6. Vector Random Variables

In the previous chapter we presented methods for dealing with two random variables. In this chapter we extend these methods to the case of n random variables in the following ways:

- By representing n random variables as a vector, we obtain a compact notation for the joint pmf, cdf, and pdf as well as marginal and conditional distributions.
- We present a general method for finding the pdf of transformations of vector random variables.
- Summary information of the distribution of a vector random variable is provided by an expected value vector and a covariance matrix.
- We use linear transformations and characteristic functions to find alternative representations of random vectors and their probabilities.
- We develop optimum estimators for estimating the value of a random variable based on observations of other random variables.
- We show how jointly Gaussian random vectors have a compact and easy-to-workwith pdf and characteristic function.

6.1 Vector Random Variables

The notion of a random variable is easily generalized to the case where several quantities are of interest. A vector random variable X is a function that assigns a vector of real numbers to each outcome ζ in S, the sample space of the random experiment. We use uppercase boldface notation for vector random variables. By convention X is a column vector (n rows by 1 column), so the vector random variable with components $X_1, X_2, \ldots, X_n$ corresponds to

$$\vec{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix} = \begin{bmatrix} X_1 & X_2 & \cdots & X_n \end{bmatrix}^T$$

We will sometimes write $X = (X_1, X_2, \ldots, X_n)$ to save space and omit the transpose unless dealing with matrices. Possible values of the vector random variable are denoted by $x = (x_1, x_2, \ldots, x_n)$ where x_i corresponds to the value of X_i.

Example 6.1 Arrivals at a Packet Switch

Packets arrive at each of three input ports of a packet switch according to independent Bernoulli trials with $p = 1/2$. Each arriving packet is equally likely to be destined to any of three output ports. Let $\mathbf{X} = (X_1, X_2, X_3)$ where X_i is the total number of packets arriving for output port i. $\mathbf{X}$ is a vector random variable whose values are determined by the pattern of arrivals at the input ports.

Example 6.3 Samples of an Audio Signal

Let the outcome ζ of a random experiment be an audio signal $X(t)$. Let the random variable $X_k = X(kT)$ be the sample of the signal taken at time kT. An MP3 codec processes the audio in blocks of n samples $\mathbf{X} = (X_1, X_2, \ldots, X_n)$. $\mathbf{X}$ is a vector random variable.

Notes and figures are based on or taken from materials in the textbook: Alberto Leon-Garcia, "Probability, Statistics, and Random Processes For Electrical Engineering, 3rd ed.", Pearson Prentice Hall, 2008, ISBN: 013-147122-8.

6.1.1 Events and Probabilities

Each event A involving $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ has a corresponding region in an n-dimensional real space R^n. As before, we use "rectangular" product-form sets in R^n as building blocks. For the n-dimensional random variable $\mathbf{X} = (X_1, X_2, \ldots, X_n)$, we are interested in events that have the **product form**

$$A = \{X_1 \text{ in } A_1\} \cap \{X_2 \text{ in } A_2\} \cap \cdots \cap \{X_n \text{ in } A_n\}, \tag{6.1}$$

where each A_k is a one-dimensional event (i.e., subset of the real line) that involves X_k only. The event A occurs when all of the events $\{X_k \text{ in } A_k\}$ occur jointly.

We are interested in obtaining the probabilities of these product-form events:

$$\begin{aligned} P[A] &= P[\mathbf{X} \in A] = P[\{X_1 \text{ in } A_1\} \cap \{X_2 \text{ in } A_2\} \cap \cdots \cap \{X_n \text{ in } A_n\}] \\ &\triangleq P[X_1 \text{ in } A_1, X_2 \text{ in } A_2, \ldots, X_n \text{ in } A_n]. \end{aligned} \tag{6.2}$$

In principle, the probability in Eq. (6.2) is obtained by finding the probability of the equivalent event in the underlying sample space, that is,

$$\begin{aligned} P[A] &= P[\{\zeta \text{ in } S : \mathbf{X}(\zeta) \text{ in } A\}] \\ &= P[\{\zeta \text{ in } S : X_1(\zeta) \in A_1, X_2(\zeta) \in A_2, \ldots, X_n(\zeta) \in A_n\}]. \end{aligned} \tag{6.3}$$

6.1.2 Joint Distribution Functions

The **joint cumulative distribution function** of $X_1, X_2, \ldots, X_n$ is defined as the probability of an n-dimensional semi-infinite rectangle associated with the point $(x_1, \ldots, x_n)$:

$$F_{\mathbf{X}}(\mathbf{x}) \triangleq F_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n) = P[X_1 \le x_1, X_2 \le x_2, \ldots, X_n \le x_n]. \tag{6.4}$$

The joint cdf is defined for discrete, continuous, and random variables of mixed type. The probability of product-form events can be expressed in terms of the joint cdf.

The joint cdf generates a family of **marginal cdf's** for subcollections of the random variables $X_1, \ldots, X_n$. These marginal cdf's are obtained by setting the appropriate entries to $+\infty$ in the joint cdf in Eq. (6.4). For example:

Joint cdf for $X_1, \ldots, X_{n-1}$ is given by $F_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_{n-1}, \infty)$ and
Joint cdf for X_1 and X_2 is given by $F_{X_1, X_2, \ldots, X_n}(x_1, x_2, \infty, \ldots, \infty)$.

The **joint probability mass function** of n discrete random variables is defined by

$$p_{\mathbf{X}}(\mathbf{x}) \triangleq p_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n) = P[X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n]. \tag{6.5}$$

The probability of any n-dimensional event A is found by summing the pmf over the points in the event

$$P[\mathbf{X} \text{ in } A] = \sum_{\mathbf{x} \text{ in } A} \cdots \sum p_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n). \tag{6.6}$$

A family of **conditional pmf's** is obtained from the joint pmf by conditioning on different subcollections of the random variables. For example, if $p_{X_1,\ldots,X_{n-1}}(x_1,\ldots,x_{n-1}) > 0$:

$$p_{X_n}(x_n \mid x_1,\ldots,x_{n-1}) = \frac{p_{X_1,\ldots,X_n}(x_1,\ldots,x_n)}{p_{X_1,\ldots,X_{n-1}}(x_1,\ldots,x_{n-1})}. \tag{6.9a}$$

Repeated applications of Eq. (6.9a) yield the following very useful expression:

$$p_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = p_{X_n}(x_n|x_1,\ldots,x_{n-1})p_{X_{n-1}}(x_{n-1}|x_1,\ldots,x_{n-2})\ldots p_{X_2}(x_2|x_1)p_{X_1}(x_1). \tag{6.9b}$$

Example 6.5 Arrivals at a Packet Switch

Find the joint pmf of $\mathbf{X} = (X_1, X_2, X_3)$ in Example 6.1. Find $P[X_1 > X_3]$.

Let N be the total number of packets arriving in the three input ports. Each input port has an arrival with probability $p = 1/2$, so N is binomial with pmf:

$$p_N(n) = \binom{3}{n}\frac{1}{2^3} \quad \text{for} \quad 0 \le n \le 3.$$

Given $N = n$, the number of packets arriving for each output port has a multinomial distribution:

$$p_{X_1,X_2,X_3}(i,j,k \mid i+j+k=n) = \begin{cases} \dfrac{n!}{i!\,j!\,k!}\dfrac{1}{3^n} & \text{for} \quad i+j+k=n, i \ge 0, j \ge 0, k \ge 0 \\ 0 & \text{otherwise.} \end{cases}$$

The joint pmf of $\mathbf{X}$ is then:

$$p_{\mathbf{X}}(i,j,k) = p_{\mathbf{X}}(i,j,k \mid n)\binom{3}{n}\frac{1}{2^3} \quad \text{for} \quad i \ge 0, j \ge 0, k \ge 0, i+j+k = n \le 3.$$

The explicit values of the joint pmf are:

$$p_{\mathbf{X}}(0,0,0) = \frac{0!}{0!\,0!\,0!}\frac{1}{3^0}\binom{3}{0}\frac{1}{2^3} = \frac{1}{8}$$

$$p_{\mathbf{X}}(1,0,0) = p_{\mathbf{X}}(0,1,0) = p_{\mathbf{X}}(0,0,1) = \frac{1!}{0!\,0!\,1!}\frac{1}{3^1}\binom{3}{1}\frac{1}{2^3} = \frac{3}{24}$$

$$p_{\mathbf{X}}(1,1,0) = p_{\mathbf{X}}(1,0,1) = p_{\mathbf{X}}(0,1,1) = \frac{2!}{0!\,1!\,1!}\frac{1}{3^2}\binom{3}{2}\frac{1}{2^3} = \frac{6}{72}$$

$$p_{\mathbf{X}}(2,0,0) = p_{\mathbf{X}}(0,2,0) = p_{\mathbf{X}}(0,0,2) = 3/72$$

$$p_{\mathbf{X}}(1,1,1) = 6/216$$

$$p_{\mathbf{X}}(0,1,2) = p_{\mathbf{X}}(0,2,1) = p_{\mathbf{X}}(1,0,2) = p_{\mathbf{X}}(1,2,0) = p_{\mathbf{X}}(2,0,1) = p_{\mathbf{X}}(2,1,0) = 3/216$$

$$p_{\mathbf{X}}(3,0,0) = p_{\mathbf{X}}(0,3,0) = p_{\mathbf{X}}(0,0,3) = 1/216.$$

Finally:

$$\begin{aligned} P[X_1 > X_3] &= p_{\mathbf{X}}(1,0,0) + p_{\mathbf{X}}(1,1,0) + p_{\mathbf{X}}(2,0,0) + p_{\mathbf{X}}(1,2,0) \\ &\quad + p_{\mathbf{X}}(2,0,1) + p_{\mathbf{X}}(2,1,0) + p_{\mathbf{X}}(3,0,0) \\ &= 8/27. \end{aligned}$$

We say that the random variables $X_1, X_2, \ldots, X_n$ are **jointly continuous random variables** if the probability of any n-dimensional event A is given by an n-dimensional integral of a probability density function:

$$P[\mathbf{X} \text{ in } A] = \int_{\mathbf{x} \text{ in } A} \cdots \int f_{X_1,\ldots,X_n}(x'_1, \ldots, x'_n)\, dx'_1 \ldots dx'_n, \tag{6.10}$$

where $f_{X_1,\ldots,X_n}(x_1, \ldots, x_n)$ is the **joint probability density function**.

