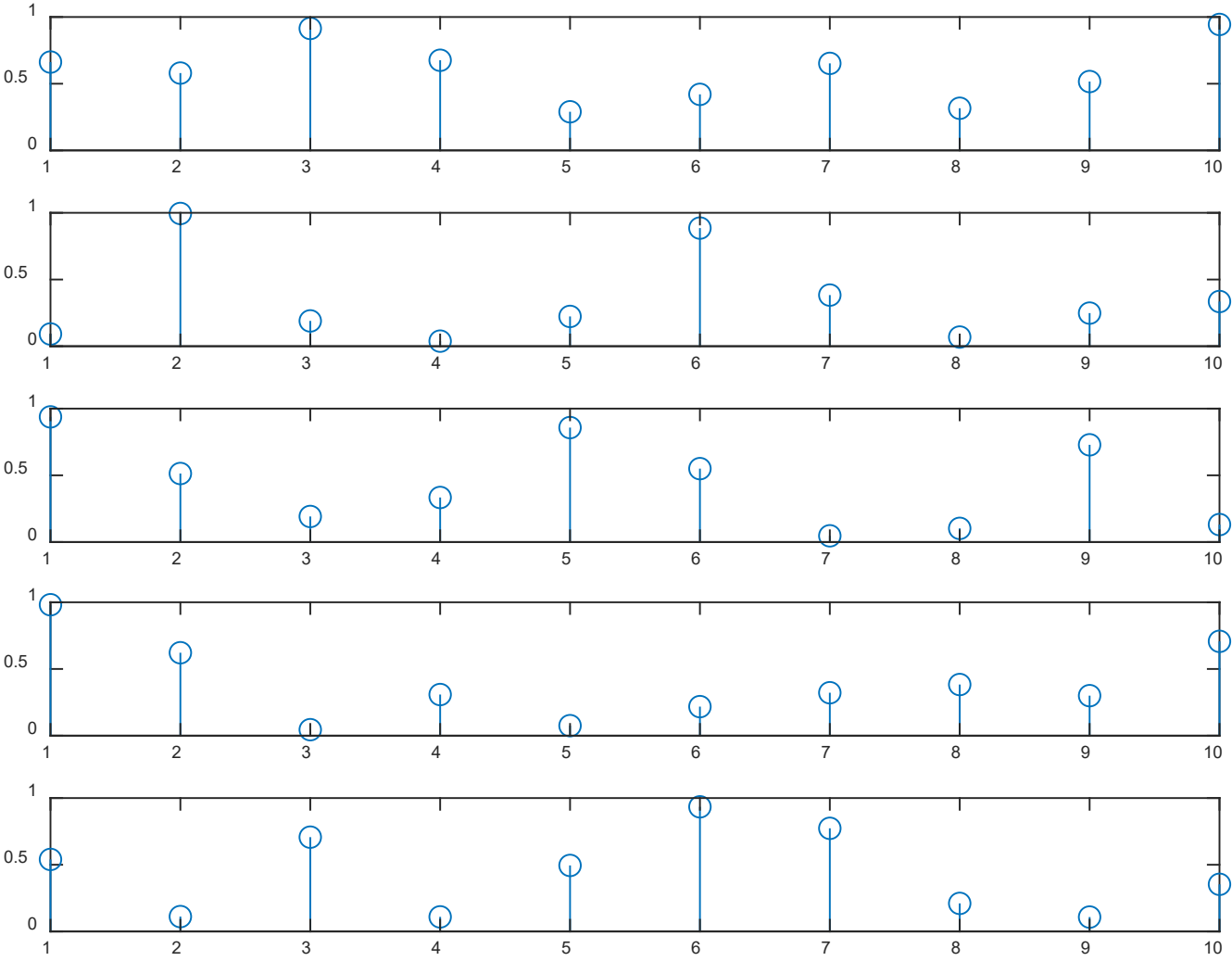
$$\sigma^2 = \text{var}(X_i) = 1/3 - 0.5^2 = 1/12$$

One realization of X_1, \dots, X_{10} can be:

```
>> X=rand(1,10)
X = 0.2460    0.5852    0.8008    0.6085    0.6470
0.0028    0.0935    0.9088    0.9466    0.4870
>> barX = mean(X)
barX = 0.5326
```

$$\Rightarrow \bar{X} \approx \mu$$

Use multiple realizations:



Considering 1000 sets of realization

```
>> X=rand(1000,10); %each row is 1 realization;  
>> barX=mean(X,2); %vector with 1000 mean values
```

```
>> mean(barX) %mean of 1000 mean values  
ans = 0.5037
```

which is close to μ

```
>> var(barX,1) %variance of 1000 mean values  
ans = 0.0087
```

which is close to $\sigma^2/n = 1/12000$

Example 3.12

Use the concepts in joint distribution to find $\mathbb{E}\{X\}$ and variance for binomial RV X which is characterized by n and p .

Recall that the binomial RV extends the Bernoulli RV by combining n independent Bernoulli trials, we can write X as:

$$X = X_1 + \cdots + X_n$$

where each of the X_1, \cdots, X_n is a Bernoulli RV according to (2.4) whose value is 1 for a success and 0 otherwise.

The expected value is:

$$\mathbb{E}\{X_1\} = \cdots = \mathbb{E}\{X_n\} = 1 \cdot p + 0 \cdot (1 - p) = p$$

Applying (3.17) yields:

$$\mathbb{E}\{X\} = \mathbb{E}\{X_1 + \cdots + X_n\} = \mathbb{E}\{X_1\} + \cdots + \mathbb{E}\{X_n\} = np$$

Since X_1, \cdots, X_n are independent, we use (3.24):

$$\text{var}(X) = \text{var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \text{var}(X_i)$$

Noting from (2.23) that:

$$\begin{aligned}\text{var}(X_i) &= \mathbb{E}\{X_i^2\} - (\mathbb{E}\{X_i\})^2 \\ &= \mathbb{E}\{X_i\} - (\mathbb{E}\{X_i\})^2, \quad X_i^2 = X_i \\ &= p - p^2 = p(1 - p)\end{aligned}$$

Hence

$$\text{var}(X) = np(1 - p)$$

The results align with Example 2.30.

