CHAPTER 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-work-with 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 $\mathbf{X}$ 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

$$\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix} = [X_1, X_2, \ldots, X_n]^{\mathrm{T}},$$

where "T" denotes the transpose of a matrix or vector. We will sometimes write $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ to save space and omit the transpose unless dealing with matrices. Possible values of the vector random variable are denoted by $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ 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 6.2 Joint Poisson Counts

A random experiment consists of finding the number of defects in a semiconductor chip and identifying their locations. The outcome of this experiment consists of the vector $\zeta = (n, \mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n)$, where the first component specifies the total number of defects and the remaining components specify the coordinates of their location. Suppose that the chip consists of M regions. Let $N_1(\zeta), N_2(\zeta), \ldots, N_M(\zeta)$ be the number of defects in each of these regions, that is, $N_k(\zeta)$ is the number of $\mathbf{y}$'s that fall in region k. The vector $\mathbf{N}(\zeta) = (N_1, N_2, \ldots, N_M)$ is then a vector random variable.

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.

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,

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

Equation (6.2) forms the basis for the definition of the n-dimensional joint probability mass function, cumulative distribution function, and probability density function. The probabilities of other events can be expressed in terms of these three functions.

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 $F_{X_1, X_2, \ldots, X_n}(x_1, x_2, \infty, \ldots, \infty)$.

Example 6.4

A radio transmitter sends a signal to a receiver using three paths. Let X_1, X_2, and X_3 be the signals that arrive at the receiver along each path. Find $P[\max(X_1, X_2, X_3) \le 5]$.

The maximum of three numbers is less than 5 if and only if each of the three numbers is less than 5; therefore

$$\begin{aligned} P[A] &= P[\{X_1 \le 5\} \cap \{X_2 \le 5\} \cap \{X_3 \le 5\}] \\ &= F_{X_1, X_2, X_3}(5, 5, 5). \end{aligned}$$

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

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

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

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

The joint pmf generates a family of **marginal pmf's** that specifies the joint probabilities for subcollections of the n random variables. For example, the one-dimensional pmf of X_j is found by adding the joint pmf over all variables other than x_j:

$$p_{X_j}(x_j) = P[X_j = x_j] = \sum_{x_1} \cdots \sum_{x_{j-1}} \sum_{x_{j+1}} \cdots \sum_{x_n} p_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n). \quad (6.7)$$

The two-dimensional joint pmf of any pair X_j and X_k is found by adding the joint pmf over all $n - 2$ other variables, and so on. Thus, the marginal pmf for $X_1, \ldots, X_{n-1}$ is given by

$$p_{X_1, \ldots, X_{n-1}}(x_1, x_2, \ldots, x_{n-1}) = \sum_{x_n} p_{X_1, \ldots, X_n}(x_1, x_2, \ldots, x_n). \quad (6.8)$$

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

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

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

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

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:

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

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 **marginal pdf's** is associated with the joint pdf in Eq. (6.12). The marginal pdf for a subset of the random variables is obtained by integrating the other variables out. For example, the marginal pdf of X_1 is

$$f_{X_1}(x_1) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1,X_2,\ldots,X_n}(x_1, x'_2, \ldots, x'_n)\, dx'_2 \ldots dx'_n. \tag{6.13}$$

As another example, the marginal pdf for $X_1, \ldots, X_{n-1}$ is given by

$$f_{X_1,\ldots,X_{n-1}}(x_1, \ldots, x_{n-1}) = \int_{-\infty}^{\infty} f_{X_1,\ldots,X_n}(x_1, \ldots, x_{n-1}, x'_n)\, dx'_n. \tag{6.14}$$

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

$$f_{X_1,X_2,X_3}(x_1,x_2,x_3) = \frac{e^{-(x_1^2+x_2^2-\sqrt{2}\,x_1x_2+{}^1\!/_2 x_3^2)}}{2\pi\sqrt{\pi}}.$$

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:

$$f_{X_1,X_3}(x_1,x_3) = \frac{e^{-x_3^2/2}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{e^{-(x_1^2+x_2^2-\sqrt{2}x_1x_2)}}{2\pi/\sqrt{2}}\,dx_2.$$

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:

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

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

Example 6.7 Multiplicative Sequence

Let X_1 be uniform in [0, 1], X_2 be uniform in $[0, X_1]$, and X_3 be uniform in $[0, X_2]$. (Note that X_3 is also the product of three uniform random variables.) Find the joint pdf of $\mathbf{X}$ and the marginal pdf of X_3.