The joint cdf of $\mathbf{X}$ is obtained from the joint pdf by integration:

$$F_{\mathbf{X}}(\mathbf{x}) = F_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f_{X_1,\ldots,X_n}(x'_1, \ldots, x'_n)\, dx'_1 \ldots dx'_n. \tag{6.11}$$

The joint pdf (if the derivative exists) is given by

$$f_{\mathbf{X}}(\mathbf{x}) \triangleq f_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n) = \frac{\partial^n}{\partial x_1 \ldots \partial x_n} F_{X_1,\ldots,X_n}(x_1, \ldots, x_n). \tag{6.12}$$

A family of **conditional pdf's** is also associated with the joint pdf. For example, the pdf of X_n given the values of $X_1, \ldots, X_{n-1}$ is given by

$$f_{X_n}(x_n \mid x_1, \ldots, x_{n-1}) = \frac{f_{X_1,\ldots,X_n}(x_1, \ldots, x_n)}{f_{X_1,\ldots,X_{n-1}}(x_1, \ldots, x_{n-1})} \tag{6.15a}$$

if $f_{X_1,\ldots,X_{n-1}}(x_1, \ldots, x_{n-1}) > 0$.

Repeated applications of Eq. (6.15a) yield an expression analogous to Eq. (6.9b):

$$f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = f_{X_n}(x_n \mid x_1, \ldots, x_{n-1}) f_{X_{n-1}}(x_{n-1} \mid x_1, \ldots, x_{n-2}) \ldots f_{X_2}(x_2 \mid x_1) f_{X_1}(x_1). \tag{6.15b}$$

Example 6.6

The random variables X_1, X_2, and X_3 have the joint Gaussian pdf

$$f_{X_1,X_2,X_3}(x_1, x_2, x_3) = \frac{e^{-(x_1^2 + x_2^2 - \sqrt{2}\, x_1 x_2 + \frac{1}{2} x_3^2)}}{2\pi\sqrt{\pi}}.$$

Find the marginal pdf of X_1 and X_3. Find the conditional pdf of X_2 given X_1 and X_3.

The marginal pdf for the pair X_1 and X_3 is found by integrating the joint pdf over x_2:

$$f_{X_1,X_3}(x_1, x_3) = \frac{e^{-x_3^2/2}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{e^{-(x_1^2 + x_2^2 - \sqrt{2} x_1 x_2)}}{2\pi/\sqrt{2}}\, dx_2.$$

The above integral was carried out in Example 5.18 with $\rho = -1/\sqrt{2}$. By substituting the result of the integration above, we obtain

$$f_{X_1,X_3}(x_1, x_3) = \frac{e^{-x_3^2/2}}{\sqrt{2\pi}} \frac{e^{-x_1^2/2}}{\sqrt{2\pi}}.$$

Therefore X_1 and X_3 are independent zero-mean, unit-variance Gaussian random variables.

The conditional pdf of X_2 given X_1 and X_3 is:

$$f_{X_2}(x_2 \mid x_1, x_3) = \frac{e^{-(x_1^2 + x_2^2 - \sqrt{2}x_1x_2 + {}^1\!/_2 x_3^2)}}{2\pi\sqrt{\pi}} \frac{\sqrt{2\pi}\sqrt{2\pi}}{e^{-x_3^2/2}e^{-x_1^2/2}}$$

$$= \frac{e^{-({}^1\!/_2 x_1^2 + x_2^2 - \sqrt{2}x_1x_2)}}{\sqrt{\pi}} = \frac{e^{-(x_2 - x_1/\sqrt{2}x_1)^2}}{\sqrt{\pi}}.$$

We conclude that X_2 given and X_3 is a Gaussian random variable with mean $x_1/\sqrt{2}$ and variance 1/2.

6.1.3 Independence

The collection of random variables $X_1, \ldots, X_n$ is **independent** if

$$P[X_1 \text{ in } A_1, X_2 \text{ in } A_2, \ldots, X_n \text{ in } A_n] = P[X_1 \text{ in } A_1]P[X_2 \text{ in } A_2] \ldots P[X_n \text{ in } A_n]$$

for *any* one-dimensional events $A_1, \ldots, A_n$. It can be shown that $X_1, \ldots, X_n$ *are independent if and only if*

$$F_{X_1, \ldots, X_n}(x_1, \ldots, x_n) = F_{X_1}(x_1) \ldots F_{X_n}(x_n) \tag{6.16}$$

for all $x_1, \ldots, x_n$. If the random variables are discrete, Eq. (6.16) is equivalent to

$$p_{X_1, \ldots, X_n}(x_1, \ldots, x_n) = p_{X_1}(x_1) \ldots p_{X_n}(x_n) \qquad \text{for all } x_1, \ldots, x_n.$$

If the random variables are jointly continuous, Eq. (6.16) is equivalent to

$$f_{X_1, \ldots, X_n}(x_1, \ldots, x_n) = f_{X_1}(x_1) \ldots f_{X_n}(x_n)$$

for all $x_1, \ldots, x_n$.

Example 6.8

The n samples $X_1, X_2, \ldots, X_n$ of a noise signal have joint pdf given by

$$f_{X_1, \ldots, X_n}(x_1, \ldots, x_n) = \frac{e^{-(x_1^2 + \ldots + x_n^2)/2}}{(2\pi)^{n/2}} \qquad \text{for all } x_1, \ldots, x_n.$$

It is clear that the above is the product of n one-dimensional Gaussian pdf's. Thus $X_1, \ldots, X_n$ are independent Gaussian random variables.

6.2 Functions of Several Random Variables

Functions of vector random variables arise naturally in random experiments. For example $X = (X_1, X_2, \ldots, X_n)$ may correspond to observations from n repetitions of an experiment that generates a given random variable. We are almost always interested in the sample mean and the sample variance of the observations. In another example $X = (X_1, X_2, \ldots, X_n)$ may correspond to samples of a speech waveform and we may be interested in extracting features that are defined as functions of X for use in a speech recognition system.

6.2.1 One Function of Several Random Variables

Example 6.9 Maximum and Minimum of *n* Random Variables

Let $W = \max(X_1, X_2, \ldots, X_n)$ and $Z = \min(X_1, X_2, \ldots, X_n)$, where the X_i are independent random variables with the same distribution. Find $F_W(w)$ and $F_Z(z)$.

The maximum of $X_1, X_2, \ldots, X_n$ is less than x if and only if each X_i is less than x, so:

$$\begin{aligned} F_W(w) &= P[\max(X_1, X_2, \ldots, X_n) \le w] \\ &= P[X_1 \le w]P[X_2 \le w] \ldots P[X_n \le w] = (F_X(w))^n. \end{aligned}$$

The minimum of $X_1, X_2, \ldots, X_n$ is greater than x if and only if each X_i is greater than x, so:

$$\begin{aligned} 1 - F_Z(z) &= P[\min(X_1, X_2, \ldots, X_n) > z] \\ &= P[X_1 > z]P[X_2 > z] \ldots P[X_n > z] = (1 - F_X(z))^n \end{aligned}$$

and

$$F_Z(z) = 1 - (1 - F_X(z))^n.$$

6.2.2 Transformations of Random Vectors

Example 6.12

Given a random vector $\mathbf{X}$, find the joint pdf of the following transformation:

$$
\begin{aligned}
Z_1 &= g_1(X_1) = a_1X_1 + b_1, \\
Z_2 &= g_2(X_2) = a_2X_2 + b_2, \\
&\vdots \\
Z_n &= g_n(X_n) = a_nX_n + b_n.
\end{aligned}
$$

Note that $Z_k = a_kX_k + b_k, \leq z_k$, if and only if $X_k \leq (z_k - b_k)/a_k$, if $a_k > 0$, so

$$
\begin{aligned}
F_{Z_1,Z_2,\ldots,Z_n}(z_1, z_2, \ldots, z_n) &= P\left[X_1 \leq \frac{z_1 - b_1}{a_1}, X_2 \leq \frac{z_2 - b_2}{a_2}, \ldots, X_n \leq \frac{z_n - b_n}{a_n}\right] \\
&= F_{X_1,X_2,\ldots,X_n}\left(\frac{z_1 - b_1}{a_1}, \frac{z_2 - b_2}{a_2}, \ldots, \frac{z_n - b_n}{a_n}\right) \\
f_{Z_1,Z_2,\ldots,Z_n}(z_1, z_2, \ldots, z_n) &= \frac{\partial^n}{\partial z_1 \ldots \partial z_n} F_{Z_1,Z_2,\ldots,Z_n}(z_1, z_2, \ldots, z_n) \\
&= \frac{1}{a_1 \ldots a_n} f_{X_1,X_2,\ldots,X_n}\left(\frac{z_1 - b_1}{a_1}, \frac{z_2 - b_2}{a_2}, \ldots, \frac{z_n - b_n}{a_n}\right).
\end{aligned}
$$

6.2.3 *pdf of General Transformations

We now introduce a general method for finding the pdf of a transformation of n jointly continuous random variables. We first develop the two-dimensional case. Let the random variables V and W be defined by two functions of X and Y:

$$V = g_1(X, Y) \qquad \text{and} \qquad W = g_2(X, Y). \tag{6.20}$$

Assume that the functions $v(x, y)$ and $w(x, y)$ are invertible in the sense that the equations $v = g_1(x, y)$ and $w = g_2(x, y)$ can be solved for x and y, that is,

$$x = h_1(v, w) \text{ and } y = h_2(v, w).$$

The joint pdf of X and Y is found by finding the equivalent event of infinitesimal rectangles. The image of the infinitesimal rectangle is shown in Fig. 6.1(a). The image can be approximated by the parallelogram shown in Fig. 6.1(b) by making the approximation

$$g_k(x + dx, y) \simeq g_k(x, y) + \frac{\partial}{\partial x} g_k(x, y)\, dx \qquad k = 1, 2$$

and similarly for the y variable. The probabilities of the infinitesimal rectangle and the parallelogram are approximately equal, therefore

$$f_{X,Y}(x, y)\, dx\, dy = f_{V,W}(v, w)\, dP$$

and

$$f_{V,W}(v, w) = \frac{f_{X,Y}(h_1(v, w), (h_2(v, w))}{\left|\dfrac{dP}{dxdy}\right|}, \tag{6.21}$$

where dP is the area of the parallelogram. By analogy with the case of a linear transformation (see Eq. 5.59), we can match the derivatives in the above approximations with the coefficients in the linear transformations and conclude that the "stretch factor" at the point (v, w) is given by the determinant of a matrix of partial derivatives:

$$J(x, y) = \det \begin{bmatrix} \dfrac{\partial v}{\partial x} & \dfrac{\partial v}{\partial y} \\ \dfrac{\partial w}{\partial x} & \dfrac{\partial w}{\partial y} \end{bmatrix}.$$