Example 3.13

Repeat Example 3.12 for

$$\bar{X} = \frac{1}{n}X = \frac{X_1 + \cdots + X_n}{n}$$

We can make use of the previous results directly:

$$\mathbb{E}\{\bar{X}\} = \mathbb{E}\left\{\frac{1}{n}X\right\} = \frac{1}{n}\mathbb{E}\{X\} = \frac{1}{n} \cdot np = p$$

$$\text{var}(\bar{X}) = \text{var}\left(\frac{1}{n}X\right) = \text{cov}\left(\frac{1}{n}X, \frac{1}{n}X\right) = \frac{1}{n^2}\text{cov}(X, X) = \frac{p(1-p)}{n}$$

When an average of n RVs is used, the mean is decreased by n times and the variance is reduced by n^2 . But comparing with X_i , the mean is the same while the variance is reduced by n .

Example 3.14

Suppose a company Web has 3 pages, requiring 750 kilobytes, 1500 kilobytes, and 2500 kilobytes for transmission. The transmission speeds are 5 and 10 megabits per second external and internal requests, respectively. Requests arrive randomly from inside and outside the company independent of page length, which is also random. The probability models for transmission speed and page length, denoted by R and L , respectively, are:

$$P_R(r) = \begin{cases} 0.4, & r = 5 \\ 0.6, & r = 10 \\ 0, & \text{otherwise} \end{cases} \quad P_L(l) = \begin{cases} 0.3, & l = 750 \\ 0.5, & l = 1500 \\ 0.2, & l = 2500 \\ 0, & \text{otherwise} \end{cases}$$

Determine the transmission time $T = g(R, L)$. Then find $\mathbb{E}\{T\}$.

Aligning the units, i.e., 1 byte = 8 bits, and 1 mega = 1000 kilo, we easily obtain:

$$T = g(R, L) = \frac{8L}{1000R}$$

As R and L are independent, $P_{R,L}(r, l) = P_R(r)P_L(l)$, we then apply (3.15):

$$\begin{aligned}\mathbb{E}\{g(R, L)\} &= \sum_r \sum_l g(r, l)P_{R,L}(r, l) = \sum_r \sum_l g(r, l)P_R(r)P_L(l) \\ &= \frac{8}{1000} \left(\sum_r \frac{P_R(r)}{r} \right) \left(\sum_l P_L(l)l \right) \\ &= 1.652\end{aligned}$$

Note that $\mathbb{E}\{g(R, L)\} \neq g(\mathbb{E}\{R\}, \mathbb{E}\{L\})$.

The **normalized** version of $\text{cov}(X, Y)$ in (3.21) is called **correlation coefficient** $\rho_{X,Y}$:

$$\rho_{X,Y} = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X)}\sqrt{\text{var}(Y)}} = \frac{\text{cov}(X, Y)}{\sigma_X\sigma_Y} \in [-1, 1] \quad (3.25)$$

Apart from bounded between 1 and -1, $\rho_{X,Y}$ has no unit.

Note that the **correlation** of X and Y is defined as:

$$r_{X,Y} = \mathbb{E}\{XY\} \quad (3.26)$$

If $r_{X,Y} = 0$, X and Y are said to be **orthogonal**.

If $\text{cov}(X, Y) = 0$ or $\rho_{X,Y} = 0$, X and Y are said to be **uncorrelated**.

Hence if X and Y are **independent**, they are **uncorrelated**.

Example 3.15

Given a noisy observation $Y = S + N$ where $S \sim \mathcal{N}(0, \sigma_s^2)$ is the signal component and $N \sim \mathcal{N}(0, \sigma_n^2)$ is the additive noise. Furthermore, S and N are independent. Determine $\text{cov}(S, Y)$, $\rho_{S,N}$ and $\rho_{S,Y}$.

As S and N are independent, $\text{cov}(S, N) = 0$. Hence we have:

$$\rho_{S,N} = 0$$

$$\text{cov}(S, Y) = \text{cov}(S, S + N) = \text{cov}(S, S) + \text{cov}(S, N) = \text{var}(S) = \sigma_s^2$$

Using (3.24), we obtain:

$$\text{var}(Y) = \text{var}(S + N) = \text{var}(S) + \text{var}(N) = \sigma_s^2 + \sigma_n^2$$

$$\Rightarrow \rho_{S,Y} = \frac{\text{cov}(S, Y)}{\sqrt{\text{var}(S)}\sqrt{\text{var}(Y)}} = \frac{\sigma_s^2}{\sqrt{\sigma_s^2 \cdot (\sigma_s^2 + \sigma_n^2)}} = \sqrt{\frac{\text{SNR}}{1 + \text{SNR}}}, \quad \text{SNR} = \frac{\sigma_s^2}{\sigma_n^2}$$

SNR stands for **signal-to-noise ratio** which is the ratio of signal power and noise power.

When $\text{SNR} \gg 1$ or signal dominates, $\rho_{S,Y} \rightarrow 1$. When $\text{SNR} \ll 1$, or noise dominates, $\rho_{S,Y} \rightarrow 0$.

What do $\rho_{S,Y} \rightarrow 1$ and $\rho_{S,Y} \rightarrow 0$ mean for S and Y ?

```
>> sigmaS2=100;
sigmaN2=10;
S=sqrt(sigmaS2).*randn(1,100000);
N=sqrt(sigmaN2).*randn(1,100000);
Y=S+N;
>> cYS=mean(Y.*S)-mean(Y)*mean(S)
cYS = 99.8749
var(Y,1) = 109.7872
var(S,1) = 99.9341
cYS/sqrt(var(Y,1)*var(S,1)) = 0.9535
```

which is close to $\sqrt{10/11}$.