The determinant $J(x, y)$ is called the **Jacobian** of the transformation. The Jacobian of the inverse transformation is given by

$$J(v, w) = \det\begin{bmatrix} \dfrac{\partial x}{\partial v} & \dfrac{\partial x}{\partial w} \\ \dfrac{\partial y}{\partial v} & \dfrac{\partial y}{\partial w} \end{bmatrix}.$$

It can be shown that

$$|J(v, w)| = \frac{1}{|J(x, y)|}.$$

We therefore conclude that the joint pdf of V and W can be found using either of the following expressions:

$$f_{V,W}(v, w) = \frac{f_{X,Y}(h_1(v, w), (h_2(v, w))}{|J(x, y)|} \tag{6.22a}$$

$$= f_{X,Y}(h_1(v, w), (h_2(v, w))|J(v, w)|. \tag{6.22b}$$

It should be noted that Eq. (6.21) is applicable even if Eq. (6.20) has more than one solution; the pdf is then equal to the sum of terms of the form given by Eqs. (6.22a) and (6.22b), with each solution providing one such term.

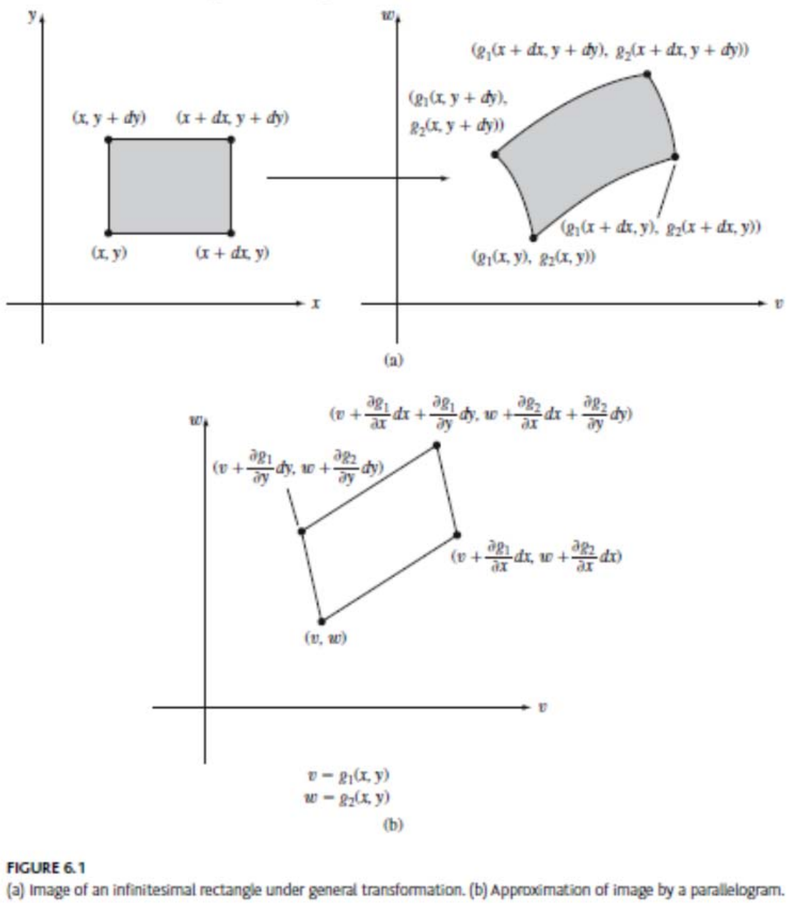


FIGURE 6.1
(a) Image of an infinitesimal rectangle under general transformation. (b) Approximation of image by a parallelogram.

Next consider the problem of finding the joint pdf for n functions of n random variables $\mathbf{X} = (X_1, \ldots, X_n)$:

$$Z_1 = g_1(\mathbf{X}), \quad Z_2 = g_2(\mathbf{X}), \ldots, \quad Z_n = g_n(\mathbf{X}).$$

We assume as before that the set of equations

$$z_1 = g_1(\mathbf{x}), \quad z_2 = g_2(\mathbf{x}), \ldots, \quad z_n = g_n(\mathbf{x}). \tag{6.23}$$

has a unique solution given by

$$x_1 = h_1(\mathbf{x}), \quad x_2 = h_2(\mathbf{x}), \ldots, \quad x_n = h_n(\mathbf{x}).$$

The joint pdf of $\mathbf{Z}$ is then given by

$$f_{Z_1,\ldots,Z_n}(z_1, \ldots, z_n) = \frac{f_{X_1,\ldots,X_n}(h_1(\mathbf{z}), h_2(\mathbf{z}), \ldots, h_n(\mathbf{z}))}{|J(x_1, x_2, \ldots, x_n)|} \tag{6.24a}$$

$$= f_{X_1,\ldots,X_n}(h_1(\mathbf{z}), h_2(\mathbf{z}), \ldots, h_n(\mathbf{z}))|J(z_1, z_2, \ldots, z_n)|, \tag{6.24b}$$

where $|J(x_1, \ldots, x_n)|$ and $|J(z_1, \ldots, z_n)|$ are the determinants of the transformation and the inverse transformation, respectively,

$$J(x_1, \ldots, x_n) = \det\begin{bmatrix} \dfrac{\partial g_1}{\partial x_1} & \cdots & \dfrac{\partial g_1}{\partial x_n} \\ \vdots & & \vdots \\ \dfrac{\partial g_n}{\partial x_1} & \cdots & \dfrac{\partial g_n}{\partial x_n} \end{bmatrix}$$

and

$$J(z_1, \ldots, z_n) = \det\begin{bmatrix} \dfrac{\partial h_1}{\partial z_1} & \cdots & \dfrac{\partial h_1}{\partial z_n} \\ \vdots & & \vdots \\ \dfrac{\partial h_n}{\partial z_1} & \cdots & \dfrac{\partial h_n}{\partial z_n} \end{bmatrix}.$$

In the special case of a linear transformation we have:

$$\mathbf{Z} = \mathbf{A}\mathbf{X} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdot & \cdot & \cdots & \cdot \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}\begin{bmatrix} X_1 \\ X_2 \\ \cdots \\ X_n \end{bmatrix}.$$

The components of $\mathbf{Z}$ are:

$$Z_j = a_{j1}X_1 + a_{j2}X_2 + \dots + a_{jn}X_n.$$

Since $dz_j/dx_i = a_{ji}$, the Jacobian is then simply:

$$J(x_1, x_2, \dots, x_n) = \det\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ . & . & \dots & . \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} = \det \mathbf{A}.$$

Assuming that $\mathbf{A}$ is invertible,[1] we then have that:

$$f_{\mathbf{Z}}(\mathbf{z}) = \frac{f_{\mathbf{X}}(\mathbf{x})}{|\det \mathbf{A}|}\bigg|_{\mathbf{x}=\mathbf{A}^{-1}\mathbf{z}} = \frac{f_{\mathbf{X}}(\mathbf{A}^{-1}\mathbf{z})}{|\det \mathbf{A}|}.$$

6.3 Expected Values of Vector Random Variables

The expected value of a function $g(\mathbf{X}) = g(X_1, \ldots, X_n)$ of a vector random variable $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ is given by:

$$E[Z] = \begin{cases} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1, x_2, \ldots, x_n) f_{\mathbf{X}}(x_1, x_2, \ldots, x_n)\, dx_1\, dx_2 \ldots dx_n & \mathbf{X} \text{ jointly continuous} \\ \sum_{x_1} \cdots \sum_{x_n} g(x_1, x_2, \ldots, x_n) p_{\mathbf{X}}(x_1, x_2, \ldots, x_n) & \mathbf{X} \text{ discrete.} \end{cases} \tag{6.25}$$

An important example is $g(\mathbf{X})$ equal to the sum of functions of $\mathbf{X}$. The procedure leading to Eq. (5.26) and a simple induction argument show that:

$$E[g_1(\mathbf{X}) + g_2(\mathbf{X}) + \cdots + g_n(\mathbf{X})] = E[g_1(\mathbf{X})] + \cdots + E[g_n(\mathbf{X})]. \tag{6.26}$$

Another important example is $g(\mathbf{X})$ equal to the product of n individual functions of the components. If $X_1, \ldots, X_n$ are *independent* random variables, then

$$E[g_1(X_1)g_2(X_2)\ldots g_n(X_n)] = E[g_1(X_1)]E[g_2(X_2)]\ldots E[g_n(X_n)]. \tag{6.27}$$

6.3.1 Mean Vector and Covariance Matrix

For $\mathbf{X} = (X_1, X_2, \ldots, X_n)$, the **mean vector** is defined as the column vector of expected values of the components X_k:

$$\mathbf{m}_{\mathbf{X}} = E[\mathbf{X}] = E\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix} \triangleq \begin{bmatrix} E[X_1] \\ E[X_2] \\ \vdots \\ E[X_n] \end{bmatrix}. \tag{6.28a}$$

Note that we define the vector of expected values as a column vector. In previous sections we have sometimes written $\mathbf{X}$ as a row vector, but in this section and wherever we deal with matrix transformations, we will represent $\mathbf{X}$ and its expected value as a column vector.

The **correlation matrix** has the second moments of $\mathbf{X}$ as its entries:

$$\mathbf{R}_{\mathbf{X}} = \begin{bmatrix} E[X_1^2] & E[X_1X_2] & \ldots & E[X_1X_n] \\ E[X_2X_1] & E[X_2^2] & \ldots & E[X_2X_n] \\ . & . & \ldots & . \\ E[X_nX_1] & E[X_nX_2] & \ldots & E[X_n^2] \end{bmatrix}. \tag{6.28b}$$

The **covariance matrix** has the second-order central moments as its entries:

$$\mathbf{K_X} = \begin{bmatrix} E[(X_1-m_1)^2] & E[(X_1-m_1)(X_2-m_2)] & \dots & E[(X_1-m_1)(X_n-m_n)] \\ E[(X_2-m_2)(X_1-m_1)] & E[(X_2-m_2)^2] & \dots & E[(X_2-m_2)(X_n-m_n)] \\ . & . & \dots & . \\ E[(X_n-m_n)(X_1-m_1)] & E[(X_n-m_n)(X_2-m_2)] & \dots & E[(X_n-m_n)^2] \end{bmatrix}. \tag{6.28c}$$

Both $\mathbf{R_X}$ and $\mathbf{K_X}$ are $n \times n$ symmetric matrices. The diagonal elements of $\mathbf{K_X}$ are given by the variances $\text{VAR}[X_k] = E[(X_k - m_k)^2]$ of the elements of $\mathbf{X}$. If these elements are uncorrelated, then $\text{COV}(X_j, X_k) = 0$ for $j \neq k$, and $\mathbf{K_X}$ is a diagonal matrix. If the random variables $X_1, \dots, X_n$ are independent, then they are uncorrelated and $\mathbf{K_X}$ is diagonal. Finally, if the vector of expected values is $\mathbf{0}$, that is, $m_k = E[X_k] = 0$ for all k, then $\mathbf{R_X} = \mathbf{K_X}$.

Example 6.16

Let $\mathbf{X} = (X_1, X_2, X_3)$ be the jointly Gaussian random vector from Example 6.6. Find $E[\mathbf{X}]$ and $\mathbf{K_X}$.

We rewrite the joint pdf as follows:

$$f_{X_1,X_2,X_3}(x_1, x_2, x_3) = \frac{e^{-(x_1^2 + x_2^2 - 2\frac{1}{\sqrt{2}}x_1x_2)}}{2\pi\sqrt{1 - \left(-\frac{1}{\sqrt{2}}\right)^2}} \frac{e^{-x_3^2/2}}{\sqrt{2\pi}}.$$

We see that X_3 is a Gaussian random variable with zero mean and unit variance, and that it is independent of X_1 and X_2. We also see that X_1 and X_2 are jointly Gaussian with zero mean and unit variance, and with correlation coefficient

$$\rho_{X_1X_2} = -\frac{1}{\sqrt{2}} = \frac{\text{COV}(X_1, X_2)}{\sigma_{X_1}\sigma_{X_2}} = \text{COV}(X_1, X_2).$$

Therefore the vector of expected values is: $\mathbf{m_X} = \mathbf{0}$, and

$$\mathbf{K_X} = \begin{bmatrix} 1 & -\frac{1}{\sqrt{2}} & 0 \\ -\frac{1}{\sqrt{2}} & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

We now develop compact expressions for $\mathbf{R_X}$ and $\mathbf{K_X}$. If we multiply $\mathbf{X}$, an $n \times 1$ matrix, and $\mathbf{X}^{\mathrm{T}}$, a $1 \times n$ matrix, we obtain the following $n \times n$ matrix:

$$\mathbf{XX}^{\mathrm{T}} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix} [X_1, X_2, \ldots, X_n] = \begin{bmatrix} X_1^2 & X_1X_2 & \ldots & X_1X_n \\ X_2X_1 & X_2^2 & \ldots & X_2X_n \\ . & . & \ldots & . \\ X_nX_1 & X_nX_2 & \ldots & X_n^2 \end{bmatrix}.$$

If we define the expected value of a matrix to be the matrix of expected values of the matrix elements, then we can write the correlation matrix as:

$$\mathbf{R_X} = E[\mathbf{XX}^{\mathrm{T}}]. \tag{6.29a}$$

The covariance matrix is then:

$$\begin{aligned} \mathbf{K_X} &= E[(\mathbf{X} - \mathbf{m_X})(\mathbf{X} - \mathbf{m_X})^{\mathrm{T}}] \\ &= E[\mathbf{XX}^{\mathrm{T}}] - \mathbf{m_X}\, E[\mathbf{X}^{\mathrm{T}}] - E[\mathbf{X}]\mathbf{m_X}^{\mathrm{T}} + \mathbf{m_X}\mathbf{m_X}^{\mathrm{T}} \\ &= \mathbf{R_X} - \mathbf{m_X}\mathbf{m_X}^{\mathrm{T}}. \end{aligned} \tag{6.29b}$$

6.3.2 Linear Transformations of Random Vectors

Many engineering systems are linear in the sense that will be elaborated on in Chapter 10. Frequently these systems can be reduced to a linear transformation of a vector of random variables where the "input" is $\mathbf{X}$ and the "output" is $\mathbf{Y}$:

$$\mathbf{Y} = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_n \\ a_{21} & a_{22} & \ldots & a_{2n} \\ . & . & \ldots & . \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix} = \mathbf{AX}.$$

The expected value of the kth component of $\mathbf{Y}$ is the inner product (dot product) of the kth row of $\mathbf{A}$ and $\mathbf{X}$:

$$E[Y_k] = E\left[\sum_{j=1}^{n} a_{kj}X_j\right] = \sum_{j=1}^{n} a_{kj}E[X_j].$$

Each component of $E[\mathbf{Y}]$ is obtained in this manner, so:

$$\begin{aligned} \mathbf{m_Y} = E[\mathbf{Y}] = \begin{bmatrix} \sum_{j=1}^{n} a_{1j}E[X_j] \\ \sum_{j=1}^{n} a_{2j}E[X_j] \\ \vdots \\ \sum_{j=1}^{n} a_{nj}E[X_j] \end{bmatrix} &= \begin{bmatrix} a_{11} & a_{12} & \ldots & a_n \\ a_{21} & a_{22} & \ldots & a_{2n} \\ . & . & \ldots & . \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{bmatrix} \begin{bmatrix} E[X_1] \\ E[X_2] \\ \vdots \\ E[X_n] \end{bmatrix} \\ &= \mathbf{A}E[\mathbf{X}] = \mathbf{Am_X}. \end{aligned}$$

The covariance matrix of $\mathbf{Y}$ is then:

$$\begin{aligned}\mathbf{K_Y} &= E[(\mathbf{Y} - \mathbf{m_Y})(\mathbf{Y} - \mathbf{m_Y})^{\mathrm{T}}] = E[(\mathbf{AX} - \mathbf{Am_X})(\mathbf{AX} - \mathbf{Am_X})^{\mathrm{T}}] \\ &= E[\mathbf{A}(\mathbf{X} - \mathbf{m_X})(\mathbf{X} - \mathbf{m_X})^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}] = \mathbf{A}E[(\mathbf{X} - \mathbf{m_X})(\mathbf{X} - \mathbf{m_X})^{\mathrm{T}}]\mathbf{A}^{\mathrm{T}} \\ &= \mathbf{AK_XA}^{\mathrm{T}}, \end{aligned} \tag{6.30b}$$

where we used the fact that the transpose of a matrix multiplication is the product of the transposed matrices in reverse order: $\{\mathbf{A}(\mathbf{X} - \mathbf{m_X})\}^{\mathrm{T}} = (\mathbf{X} - \mathbf{m_X})^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}$.

The **cross-covariance** matrix of two random vectors $\mathbf{X}$ and $\mathbf{Y}$ is defined as:

$$\mathbf{K_{XY}} = E[(\mathbf{X} - \mathbf{m_X})(\mathbf{Y} - \mathbf{m_Y})^{\mathrm{T}}] = E[\mathbf{XY}^{\mathrm{T}}] - \mathbf{m_X m_Y}^{\mathrm{T}} = \mathbf{R_{XY}} - \mathbf{m_X m_Y}^{\mathrm{T}}.$$

We are interested in the cross-covariance between $\mathbf{X}$ and $\mathbf{Y} = \mathbf{AX}$:

$$\begin{aligned}\mathbf{K_{XY}} &= E[\mathbf{X} - \mathbf{m_X})(\mathbf{Y} - \mathbf{m_Y})^{\mathrm{T}}] = E[(\mathbf{X} - \mathbf{m_X})(\mathbf{X} - \mathbf{m_X})^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}] \\ &= \mathbf{K_X A}^{\mathrm{T}}. \end{aligned} \tag{6.30c}$$

Suppose that the components of $\mathbf{X}$ are correlated so $\mathbf{K_X}$ is not a diagonal matrix. In many situations we are interested in finding a transformation matrix $\mathbf{A}$ so that $\mathbf{Y} = \mathbf{AX}$ has uncorrelated components. This requires finding $\mathbf{A}$ so that $\mathbf{K_Y} = \mathbf{AK_XA}^{\mathrm{T}}$ is a diagonal matrix. In the last part of this section we show how to find such a matrix $\mathbf{A}$.

6.3.3 *Joint Characteristic Function

6.3.4 *Diagonalization of Covariance Matrix

Let $\mathbf{X}$ be a random vector with covariance $\mathbf{K_X}$. We are interested in finding an $n \times n$ matrix $\mathbf{A}$ such that $\mathbf{Y} = \mathbf{AX}$ has a covariance matrix that is diagonal. The components of $\mathbf{Y}$ are then uncorrelated.

We saw that $\mathbf{K_X}$ is a real-valued symmetric matrix. In Appendix C we state results from linear algebra that $\mathbf{K_X}$ is then a diagonalizable matrix, that is, there is a matrix $\mathbf{P}$ such that:

$$\mathbf{P}^{\mathrm{T}}\mathbf{K_X P} = \mathbf{\Lambda} \quad \text{and} \quad \mathbf{P}^{\mathrm{T}}\mathbf{P} = \mathbf{I} \tag{6.38a}$$

where $\mathbf{\Lambda}$ is a diagonal matrix and $\mathbf{I}$ is the identity matrix. Therefore if we let $\mathbf{A} = \mathbf{P}^{\mathrm{T}}$, then from Eq. (6.30b) we obtain a diagonal $\mathbf{K_Y}$.