Bias and Mean Square Error

In practical scenarios, we may not be able to obtain a value of interest, say, A , from our observations, without any uncertainty. Instead, we can get an **estimate** of A , denoted by \hat{A} , and it is desired that $\hat{A} \approx A$.

There are two conventional metrics to assess \hat{A} : **bias** and **mean square error (MSE)**, denoted by $\text{bias}(\hat{A})$ and $\text{MSE}(\hat{A})$:

$$\text{bias}(\hat{A}) = \mathbb{E}\{\hat{A}\} - A \quad (3.27)$$

and

$$\text{MSE}(\hat{A}) = \mathbb{E}\{(\hat{A} - A)^2\} \quad (3.28)$$

When $\text{bias}(\hat{A}) = 0$ or $E\{\hat{A}\} = A$, \hat{A} is an **unbiased** estimate of A . Furthermore, we desire $\text{MSE}(\hat{A})$ as **small** as possible.

From (3.28), we have:

$$\begin{aligned}\text{MSE}(\hat{A}) &= \mathbb{E}\{([\hat{A} - \mathbb{E}\{\hat{A}\}] + [\mathbb{E}\{\hat{A}\} - A])^2\} \\ &= \mathbb{E}\{(\hat{A} - \mathbb{E}\{\hat{A}\})^2\} + 2\mathbb{E}\{[\hat{A} - \mathbb{E}\{\hat{A}\}][\mathbb{E}\{\hat{A}\} - A]\} + (\mathbb{E}\{\hat{A}\} - A)^2 \\ &= \text{var}(\hat{A}) + 2[\mathbb{E}\{\hat{A}\} - \mathbb{E}\{\hat{A}\}][\mathbb{E}\{\hat{A}\} - A] + (\text{bias}(\hat{A}))^2 \\ &= \text{var}(\hat{A}) + (\text{bias}(\hat{A}))^2\end{aligned}$$

which can also be written as:

$$\text{var}(\hat{A}) = \text{MSE}(\hat{A}) - (\text{bias}(\hat{A}))^2 \quad (3.29)$$

which is analogous to (2.23). Note that when \hat{A} is an **unbiased** estimate, $\text{bias}(\hat{A}) = 0$, (3.29) becomes:

$$\text{var}(\hat{A}) = \text{MSE}(\hat{A}) \quad (3.30)$$

Example 3.16

Consider a very large population of N people in a city. Among them, Np people support a certain proposition while $N(1 - p)$ oppose it. The task is to find p . As it is not feasible to ask all individuals, only a small fraction of the population, n of them, are randomly selected for estimating p . That is, if $m \leq n$ among the sampled population support the proposition, the estimate of p is determined as $\hat{p} = m/n$. Compute the mean, variance and mean square error of \hat{p} .

Assign RVs $X_i, i = 1, \dots, n$, for the n sampled people such that $X_i = 1$ if he supports the proposition and $X_i = 0$ if he opposes it. Then \hat{p} is computed as:

$$\hat{p} = \frac{1}{n} \sum_{i=1}^n X_i$$

$\mathbb{E}\{\hat{p}\}$ is easily obtained as:

$$\mathbb{E}\{\hat{p}\} = \mathbb{E}\left\{\frac{1}{n} \sum_{i=1}^n X_i\right\} = \frac{1}{n} \sum_{i=1}^n \mathbb{E}\{X_i\} = \frac{1}{n} \sum_{i=1}^n p = p$$

Next, we compute $\text{var}\left(\sum_{i=1}^n X_i\right)$ using (3.23):

$$\text{var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \text{var}(X_i) + 2 \sum_{i=1}^n \sum_{j<i} \text{cov}(X_i, X_j)$$

For the first term, we apply $\text{var}(X_i) = p(1 - p)$ for Bernoulli events in Example 3.12:

$$\sum_{i=1}^n \text{var}(X_i) = np(1 - p)$$

For $i \neq j$, we use (3.21) to compute $\text{cov}(X_i, X_j)$:

$$\begin{aligned}\text{cov}(X_i, X_j) &= \mathbb{E}\{X_i X_j\} - E\{X_i\}E\{X_j\} \\ &= P(X_i = 1 \cap X_j = 1) - p^2 \\ &= P(X_i = 1) \cdot P(X_j = 1 | X_i = 1) - p^2 \\ &= \frac{Np}{N} \cdot \frac{Np - 1}{N - 1} - p^2\end{aligned}$$

Combining the results yields:

$$\begin{aligned}\text{var}\left(\sum_{i=1}^n X_i\right) &= np(1-p) + n(n-1) \left[\frac{Np}{N} \cdot \frac{Np-1}{N-1} - p^2 \right] \\ &= np(1-p) - \frac{n(n-1)p(1-p)}{N-1}\end{aligned}$$

Finally, we get:

$$\text{var}(\hat{p}) = \text{var}\left(\sum_{i=1}^n X_i/n\right) = \frac{p(1-p)}{n} - \frac{(n-1)p(1-p)}{n(N-1)}$$

Hence we see that \hat{p} is an unbiased estimate of p while $\text{var}(\hat{p}) = \text{MSE}(\hat{p})$ decreases with increasing n .

Note that

$$\text{var}(\hat{p}) = \frac{p(1-p)}{n}, \quad N \rightarrow \infty$$

The variance result aligns with Example 3.13.

What is the difference between Examples 3.13 and 3.16?