We now show how $\mathbf{P}$ is obtained. First, we find the eigenvalues and eigenvectors of $\mathbf{K_X}$ from:

$$\mathbf{K_X e}_i = \lambda_i \mathbf{e}_i \tag{6.38b}$$

where $\mathbf{e}_i$ are $n \times 1$ column vectors.[2] We can normalize each eigenvector $\mathbf{e}_i$ so that $\mathbf{e}_i^{\mathrm{T}}\mathbf{e}_i$, the sum of the square of its components, is 1. The normalized eigenvectors are then orthonormal, that is,

$$\mathbf{e}_i^{\mathrm{T}}\mathbf{e}_j = \delta_{i,j} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j. \end{cases} \tag{6.38c}$$

Let $\mathbf{P}$ be the matrix whose columns are the eigenvectors of $\mathbf{K_X}$ and let $\mathbf{\Lambda}$ be the diagonal matrix of eigenvalues:

$$\mathbf{P} = [\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n] \qquad \mathbf{\Lambda} = \text{diag}[\lambda_1].$$

From Eq. (6.38b) we have:

$$\begin{aligned}\mathbf{K_X P} &= \mathbf{K_X}[\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n] = [\mathbf{K_X e}_1, \mathbf{K_X e}_2, \ldots, \mathbf{K_X e}_n] \\ &= [\lambda_1 \mathbf{e}_1, \lambda_2 \mathbf{e}_2, \ldots, \lambda_n \mathbf{e}_n] = \mathbf{P\Lambda} \qquad (6.39a)\end{aligned}$$

where the second equality follows from the fact that each column of $\mathbf{K_X P}$ is obtained by multiplying a column of $\mathbf{P}$ by $\mathbf{K_X}$. By premultiplying both sides of the above equations by $\mathbf{P}^{\mathrm{T}}$, we obtain:

$$\mathbf{P}^{\mathrm{T}}\mathbf{K_X P} = \mathbf{P}^{\mathrm{T}}\mathbf{P\Lambda} = \mathbf{\Lambda}. \qquad (6.39b)$$

We conclude that if we let $\mathbf{A} = \mathbf{P}^{\mathrm{T}}$, and

$$\mathbf{Y} = \mathbf{AX} = \mathbf{P}^{\mathrm{T}}\mathbf{X}, \qquad (6.40a)$$

then the random variables in $\mathbf{Y}$ are uncorrelated since

$$\mathbf{K_Y} = \mathbf{P}^{\mathrm{T}}\mathbf{K_X P} = \mathbf{\Lambda}. \qquad (6.40b)$$

In summary, any covariance matrix $\mathbf{K_X}$. can be diagonalized by a linear transformation. The matrix $\mathbf{A}$ in the transformation is obtained from the eigenvectors of $\mathbf{K_X}$.

It is interesting to look at the vector $\mathbf{X}$ expressed in terms of $\mathbf{Y}$. Multiply both sides of Eq. (6.40a) by $\mathbf{P}$ and use the fact that $\mathbf{PP}^{\mathrm{T}} = \mathbf{I}$:

$$\mathbf{X} = \mathbf{PP}^{\mathrm{T}}\mathbf{X} = \mathbf{PY} = [\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n]\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \sum_{k=1}^{n} Y_k \mathbf{e}_k. \qquad (6.41)$$

This equation is called the **Karhunen-Loeve expansion**. The equation shows that a random vector $\mathbf{X}$ can be expressed as a weighted sum of the eigenvectors of $\mathbf{K_X}$, where the coefficients are *uncorrelated* random variables Y_k. Furthermore, the eigenvectors form an orthonormal set. Note that if any of the eigenvalues are zero, $\text{VAR}[Y_k] = \lambda_k = 0$, then $Y_k = 0$, and the corresponding term can be dropped from the expansion in Eq. (6.41). In Chapter 10, we will see that this expansion is very useful in the processing of random signals.

6.4 Jointly Gaussian Random Vectors

The random variables $X_1, X_2, \ldots, X_n$ are said to be jointly Gaussian if their joint pdf is given by

$$f_{\mathbf{X}}(\mathbf{x}) \triangleq f_{X_1,X_2,\ldots,X_n}(x_1,\ldots,x_n) = \frac{\exp\{-\frac{1}{2}(\mathbf{x}-\mathbf{m})^{\mathrm{T}}K^{-1}(\mathbf{x}-\mathbf{m})\}}{(2\pi)^{n/2}|K|^{1/2}}, \quad (6.42a)$$

where $\mathbf{x}$ and $\mathbf{m}$ are column vectors defined by

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \qquad \mathbf{m} = \begin{bmatrix} m_1 \\ m_2 \\ \vdots \\ m_n \end{bmatrix} = \begin{bmatrix} E[X_1] \\ E[X_2] \\ \vdots \\ E[X_n] \end{bmatrix}$$

and K is the covariance matrix that is defined by

$$K = \begin{bmatrix} \text{VAR}(X_1) & \text{COV}(X_1, X_2) & \ldots & \text{COV}(X_1, X_n) \\ \text{COV}(X_2, X_1) & \text{VAR}(X_2) & \ldots & \text{COV}(X_2, X_n) \\ \vdots & \vdots & & \vdots \\ \text{COV}(X_n, X_1) & \ldots & & \text{VAR}(X_n) \end{bmatrix}. \quad (6.42b)$$

The $(.)^{\mathrm{T}}$ in Eq. (6.42a) denotes the transpose of a matrix or vector. Note that the covariance matrix is a symmetric matrix since $\text{COV}(X_i, X_j) = \text{COV}(X_j, X_i)$.

Equation (6.42a) shows that *the pdf of jointly Gaussian random variables is completely specified by the individual means and variances and the pairwise covariances*. It can be shown using the joint characteristic function that all the marginal pdf's associated with Eq. (6.42a) are also Gaussian and that these too are completely specified by the same set of means, variances, and covariances.

Example 6.21

The vector of random variables (X, Y, Z) is jointly Gaussian with zero means and covariance matrix:

$$K = \begin{bmatrix} \text{VAR}(X) & \text{COV}(X, Y) & \text{COV}(X, Z) \\ \text{COV}(Y, X) & \text{VAR}(Y) & \text{COV}(Y, Z) \\ \text{COV}(Z, X) & \text{COV}(Z, Y) & \text{VAR}(Z) \end{bmatrix} = \begin{bmatrix} 1.0 & 0.2 & 0.3 \\ 0.2 & 1.0 & 0.4 \\ 0.3 & 0.4 & 1.0 \end{bmatrix}.$$

Find the marginal pdf of X and Z.

We can solve this problem two ways. The first involves integrating the pdf directly to obtain the marginal pdf. The second involves using the fact that the marginal pdf for X and Z is also Gaussian and has the same set of means, variances, and covariances. We will use the second approach.

The pair (X, Z) has zero-mean vector and covariance matrix:

$$K' = \begin{bmatrix} \text{VAR}(X) & \text{COV}(X, Z) \\ \text{COV}(Z, X) & \text{VAR}(Z) \end{bmatrix} = \begin{bmatrix} 1.0 & 0.3 \\ 0.3 & 1.0 \end{bmatrix}.$$

The joint pdf of X and Z is found by substituting a zero-mean vector and this covariance matrix into Eq. (6.42a).

6.4.1 *Linear Transformation of Gaussian Random Variables

6.4.2 *Joint Characteristic Function of a Gaussian Random Variable

6.5 Estimation of Random Variables

In this book we will encounter two basic types of estimation problems. In the first type, we are interested in estimating the parameters of one or more random variables, e.g., probabilities, means, variances, or covariances.

In Chapter 1, we stated that relative frequencies can be used to estimate the probabilities of events, and that sample averages can be used to estimate the mean and other moments of a random variable. In Chapters 7 and 8 we will consider this type of estimation further.

In this section, we are concerned with the second type of estimation problem, where we are interested in estimating the value of an inaccessible random variable X in terms of the observation of an accessible random variable Y. For example, X could be the input to a communication channel and Y could be the observed output. In a prediction application, X could be a future value of some quantity and Y its present value.

Estimators:

Maximum a posteriori (MAP) estimator

Maximum likelihood (ML) estimator

Minimum MSE Estimator

6.5.1 MAP and ML Estimators

We have considered estimation problems informally earlier in the book. For example, in estimating the output of a discrete communications channel we are interested in finding the most probable input given the observation Y=y, that is, the value of input x that maximizes $P[X = x \mid Y = y]$:

$$\max_X P[X = x \mid Y = y]$$

In general we refer to the above estimator for X in terms of Y as the maximum a posteriori (MAP) estimator. The a posteriori probability is given by:

$$P[X = x \mid Y = y] = \frac{P[Y = y \mid X = x] \cdot P[X = x]}{P[Y = y]}$$

and so the MAP estimator requires that we know the a priori probabilities $P[X = x]$.

In some situations we know $P[Y = y \mid X = x]$ but we do not know the a priori probabilities, so we select the estimator value x as the value that maximizes the likelihood of the observed value Y=y:

$$\max_X P[Y = y \mid X = x]$$

We refer to this estimator of X in terms of Y as the maximum likelihood (ML) estimator.

We can define MAP and ML estimators when X and Y are continuous random variables by replacing events of the form $\{Y = y\}$ by $\{y < Y < y + dy\}$. If X and Y are continuous, the MAP estimator for X given the observation Y is given by:

$$\max_X f_X(X = x \mid Y = y)$$

and the ML estimator for X given the observation Y is given by:

$$\max_X f_Y(Y = y \mid X = x)$$

Example 6.25 Comparison of ML and MAP Estimators

Let X and Y be the random pair in Example 5.16. Find the MAP and ML estimators for X in terms of Y.

From Example 5.16 and 5.32, the conditional pdf of X given Y is given by:

$$f_{X,Y}(x,y)=\begin{cases}2\cdot e^{-x}\cdot e^{-y}, & 0\le y\le x<\infty\\ 0, & else\end{cases}$$

$$f_Y(y)=2\cdot e^{-2y} \quad \text{and} \quad f_X(x)=2\cdot e^{-x}\cdot\left(1-e^{-x}\right)$$

the conditional pdfs are:

$$f_Y(y\mid x)=\frac{f_{X,Y}(x,y)}{f_X(x)}=\frac{2\cdot e^{-x}\cdot e^{-y}}{2\cdot e^{-x}\cdot\left(1-e^{-x}\right)}=\frac{e^{-y}}{\left(1-e^{-x}\right)}, \quad for\ 0<y\le x$$

$$f_X(x\mid y)=\frac{f_{X,Y}(x,y)}{f_Y(y)}=\frac{2\cdot e^{-x}\cdot e^{-y}}{2\cdot e^{-2y}}=e^{-(x-y)}, \ for\ x\ge y$$

The MAP estimator

$$\max_X f_X(X=x\mid Y=y)$$

$$f_X(x\mid y)=e^{-(x-y)}, \qquad for\ x\ge y$$

As the function is a smooth curve in x, using the derivative to determine the max and min does not help and the bounds must be at either y or ∞. As the function decreases as x increases beyond y, the maximum must occur at y. Therefore the MAP estimator is $\hat{X}_{MAP}=y$.