White Noise

It is an **independent and identically distributed (IID)** random sequence $\{q_1, \dots, q_N\}$, typically used in modeling **additive noise**, which has the following properties:

$$\mathbb{E}\{q_n\} = 0 \quad (3.31)$$

and

$$\mathbb{E}\{q_n q_m\} = \begin{cases} \sigma_q^2, & n = m \\ 0, & \text{otherwise} \end{cases} \quad (3.32)$$

Note that because of **independence**, $\mathbb{E}\{q_n q_m\} = \mathbb{E}\{q_n\}\mathbb{E}\{q_m\}$ for $m \neq n$, and $\mathbb{E}\{q_n^2\}$ is the variance or power.

Note also that (3.31) and (3.32) do not specify the distribution. To generate Gaussian white noise, we can use `randn` while `rand-0.5` for uniform white noise.

Example 3.17

Use the MATLAB `randn` to evaluate the whiteness property of a Gaussian distributed sequence with length 2.

We generate 10000 sets of $\{q_1, q_2\}$ with $\sigma_q^2 = 100$, and use:

$$\frac{1}{10000} \sum_{i=1}^{10000} q_{1,i}^2, \quad \frac{1}{10000} \sum_{i=1}^{10000} q_{2,i}^2, \quad \frac{1}{10000} \sum_{i=1}^{10000} q_{1,i}q_{2,i}$$

where $\{q_{1,i}, q_{2,i}\}$ denotes the i th realization sequence, to estimate $\mathbb{E}\{q_1^2\}$, $\mathbb{E}\{q_2^2\}$ and $\mathbb{E}\{q_1q_2\}$.

```
>> Q=10*randn(2,10000); %each column is a sequence
>> mean(Q(1,:).*Q(1,:))
ans = 100.1395
>> mean(Q(2,:).*Q(2,:))
ans = 97.0205
```

```
>> mean(Q(1,:) .* Q(2,:))  
ans = -0.8513
```

If we assign the column vector $\mathbf{q} = [q_1 \ q_2]^T$, these expected values are elements in the **covariance matrix** of \mathbf{q} , denoted by \mathbf{C}_q :

$$\begin{aligned}\mathbf{C}_q &= \mathbb{E}\{(\mathbf{q} - \mathbb{E}\{\mathbf{q}\})(\mathbf{q} - \mathbb{E}\{\mathbf{q}\})^T\} \\ &= \mathbb{E}\{\mathbf{q}\mathbf{q}^T\} \\ &= \begin{bmatrix} \mathbb{E}\{q_1^2\} & \mathbb{E}\{q_1q_2\} \\ \mathbb{E}\{q_1q_2\} & \mathbb{E}\{q_2^2\} \end{bmatrix} \\ &= \begin{bmatrix} \sigma_q^2 & 0 \\ 0 & \sigma_q^2 \end{bmatrix} \\ &= \sigma_q^2 \mathbf{I}_2, \quad \mathbf{I}_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\end{aligned}$$

which is a **diagonal** matrix. Note that \mathbf{I} is the **identity** matrix.

Example 3.18

Consider the problem of estimating a constant A from N observations:

$$r_n = A + w_n, \quad n = 0, 1, \dots, N - 1$$

where w_n is a white noise with mean 0 and variance σ_w^2 . Two estimators are suggested to estimate A :

$$\hat{A}_1 = r_0$$

$$\hat{A}_2 = \frac{1}{N} \sum_{n=0}^{N-1} r_n$$

Compute the mean, variance and mean square error for each of the two estimators.

Recall the whiteness property, we have $\mathbb{E}\{w_n\} = 0$, $\mathbb{E}\{w_n^2\} = \sigma_w^2$ and $\mathbb{E}\{w_n w_m\} = 0$ when $n \neq m$.

For \hat{A}_1 , we obtain:

$$\mathbb{E}\{\hat{A}_1\} = \mathbb{E}\{r_0\} = \mathbb{E}\{A + w_0\} = A + \mathbb{E}\{w_0\} = A + 0 = A$$

$$\text{MSE}(\hat{A}_1) = \mathbb{E}\{(r_0 - A)^2\} = \mathbb{E}\{(A + w_0 - A)^2\} = \mathbb{E}\{w_0^2\} = \sigma_w^2$$

As \hat{A}_1 is unbiased, $\text{var}(\hat{A}_1) = \text{MSE}(\hat{A}_1) = \sigma_w^2$.

Similarly, for \hat{A}_2 :

$$\begin{aligned}\mathbb{E}\{\hat{A}_2\} &= \mathbb{E}\left\{\frac{1}{N} \sum_{n=0}^{N-1} r_n\right\} = \frac{1}{N} \sum_{n=0}^{N-1} \mathbb{E}\{r_n\} = \frac{1}{N} \sum_{n=0}^{N-1} \mathbb{E}\{A + w_n\} \\ &= \frac{1}{N} \sum_{n=0}^{N-1} (A + \mathbb{E}\{w_n\}) = \frac{1}{N} \sum_{n=0}^{N-1} (A + 0) = A\end{aligned}$$

$$\begin{aligned}
\text{MSE}(\hat{A}_2) &= \text{var}(\hat{A}_2) = \mathbb{E} \left\{ \left(\frac{1}{N} \sum_{n=0}^{N-1} r_n - A \right)^2 \right\} \\
&= \mathbb{E} \left\{ \left(\frac{1}{N} \sum_{n=0}^{N-1} w_n + A - A \right)^2 \right\} = \frac{1}{N^2} \mathbb{E} \left\{ \left(\sum_{n=0}^{N-1} w_n \right)^2 \right\} \\
&= \frac{1}{N^2} \sum_{n=0}^{N-1} \sum_{m=0}^{N-1} \mathbb{E} \{ w_n w_m \} = \frac{1}{N^2} \sum_{n=0}^{N-1} \mathbb{E} \{ w_n^2 \} = \frac{\sigma_w^2}{N}
\end{aligned}$$

Which estimator is better? \hat{A}_1 or \hat{A}_2 ? Why?

Bivariate Gaussian Random Variables

Consider 2 **Gaussian** RVs $X \sim \mathcal{N}(\mu_x, \sigma_x^2)$ and $Y \sim \mathcal{N}(\mu_y, \sigma_y^2)$:

$$p(x) = \frac{1}{\sqrt{2\pi\sigma_x^2}} e^{-\frac{1}{2\sigma_x^2}(x-\mu_x)^2} \quad \text{and} \quad p(y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} e^{-\frac{1}{2\sigma_y^2}(y-\mu_y)^2}$$

The **joint PDF**, known as **bivariate Gaussian PDF**, is:

$$p(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \cdot \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[\frac{(x-\mu_x)^2}{\sigma_x^2} - \frac{2\rho(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} + \frac{(y-\mu_y)^2}{\sigma_y^2} \right] \right\} \quad (3.33)$$

where ρ is the correlation coefficient between X and Y :

$$\rho = \frac{\mathbb{E}\{(X - \mu_x)(Y - \mu_y)\}}{\sigma_x\sigma_y}$$

It is clear that **bivariate Gaussian** PDF is characterized by 5 parameters, namely, μ_x , μ_y , σ_x^2 , σ_y^2 , and ρ .

Viewing X and Y as a **vector** $\mathbf{X} = [X \ Y]^T$, we can write $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{C})$ where $\boldsymbol{\mu}$ is the **mean** vector and \mathbf{C} is **covariance** matrix:

$$\boldsymbol{\mu} = [\mu_x \ \mu_y]^T$$

$$\begin{aligned} \mathbf{C} &= \begin{bmatrix} \sigma_x^2 & \mathbb{E}\{(X - \mu_x)(Y - \mu_y)\} \\ \mathbb{E}\{(X - \mu_x)(Y - \mu_y)\} & \sigma_y^2 \end{bmatrix} \\ &= \begin{bmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix} \end{aligned}$$

Note that the bivariate Gaussian PDF can also be characterized by $\boldsymbol{\mu}$ and \mathbf{C} .

Example 3.19

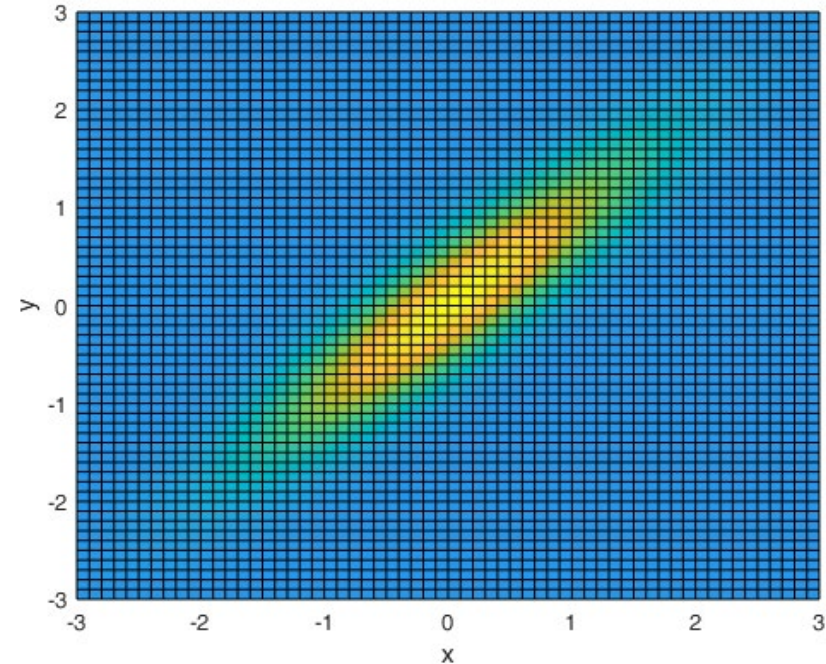
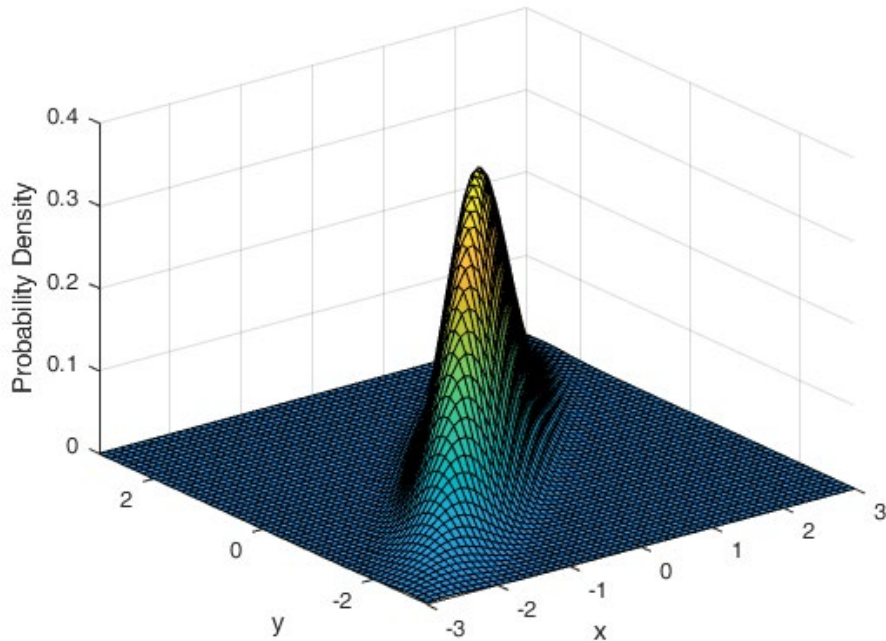
Use MATLAB to visualize a bivariate Gaussian PDF where $X \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{C})$ with

$$\boldsymbol{\mu} = [0 \ 0]^T \quad \text{and} \quad \mathbf{C} = \begin{bmatrix} 1 & 0.9 \\ 0.9 & 1 \end{bmatrix}$$