The ML estimator

$$\max_X f_Y(Y=y\mid X=x)$$

$$f_Y(y\mid x)=\frac{e^{-y}}{\left(1-e^{-x}\right)}, \quad for\ 0<y\le x$$

The derivative is again not useful; however, as x increases beyond y, the denominator becomes larger so the conditional pdf decreases. Therefore the ML estimator is $\hat{X}_{ML}=y$.

In this example the ML and MAP estimators agree.

Example 6.26 Jointly Gaussian Random Variables

Find the MAP and ML estimator of X in terms of Y when X and Y are jointly Gaussian random variables.

Jointly Gaussian RV from section 5.9:

$$f_{X,Y}(x,y)=\frac{1}{2\pi\cdot\sigma_x\cdot\sigma_y\cdot\sqrt{1-\rho^2}}\cdot\exp\left[\frac{-\left(\left(\frac{x-m_x}{\sigma_x}\right)^2-2\cdot\rho\cdot\left(\frac{x-m_x}{\sigma_x}\right)\cdot\left(\frac{y-m_y}{\sigma_y}\right)+\left(\frac{y-m_y}{\sigma_y}\right)^2\right)}{2\cdot\left(1-\rho^2\right)}\right]$$

The conditional pdf of X given Y is given by:

$$f_X(x\,|\,y)=\frac{1}{\sqrt{2\pi\cdot\sigma_x^{\;2}\cdot\left(1-\rho^2\right)}}\cdot\exp\left[\frac{-\left(\left(x-\rho\cdot\frac{\sigma_x}{\sigma_y}\cdot\left(y-m_y\right)-m_x\right)^2\right)}{2\cdot\sigma_x^{\;2}\cdot\left(1-\rho^2\right)}\right]$$

which is maximized by the value of x for which the exponent is zero. Therefore

$$\hat{X}_{MAP}-\rho\cdot\frac{\sigma_x}{\sigma_y}\cdot\left(y-m_y\right)-m_x=0$$

$$\hat{X}_{MAP}=\rho\cdot\frac{\sigma_x}{\sigma_y}\cdot\left(y-m_y\right)+m_x$$

The conditional pdf of Y given X is given by:

$$f_Y(y\,|\,x)=\frac{1}{\sqrt{2\pi\cdot\sigma_y^{\;2}\cdot\left(1-\rho^2\right)}}\cdot\exp\left[\frac{-\left(\left(y-\rho\cdot\frac{\sigma_y}{\sigma_x}\cdot\left(x-m_x\right)-m_y\right)^2\right)}{2\cdot\sigma_y^{\;2}\cdot\left(1-\rho^2\right)}\right]$$

which is again maximized by the value of x for which the exponent is zero. Therefore

$$y-\rho\cdot\frac{\sigma_y}{\sigma_x}\cdot\left(\hat{X}_{ML}-m_x\right)-m_y=0$$

$$\hat{X}_{ML}=\frac{\sigma_x}{\rho\cdot\sigma_y}\cdot\left(y-m_y\right)+m_x$$

Therefore we conclude that the two estimators are not equal. In other words, knowledge of the a priori probabilities of X will affect the estimator.

6.5.2 Minimum MSE Linear Estimator

Another estimate for X is given by a function of the observation $\hat{X} = g(Y)$. In general, the estimation error, $\tilde{X} = X - \hat{X} = X - g(Y)$, is nonzero, and there is a cost associated with the error, $\text{cost} = c \cdot \tilde{X} = c \cdot (X - g(Y))$. We are usually interested in finding the function g(Y) that minimizes the expected value of the cost, $E[c \cdot \tilde{X}] = E[c \cdot (X - g(Y))]$. For example, if X and Y are the discrete input and output of a communication channel, and c is zero when $X = g(Y)$ and one otherwise, then the expected value of the cost corresponds to the probability of error, that is, that $X \neq g(Y)$ .When X and Y are continuous random variables, we frequently use the mean square error (MSE) as the cost:

$$e = E\left[\tilde{X}^2\right] = E\left[(X - g(Y))^2\right]$$

In the remainder of this section we focus on this particular cost function. We first consider the case where g(Y) is constrained to be a linear function of Y, and then consider the case where g(Y) can be any function, whether linear or nonlinear.

First, consider the problem of estimating a random variable X by a constant a so that the mean square error is minimized:

$$\min_a E\left[(X - a)^2\right] = E\left[X^2\right] - 2 \cdot a \cdot E[X] + a^2$$

The best a is found by taking the derivative with respect to a, setting the result to zero, and solving for a. The result is

$$-2 \cdot E[X] + 2 \cdot a = 0$$
$$a^* = E[X]$$

which makes sense since the expected value of X is the center of mass of the pdf. Note that the conjugate of a is shown in case a is complex.

The mean square error for this estimator is equal to $VAR[X] = E\left[(X - a^*)^2\right]$.

Now consider estimating X by a linear function $g(Y) = a \cdot Y + b$

$$\min_{a,b} E\left[(X - a \cdot Y - b)^2\right]$$

Equation (6.53a) can be viewed as the approximation of $X - a \cdot Y$ by the constant b. And then finding the best a. This is the minimization posed in Eq. (6.51) and the best b is

$$b^* = E[X - a \cdot Y] = E[X] - a \cdot E[Y]$$

Substitution into Eq. (6.53a) implies that the best a is found by

$$\min_a E\left[((X - E[X]) - a \cdot (Y - E[Y]))^2\right]$$

We once again differentiate with respect to a, set the result to zero, and solve for a:

$$-2 \cdot E[((X - E[X]) - a \cdot (Y - E[Y])) \cdot (Y - E[Y])] = 0$$

$$-2 \cdot E[(X - E[X]) \cdot (Y - E[Y])] - 2 \cdot a \cdot E[(Y - E[Y])^2] = 0$$

$$-2 \cdot COV(X,Y) - 2 \cdot a \cdot VAR[Y] = 0$$

The best coefficient a is found to be

$$a^* = \frac{COV(X,Y)}{VAR[Y]} = \frac{\rho_{X,Y} \cdot \sigma_X \cdot \sigma_Y}{\sigma_Y^{\ 2}} = \rho_{X,Y} \cdot \frac{\sigma_X}{\sigma_Y}$$

Therefore, the minimum mean square error (mmse) linear estimator for X in terms of Y is

$$\hat{X} = a^* \cdot Y + b^*$$

$$\hat{X} = a^* \cdot Y + E[X] - a^* \cdot E[Y]$$

$$\hat{X} = \rho_{X,Y} \cdot \frac{\sigma_X}{\sigma_Y} \cdot (Y - E[Y]) + E[X]$$

The term $(Y - E[Y])/\sigma_Y$ is simply a zero-mean, unit-variance version of Y. Thus $\sigma_X \cdot (Y - E[Y])/\sigma_Y$ is a rescaled version of Y that has the variance of the random variable that is being estimated, namely $\sigma_X^{\ 2}$ .The term E[X] simply ensures that the estimator has the correct mean. The key term in the above estimator is the correlation coefficient: $\rho_{X,Y}$ that specifies the sign and extent of the estimate of Y relative to $\sigma_X \cdot (Y - E[Y])/\sigma_Y$. If X and Y are uncorrelated then the best estimate for X is its mean, E[X]. On the other hand, if $\rho_{X,Y} = \pm 1$ then the best estimate is equal to

$$\hat{X} = \pm \sigma_X \cdot (Y - E[Y])/\sigma_Y + E[X]$$

We draw our attention to the second equality in Eq. (6.54):

$$E[\{(X - E[X]) - a^* \cdot (Y - E[Y])\} \cdot (Y - E[Y])] = 0$$

This equation is called the orthogonality condition because it states that the error of the best linear estimator, the quantity inside the braces, is orthogonal to the observation $(Y - E[Y])$ The orthogonality condition is a fundamental result in mean square estimation.

Computing the cost function for this estimator

$$\begin{aligned}
e_L^* &= E[((X - E[X]) - a^*(Y - E[Y]))^2] \\
&= E[((X - E[X]) - a^*(Y - E[Y]))(X - E[X])] \\
&\qquad - a^* E[((X - E[X]) - a^*(Y - E[Y]))(Y - E[Y])] \\
&= E[((X - E[X]) - a^*(Y - E[Y]))(X - E[X])] \\
&= \text{VAR}(X) - a^* \text{COV}(X, Y) \\
&= \text{VAR}(X)(1 - \rho_{X,Y}^2) \qquad (6.57)
\end{aligned}$$

where the second equality follows from the orthogonality condition. Note that when $\rho_{X,Y} = \pm 1$ the mean square error is zero. This implies that

$$P\left[X - a^* \cdot Y - b^* = 0\right] = P\left[X = a^* \cdot Y + b^*\right] = 1$$

so that X is essentially a linear function of Y.

6.5.3 Minimum MSE Estimator (generally nonlinear)

In general the estimator for X that minimizes the mean square error is a nonlinear function of Y. The estimator g(Y) that best approximates X in the sense of minimizing mean square error must satisfy

$$\min_{g(\cdot)} E\left[(X-g(Y))^2\right]$$

The problem can be solved by using conditional expectation:

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

$$E\left[(X-g(Y))^2\right]=\int_{-\infty}^{\infty} E\left[(X-g(Y))^2 \mid Y=y\right]\cdot f_Y(y)\cdot dy$$

The integrand above is positive for all y; therefore, the integral is minimized by minimizing $E\left[(X-g(Y))^2 \mid Y=y\right]$ for each y. But g(y) is a constant as far as the conditional expectation is concerned, so the problem is equivalent to Eq. (6.51) and the "constant" that minimizes $E\left[(X-g(Y))^2 \mid Y=y\right]$ is

$$g^*(Y)=E[X \mid Y=y]$$

The function is called the regression curve which simply traces the conditional expected value of X given the observation Y=y.