It is clear that $\mu_x = \mu_y = 0$, $\sigma_x^2 = \sigma_y^2 = 1$ and $\rho = 0.9$. We can use the following MATLAB code:

```
>> mu = [0 0];  
C = [1 0.9; 0.9 1]  
C = 1.0000    0.9000  
    0.9000    1.0000  
>> x = -3:0.1:3;  
y = -3:0.1:3;  
[X,Y] = meshgrid(x,y);  
Xvec = [X(:) Y(:)];  
z = mvnpdf(Xvec,mu,C);  
z = reshape(z,length(y),length(x));
```

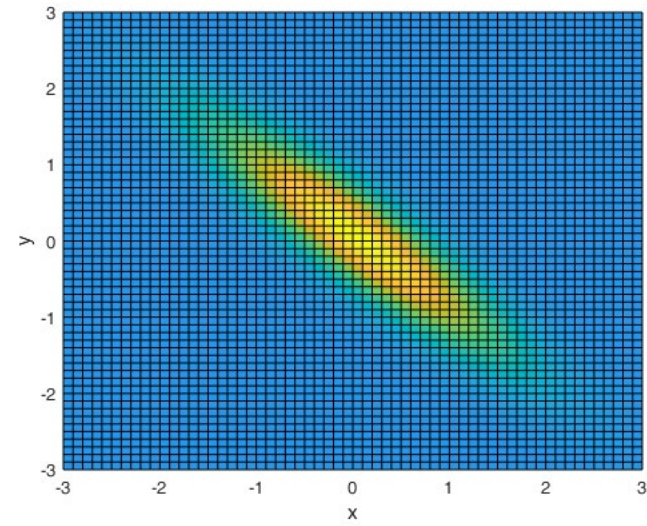
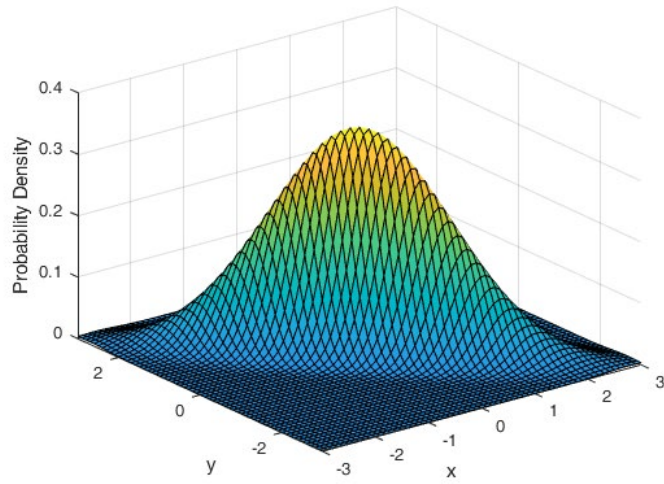
```
surf(x, y, z)
caxis([min(z(:))-0.5*range(z(:)), max(z(:))])
axis([-3 3 -3 3 0 0.4])
xlabel('x')
ylabel('y')
zlabel('Probability Density')
```



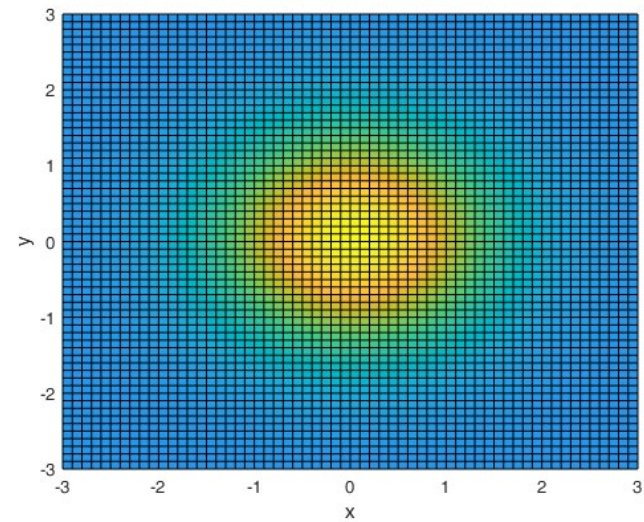
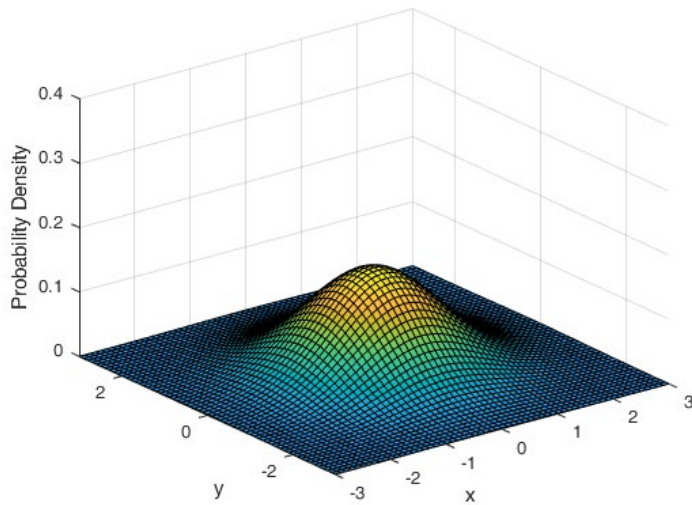
2D points appears with high probability at yellow region.

When x is positive, can you predict the sign of y ?

$\rho = -0.9$:

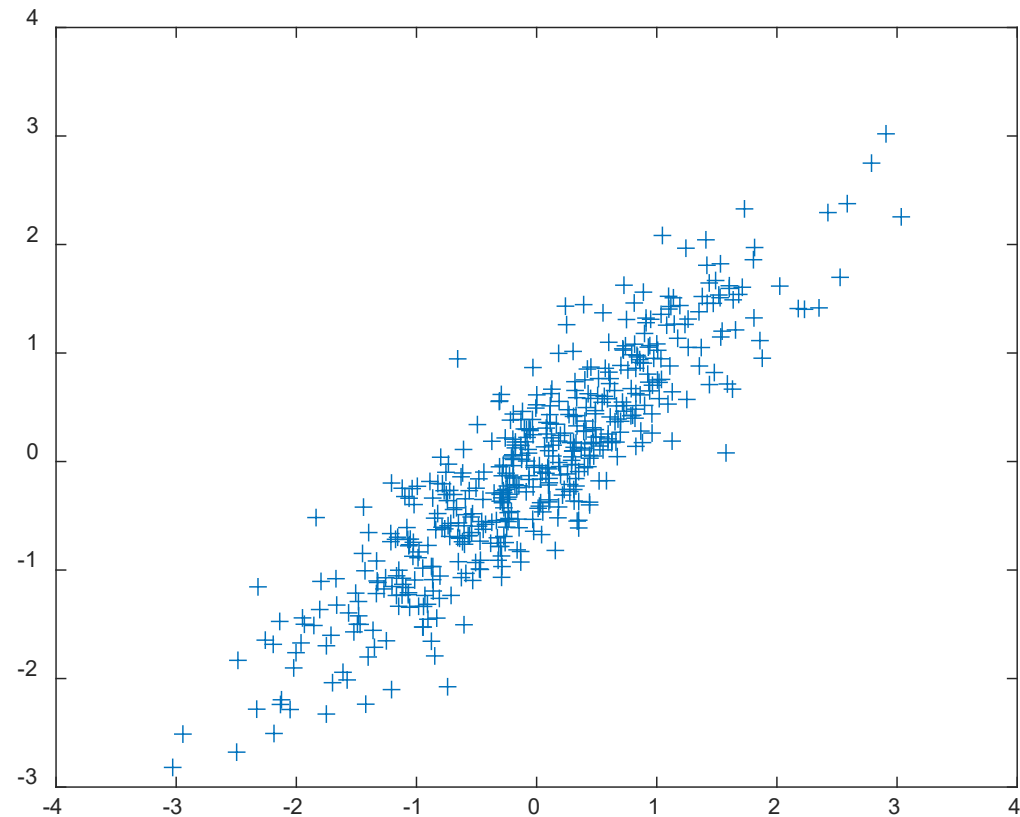


$\rho = 0$:



To generate bivariate Gaussian RVs, we can use `mvnrnd`.

```
>>mvnrnd(mu,C)
    0.5377    1.2833
>> R = mvnrnd(mu,C,500);
plot(R(:,1),R(:,2),'+')
```



If X and Y are **independent**, $\mathbb{E}\{XY\} = \mathbb{E}\{X\}\mathbb{E}\{Y\}$, we have:

$$\mathbb{E}\{(X - \mu_x)(Y - \mu_y)\} = \mathbb{E}\{(X - \mu_x)\}\mathbb{E}\{(Y - \mu_y)\} = 0 \Rightarrow \rho = 0$$

Then (3.33) is simplified to:

$$\begin{aligned} p(x, y) &= \frac{1}{2\pi\sigma_x\sigma_y} \cdot \exp \left\{ -\frac{1}{2} \left[\frac{(x - \mu_x)^2}{\sigma_x^2} + \frac{(y - \mu_y)^2}{\sigma_y^2} \right] \right\} \\ &= \frac{1}{\sqrt{2\pi\sigma_x^2}} e^{-\frac{1}{2\sigma_x^2}(x-\mu_x)^2} \cdot \frac{1}{\sqrt{2\pi\sigma_y^2}} e^{-\frac{1}{2\sigma_y^2}(y-\mu_y)^2} = p(x)p(y) \end{aligned}$$

which aligns with (3.19).

Bivariate Gaussian RVs X and Y are **uncorrelated** if and only if they are **independent**.

For **linear combinations** of bivariate Gaussian RVs:

$$W_1 = a_1X + b_1Y, \quad W_2 = a_2X + b_2Y$$

W_1 and W_2 are also **bivariate Gaussian** RVs with properties:

$$\mathbb{E}\{W_i\} = a_i\mu_x + b_i\mu_y, \quad i = 1, 2 \quad (3.34)$$

$$\text{var}(W_i) = a_i^2\sigma_x^2 + b_i^2\sigma_y^2 + 2a_ib_i\rho\sigma_x\sigma_y, \quad i = 1, 2 \quad (3.35)$$

$$\text{cov}(W_1, W_2) = a_1a_2\sigma_x^2 + b_1b_2\sigma_y^2 + (a_1b_1 + a_2b_2)\rho\sigma_x\sigma_y \quad (3.36)$$

Example 3.20

Find the PDF of $Y = S + N$ where $S \sim \mathcal{N}(0, \sigma_s^2)$ and $N \sim \mathcal{N}(0, \sigma_n^2)$ are independent. Then find the joint PDF of S and Y when $\sigma_s = 4$ and $\sigma_n = 3$.