The mean square error of the best estimator is:

$$e^*=E\left[(X-g(Y))^2\right]=\int_R E\left[(X-E[X \mid Y=y])^2 \mid Y=y\right]\cdot f_Y(y)\cdot dy$$

$$e^*=\int_R VAR[X \mid Y=y]\cdot f_Y(y)\cdot dy$$

Linear estimators in general are suboptimal and have larger mean square errors.

6.5.4 Estimation using a Vector of Observations

The MAP, ML, and mean square estimators can be extended to where a vector of observations is available. Here we focus on mean square estimation. We wish to estimate X by a function g(Y) of a random vector of observations so that the mean square error is minimized:

$$\min_{g(\cdot)} E\left[(X - g(Y))^2\right]$$

To simplify the discussion we will assume that X and the have zero means. The same derivation that led to Eq. (6.58) leads to the optimum minimum mean square estimator:

$$g^*(Y) = E[X \mid Y = y]$$

The minimum mean square error is then:

$$e^* = E\left[(X - g(Y))^2\right] = \int_R E\left[(X - E[X \mid Y = y])^2 \mid Y = y\right] \cdot f_Y(y) \cdot dy$$

$$e^* = \int_R VAR[X \mid Y = y] \cdot f_Y(y) \cdot dy$$

Now suppose the estimate is a linear function of the observations:

$$g(\mathbf{Y}) = \sum_{k=1}^{n} a_k Y_k = \mathbf{a}^{\mathrm{T}} \mathbf{Y}.$$

The mean square error is now:

$$E[(X - g(\mathbf{Y}))^2] = E\left[\left(X - \sum_{k=1}^{n} a_k Y_k\right)^2\right].$$

We take derivatives with respect to a_k and again obtain the orthogonality conditions:

$$E\left[\left(X - \sum_{k=1}^{n} a_k Y_k\right) Y_j\right] = 0 \quad \text{for } j = 1, \ldots, n.$$

The orthogonality condition becomes:

$$E[XY_j] = E\left[\left(\sum_{k=1}^{n} a_k Y_k\right) Y_j\right] = \sum_{k=1}^{n} a_k E[Y_k Y_j] \quad \text{for } j = 1, \ldots, n.$$

We obtain a compact expression by introducing matrix notation:

$$E[X\mathbf{Y}] = \mathbf{R}_{\mathbf{Y}} \boldsymbol{a} \qquad \text{where } \boldsymbol{a} = (a_1, a_2, \ldots, a_n)^{\mathrm{T}}. \tag{6.60}$$

where $E[X\mathbf{Y}] = [E[XY_1], E[XY_2], \ldots, E[XY_n]]^{\mathrm{T}}$ and $\mathbf{R}_{\mathbf{Y}}$ is the correlation matrix. Assuming $\mathbf{R}_{\mathbf{Y}}$ is invertible, the optimum coefficients are:

$$\boldsymbol{a} = \mathbf{R}_Y^{-1} E[X\mathbf{Y}]. \tag{6.61a}$$

Note that the solution is based on an "auto-correlation" matrix and a "cross-correlation" vector.

We can use the methods from Section 6.3 to invert $\mathbf{R_Y}$. The mean square error of the optimum linear estimator is:

$$E[(X - \boldsymbol{a}^{\mathrm{T}}\mathbf{Y})^2] = E[(X - \boldsymbol{a}^{\mathrm{T}}\mathbf{Y})X] - E[(X - \boldsymbol{a}^{\mathrm{T}}\mathbf{Y})\boldsymbol{a}^{\mathrm{T}}\mathbf{Y}]$$
$$= E[(X - \boldsymbol{a}^{\mathrm{T}}\mathbf{Y})X] = \mathrm{VAR}(X) - \boldsymbol{a}^{\mathrm{T}}E[\mathbf{Y}X]. \quad (6.61\mathrm{b})$$

Example 6.30 Diversity Receiver

A radio receiver has two antennas to receive noisy versions of a signal X. The desired signal X is a Gaussian random variable with zero mean and variance 2. The signals received in the first and second antennas are $Y_1 = X + N_1$ and $Y_2 = X + N_2$ where N_1 and N_2 are zero-mean, unit-variance Gaussian random variables. In addition, X, N_1, and N_2 are independent random variables. Find the optimum mean square error linear estimator for X based on a single antenna signal and the corresponding mean square error. Compare the results to the optimum mean square estimator for X based on both antenna signals $\mathbf{Y} = (Y_1, Y_2)$.

Since all random variables have zero mean, we only need the correlation matrix and the cross-correlation vector in Eq. (6.61):

$$\mathbf{R}_Y = \begin{bmatrix} E[Y_1^2] & E[Y_1Y_2] \\ E[Y_1Y_2] & E[Y_2^2] \end{bmatrix}$$
$$= \begin{bmatrix} E[(X + N_1)^2] & E[(X + N_1)(X + N_2)] \\ E[(X + N_1)(X + N_2)] & E[(X + N_2)^2] \end{bmatrix}$$
$$= \begin{bmatrix} E[X^2] + E[N_1^2] & E[X^2] \\ E[X^2] & E[X^2] + E[N_2^2] \end{bmatrix} = \begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix}$$

and

$$E[X\mathbf{Y}] = \begin{bmatrix} E[XY_1] \\ E[XY_2] \end{bmatrix} = \begin{bmatrix} E[X^2] \\ E[X^2] \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}.$$

The optimum estimator using a single antenna received signal involves solving the 1×1 version of the above system:

$$\hat{X} = \frac{E[X^2]}{E[X^2] + E[N_1^2]}Y_1 = \frac{2}{3}Y_1$$

and the associated mean square error is:

$$\mathrm{VAR}(X) - a^* \,\mathrm{COV}(Y_1, X) = 2 - \frac{2}{3}2 = \frac{2}{3}.$$

The coefficients of the optimum estimator using two antenna signals are:

$$\boldsymbol{a} = \mathbf{R_Y}^{-1}E[X\mathbf{Y}] = \begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix}^{-1}\begin{bmatrix} 2 \\ 2 \end{bmatrix} = \frac{1}{5}\begin{bmatrix} 3 & -2 \\ -2 & 3 \end{bmatrix}\begin{bmatrix} 2 \\ 2 \end{bmatrix} = \begin{bmatrix} 0.4 \\ 0.4 \end{bmatrix}$$

and the optimum estimator is:

$$\hat{X} = 0.4Y_1 + 0.4Y_2.$$

The mean square error for the two antenna estimator is:

$$E[(X - a^{\mathrm{T}}\mathbf{Y})^2] = \mathrm{VAR}(X) - a^{\mathrm{T}}E[\mathbf{Y}X] = 2 - [0.4, 0.4]\begin{bmatrix} 2 \\ 2 \end{bmatrix} = 0.4.$$

As expected, the two antenna system has a smaller mean square error. Note that the receiver adds the two received signals and scales the result by 0.4. The sum of the signals is:

$$\hat{X} = 0.4Y_1 + 0.4Y_2 = 0.4(2X + N_1 + N_2) = 0.8\left(X + \frac{N_1 + N_2}{2}\right)$$

so combining the signals keeps the desired signal portion, X, constant while averaging the two noise signals N_1 and N_2. The problems at the end of the chapter explore this topic further.

Example 6.31 Second-Order Prediction of Speech

Let $X_1, X_2, \ldots$ be a sequence of samples of a speech voltage waveform, and suppose that the samples are fed into the second-order predictor shown in Fig. 6.3. Find the set of predictor coefficients a and b that minimize the mean square value of the predictor error when X_n is estimated by $aX_{n-2} + bX_{n-1}$.

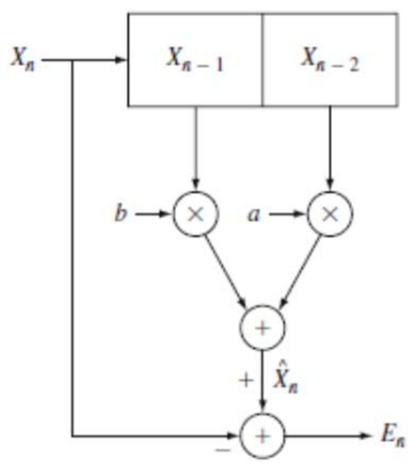


FIGURE 6.3
A two-tap linear predictor for processing speech.

We find the best predictor for X_1, X_2, and X_3 and assume that the situation is identical for X_2, X_3, and X_4 and so on. It is common practice to model speech samples as having zero mean and variance σ^2, and a covariance that does not depend on the specific index of the samples, but rather on the separation between them:

$$\text{COV}(X_j, X_k) = \rho_{|j-k|}\sigma^2.$$

The equation for the optimum linear predictor coefficients becomes

$$\sigma^2 \begin{bmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \sigma^2 \begin{bmatrix} \rho_2 \\ \rho_1 \end{bmatrix}.$$

Equation (6.61a) gives

$$a = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} \quad \text{and} \quad b = \frac{\rho_1(1 - \rho_1^2)}{1 - \rho_1^2}.$$

In Problem 6.78, you are asked to show that the mean square error using the above values of a and b is

$$\sigma^2 \left\{ 1 - \rho_1^2 - \frac{(\rho_1^2 - \rho_2)^2}{1 - \rho_1^2} \right\}. \tag{6.64}$$

Typical values for speech signals are $\rho_1 = .825$ and $\rho_2 = .562$. The mean square value of the predictor output is then $.281\sigma^2$. The lower variance of the output $(.281\sigma^2)$ relative to the input variance (σ^2) shows that the linear predictor is effective in anticipating the next sample in terms of the two previous samples. The order of the predictor can be increased by using more terms in the linear predictor. Thus a third-order predictor has three terms and involves inverting a 3×3 correlation matrix, and an n-th order predictor will involve an $n \times n$ matrix. Linear predictive techniques are used extensively in speech, audio, image and video compression systems. We discuss linear prediction methods in greater detail in Chapter 10.

6.6 *Generating Correlated Vector Random Variables

Many applications involve vectors or sequences of correlated random variables. Computer simulation models of such applications therefore require methods for generating such random variables. In this section we present methods for generating vectors of random variables with specified covariance matrices. We also discuss the generation of jointly Gaussian vector random variables.

6.6.1 Generating Random Vectors with Specified Covariance Matrix

Suppose we wish to generate a random vector $\mathbf{Y}$ with an arbitrary valid covariance matrix $\mathbf{K_Y}$. Let $\mathbf{Y} = \mathbf{A}^{\mathrm{T}}\mathbf{X}$ as in Example 6.17, where $\mathbf{X}$ is a vector random variable with components that are uncorrelated, zero mean, and unit variance. $\mathbf{X}$ has covariance matrix equal to the identity matrix $\mathbf{K_X} = \mathbf{I}$, $\mathbf{m_Y} = \mathbf{Am_X} = \mathbf{0}$, and

$$\mathbf{K_Y} = \mathbf{A}^{\mathrm{T}}\mathbf{K_X}\mathbf{A} = \mathbf{A}^{\mathrm{T}}\mathbf{A}.$$

Let $\mathbf{P}$ be the matrix whose columns are the eigenvectors of $\mathbf{K_Y}$ and let $\mathbf{\Lambda}$ be the diagonal matrix of eigenvalues, then from Eq. (6.39b) we have:

$$\mathbf{P}^{\mathrm{T}}\mathbf{K_Y}\mathbf{P} = \mathbf{P}^{\mathrm{T}}\mathbf{P}\mathbf{\Lambda} = \mathbf{\Lambda}.$$

If we premultiply the above equation by $\mathbf{P}$ and then postmultiply by $\mathbf{P}^{\mathrm{T}}$, we obtain expression for an arbitrary covariance matrix $\mathbf{K_Y}$ in terms of its eigenvalues and eigenvectors:

$$\mathbf{P\Lambda P}^{\mathrm{T}} = \mathbf{PP}^{\mathrm{T}}\mathbf{K_Y}\mathbf{PP}^{\mathrm{T}} = \mathbf{K_Y}. \tag{6.65}$$

Define the matrix $\mathbf{\Lambda}^{1/2}$ as the diagonal matrix of square roots of the eigenvalues:

$$\mathbf{\Lambda}^{1/2} \triangleq \begin{bmatrix} \sqrt{\lambda_1} & 0 & \dots & 0 \\ 0 & \sqrt{\lambda_2} & \dots & 0 \\ \cdot & \cdot & \dots & \cdot \\ 0 & 0 & \dots & \sqrt{\lambda_n} \end{bmatrix}.$$

In Problem 6.53 we show that any covariance matrix $\mathbf{K_Y}$ is positive semi-definite, which implies that it has nonnegative eigenvalues, and so taking the square root is always possible. If we now let

$$\mathbf{A} = (\mathbf{P\Lambda}^{1/2})^{\mathrm{T}} \tag{6.66}$$

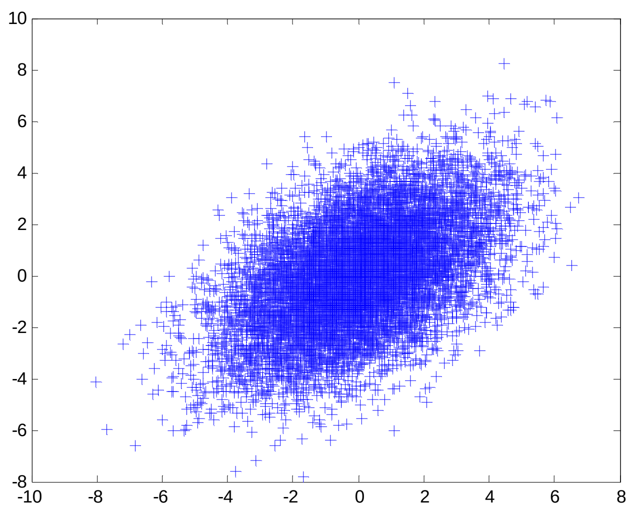
then

$$\mathbf{A}^{\mathrm{T}}\mathbf{A} = \mathbf{P\Lambda}^{1/2}\mathbf{\Lambda}^{1/2}\mathbf{P}^{\mathrm{T}} = \mathbf{P\Lambda P}^{\mathrm{T}} = \mathbf{K_Y}.$$

Therefore $\mathbf{Y}$ has the desired covariance matrix $\mathbf{K_Y}$.

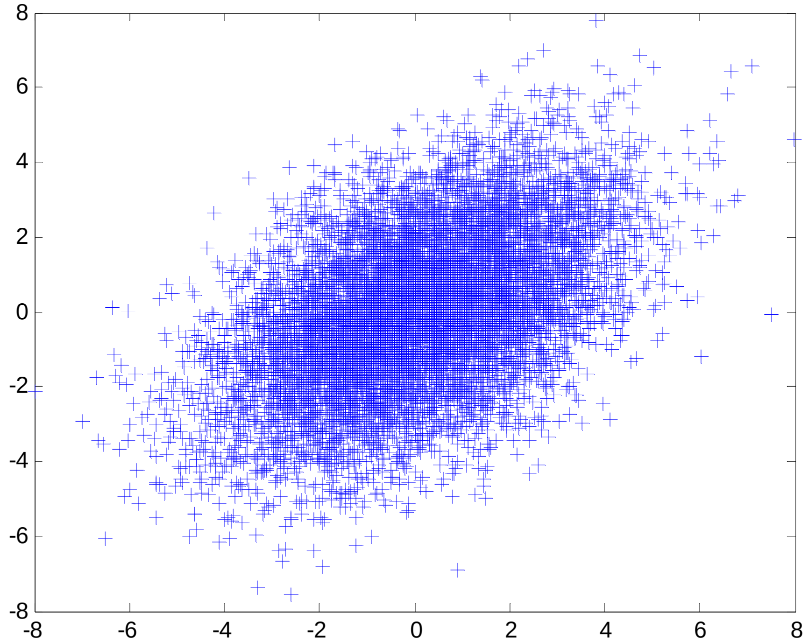
6.6.2 Generating Vectors of Jointly Gaussian Random Variables

```
%%
% Example 6.34
% The necessary steps for generating the Gaussian random variables
% with the covariance matrix from Example 6.32.

clear
%close all

U1=rand(10000, 1);        % Create a 1000-element vector U1.
U2=rand(10000, 1);       % Create a 1000-element vector U2.
R2=-2*log(U1);           % Find
TH=2*pi*U2;              % Find
X1=sqrt(R2).*sin(TH);    % Generate X1.
X2=sqrt(R2).*cos(TH);    % Generate X2.

mx1 = mean(X1)
mx2 = mean(X2)
vx1=sqrt(cov(X1))
vx2=sqrt(cov(X2))

Y1= X1+sqrt(3)*X2;         % Generate Y1.
Y2=-X1+sqrt(3)*X2;        % Generate Y2.

figure
plot(X1,X2,'+')           % Plot scattergram

figure
plot(Y1,Y2,'+')           % Plot scattergram

mx1 =    -0.0170
mx2 =    -0.0101
vx1 =     1.0009
vx2 =     0.9921
```



```
%%
% Example 6.34
% The necessary steps for generating the Gaussian random variables
% with the covariance matrix from Example 6.32.


clear
close all

X1=randn(10000, 1);     % Create a 1000-element vector U1.
X2=randn(10000, 1);     % Create a 1000-element vector U2.
mx1 = mean(X1)
mx2 = mean(X2)
vx1=sqrt(cov(X1))
vx2=sqrt(cov(X2))

Y1= X1+sqrt(3)*X2;       % Generate Y1.
Y2=-X1+sqrt(3)*X2;       % Generate Y2.

figure
plot(X1,X2,'+')          % Plot scattergram

figure
plot(Y1,Y2,'+')          % Plot scattergram
```

```
mx1 =    0.0154
mx2 =    0.0094
vx1 =    1.0072
vx2 =    0.9990
```



Notes and figures are based on or taken from materials in the textbook: Alberto Leon-Garcia, "Probability, Statistics, and Random Processes For Electrical Engineering, 3rd ed.", Pearson Prentice Hall, 2008, ISBN: 013-147122-8.

Summary

- The joint statistical behavior of a vector of random variables X is specified by the joint cumulative distribution function, the joint probability mass function, or the joint probability density function. The probability of any event involving the joint behavior of these random variables can be computed from these functions.

- The statistical behavior of subsets of random variables from a vector X is specified by the marginal cdf, marginal pdf, or marginal pmf that can be obtained from the joint cdf, joint pdf, or joint pmf of X.

- A set of random variables is independent if the probability of a product-form event is equal to the product of the probabilities of the component events. Equivalent conditions for the independence of a set of random variables are that the joint cdf, joint pdf, or joint pmf factors into the product of the corresponding marginal functions.

- The statistical behavior of a subset of random variables from a vector X, given the exact values of the other random variables in the vector, is specified by the conditional cdf, conditional pmf, or conditional pdf. Many problems naturally lend themselves to a solution that involves conditioning on the values of some of the random variables. In these problems, the expected value of random variables can be obtained through the use of conditional expectation.

- The mean vector and the covariance matrix provide summary information about a vector random variable. The joint characteristic function contains all of the information provided by the joint pdf.

- Transformations of vector random variables generate other vector random variables. Standard methods are available for finding the joint distributions of the new random vectors.

- The orthogonality condition provides a set of linear equations for finding the minimum mean square linear estimate. The best mean square estimator is given by the conditional expected value.

- The joint pdf of a vector X of jointly Gaussian random variables is determined by the vector of the means and by the covariance matrix. All marginal pdf's and conditional pdf's of subsets of X have Gaussian pdf's. Any linear function or linear transformation of jointly Gaussian random variables will result in a set of jointly Gaussian random variables.

- A vector of random variables with an arbitrary covariance matrix can be generated by taking a linear transformation of a vector of unit-variance, uncorrelated random variables. A vector of Gaussian random variables with an arbitrary covariance matrix can be generated by taking a linear transformation of a vector of independent, unit-variance jointly Gaussian random variables.

CHECKLIST OF IMPORTANT TERMS

Conditional cdf

Conditional expectation

Conditional pdf

Conditional pmf

Correlation matrix

Covariance matrix

Independent random variables

Jacobian of a transformation

Joint cdf

Joint characteristic function

Joint pdf

Joint pmf

Jointly continuous random variables

Jointly Gaussian random variables

Karhunen-Loeve expansion

MAP estimator

Marginal cdf

Marginal pdf

Marginal pmf

Maximum likelihood estimator

Mean square error

Mean vector

MMSE linear estimator

Orthogonality condition

Product-form event

Regression curve

Vector random variables