Noting that $Y = S + N$, applying (3.34) and (3.35) with $a_i = b_i = 1$ and $\rho = 0$, we obtain:

$$\mathbb{E}\{Y\} = \mathbb{E}\{S\} + \mathbb{E}\{N\} = 0$$

$$\sigma_y^2 = \text{var}(Y) = \sigma_s^2 + \sigma_n^2$$

Hence the PDF of Y is:

$$p(y) = \frac{1}{\sqrt{2\pi(\sigma_s^2 + \sigma_n^2)}} e^{-\frac{1}{2(\sigma_s^2 + \sigma_n^2)}y^2}$$

When $\sigma_s = 4$ and $\sigma_n = 3$, we have:

$$\sigma_y^2 = \text{var}(Y) = 25$$

Using the result in Example 3.15

$$\rho_{S,Y} = \frac{\sigma_s^2}{\sqrt{\sigma_s^2 \cdot (\sigma_s^2 + \sigma_n^2)}} = \frac{16}{\sqrt{16 \cdot 25}} = \frac{4}{5}$$

Applying (3.33) yields:

$$p(s, y) = \frac{1}{24\pi} e^{-\frac{1}{18} \left(\frac{25s^2}{16} - 2sy + y^2 \right)}$$

For multiple Gaussian RVs, X_1, \dots, X_N , the general form of joint PDF for $\mathbf{X} = [X_1 \cdots X_N]^T \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{C})$ is:

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{N/2} |\mathbf{C}|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \mathbf{C}^{-1}(\mathbf{x}-\boldsymbol{\mu})} \quad (3.37)$$

where

$$\boldsymbol{\mu} = [\mu_1 \ \mu_2 \ \cdots \ \mu_N]^T$$

$$\begin{aligned} \mathbf{C} &= \mathbb{E}\{(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^T\} \\ &= \begin{bmatrix} \mathbb{E}\{(X_1 - \mu_1)^2\} & \mathbb{E}\{(X_2 - \mu_2)(X_1 - \mu_1)\} & \cdots & \cdots \\ \mathbb{E}\{(X_1 - \mu_1)(X_2 - \mu_2)\} & \mathbb{E}\{(X_2 - \mu_2)^2\} & \cdots & \cdots \\ \vdots & \ddots & \ddots & \vdots \\ \mathbb{E}\{(X_1 - \mu_1)(X_N - \mu_N)\} & \cdots & \cdots & \mathbb{E}\{(X_N - \mu_N)^2\} \end{bmatrix} \end{aligned}$$

Note that $\mathbf{C} \in \mathbb{R}^{N \times N}$ and $|\mathbf{C}|$ is the determinant of \mathbf{C} .

If all X_1, \dots, X_N are **independent** with **same variance** σ^2 , then:

$$\mathbf{C} = \sigma^2 \mathbf{I}_N \Rightarrow |\mathbf{C}| = \sigma^{2N} \quad \text{and} \quad \mathbf{C}^{-1} = \sigma^{-2} \mathbf{I}_N$$

where \mathbf{I}_N is the $N \times N$ identity matrix. Then we have:

$$\begin{aligned} p(\mathbf{x}) &= \frac{1}{(2\pi)^{N/2} \sigma^N} e^{-\frac{1}{2\sigma^2} (\mathbf{x} - \boldsymbol{\mu})^T (\mathbf{x} - \boldsymbol{\mu})} = \frac{1}{(2\pi)^{N/2} \sigma^N} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^N (x_i - \mu_i)^2} \\ &= p(x_1) \cdots p(x_N) = \prod_{i=1}^N p(x_i) \end{aligned} \quad (3.38)$$

which aligns and generalizes (3.19).

On the other hand, for the bivariate case of $\mathbf{X} = [X \ Y]^T$, it is easily verified that:

$$\begin{aligned} \mathbf{C} &= \begin{bmatrix} \sigma_x^2 & \mathbb{E}\{(X - \mu_x)(Y - \mu_y)\} \\ \mathbb{E}\{(X - \mu_x)(Y - \mu_y)\} & \sigma_y^2 \end{bmatrix} \\ &= \begin{bmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix} \end{aligned}$$

and

$$|\mathbf{C}| = \sigma_x^2\sigma_y^2(1 - \rho^2)$$

$$\mathbf{C}^{-1} = \frac{1}{\sigma_x^2\sigma_y^2(1 - \rho^2)} \begin{bmatrix} \sigma_y^2 & -\rho\sigma_x\sigma_y \\ -\rho\sigma_x\sigma_y & \sigma_x^2 \end{bmatrix}$$

Using these results, we get the joint PDF $p(x, y)$ in (3.33). That is, (3.37) generalizes (3.33) for any value of N .

In fact, the joint CDF, PMF, PDF and expected values can be readily extended from 2 RVs to N RVs.

For N **independent** RVs X_1, \dots, X_N , (3.18)-(3.20) can be generalized as:

$$\begin{aligned} F(x_1, \dots, x_N) &= P(X_1 \leq x_1, \dots, X_N \leq x_N) \\ &= P(X_1 \leq x_1) \cdots P(X_N \leq x_N) = F(x_1) \cdots F(x_N) \end{aligned}$$

$$p(x_1, \dots, x_N) = p(x_1) \cdots p(x_N)$$

$$\mathbb{E}\{g_1(X_1) \cdots g_N(X_N)\} = \mathbb{E}\{g_1(X_1)\} \cdots \mathbb{E}\{g_N(X_N)\}$$

where $\{g_n\}$ are any functions.

References:

1. R. D. Yates and D. J. Goodman, *Probability and Stochastic Processes: A Friendly Introduction for Electrical and Computer Engineer*, Wiley, 2014
2. S. M. Ross, *Introduction to Probability Models*, Academic Press, 2014
3. D. G. Childers, *Probability and Random Processes: Using MATLAB with Applications to Continuous and Discrete Time Systems*, Irwin, 1997
4. S. Kay, *Intuitive Probability and Random Processes using MATLAB*, Springer, 2006
5. [Multivariate Normal Distribution - MATLAB & Simulink \(mathworks.com\)](https://www.mathworks.com/help/matlab/matlab_prog/multivariate-normal-distribution.html)