For $0 < z < y < x < 1$, the joint pdf is nonzero and given by:

$$f_{X_1,X_2,X_3}(x_1,x_2,x_3) = f_{X_3}(z \mid x,y) f_{X_2}(y \mid x) f_{X_1}(x) = \frac{1}{y}\frac{1}{x}1 = \frac{1}{xy}.$$

The joint pdf of X_2 and X_3 is nonzero for $0 < z < y < 1$ and is obtained by integrating x between y and 1:

$$f_{X_2,X_3}(x_2, x_3) = \int_y^1 \frac{1}{xy} dx = \frac{1}{y} \ln x \Big|_y^1 = \frac{1}{y} \ln \frac{1}{y}.$$

We obtain the pdf of X_3 by integrating y between z and 1:

$$f_{X_3}(x_3) = -\int_z^1 \frac{1}{y} \ln y \, dy = -\frac{1}{2}(\ln y)^2 \Big|_z^1 = \frac{1}{2}(\ln z)^2.$$

Note that the pdf of X_3 is concentrated at the values close to $x = 0$.

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 $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ 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

$\mathbf{X} = (X_1, X_2, \ldots, X_n)$ may correspond to samples of a speech waveform and we may be interested in extracting features that are defined as functions of $\mathbf{X}$ for use in a speech recognition system.

6.2.1 One Function of Several Random Variables

Let the random variable Z be defined as a function of several random variables:

$$Z = g(X_1, X_2, \ldots, X_n). \tag{6.17}$$

The cdf of Z is found by finding the equivalent event of $\{Z \le z\}$, that is, the set $R_z = \{\mathbf{x}: g(\mathbf{x}) \le z\}$, then

$$F_Z(z) = P[\mathbf{X} \text{ in } R_z] = \int \underset{\mathbf{x} \text{ in } R_z}{\cdots} \int f_{X_1, \ldots, X_n}(x_1', \ldots, x_n')\, dx_1' \ldots dx_n'. \tag{6.18}$$

The pdf of Z is then found by taking the derivative of $F_Z(z)$.

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

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

Example 6.10 Merging of Independent Poisson Arrivals

Web page requests arrive at a server from n independent sources. Source j generates packets with exponentially distributed interarrival times with rate λ_j. Find the distribution of the interarrival times between consecutive requests at the server.

Let the interarrival times for the different sources be given by $X_1, X_2, \ldots, X_n$. Each X_j satisfies the memoryless property, so the time that has elapsed since the last arrival from each source is irrelevant. The time until the next arrival at the multiplexer is then:

$$Z = \min(X_1, X_2, \ldots, X_n).$$

Therefore the pdf of Z is:

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

$$= \big(1 - F_{X_1}(z)\big)\big(1 - F_{X_2}(z)\big)\dots\big(1 - F_{X_n}(z)\big)$$
$$= e^{-\lambda_1 z}e^{-\lambda_2 z}\dots e^{-\lambda_n z} = e^{-(\lambda_1+\lambda_2+\dots+\lambda_n)z}.$$

The interarrival time is an exponential random variable with rate $\lambda_1 + \lambda_2 + \dots + \lambda_n$.

Example 6.11 Reliability of Redundant Systems

A computing cluster has n independent redundant subsystems. Each subsystem has an exponentially distributed lifetime with parameter λ. The cluster will operate as long as at least one subsystem is functioning. Find the cdf of the time until the system fails.

Let the lifetime of each subsystem be given by $X_1, X_2, \dots, X_n$. The time until the last subsystem fails is:

$$W = \max(X_1, X_2, \dots, X_n).$$

Therefore the cdf of W is:

$$F_W(w) = \big(F_X(w)\big)^n = (1 - e^{-\lambda w})^n = 1 - \binom{n}{1}e^{-\lambda w} + \binom{n}{2}e^{-2\lambda w} + \dots.$$

6.2.2 Transformations of Random Vectors

Let $X_1, \dots, X_n$ be random variables in some experiment, and let the random variables $Z_1, \dots, Z_n$ be defined by a transformation that consists of n functions of $\mathbf{X} = (X_1, \dots, X_n)$:

$$Z_1 = g_1(\mathbf{X}) \qquad Z_2 = g_2(\mathbf{X}) \qquad \dots \qquad Z_n = g_n(\mathbf{X}).$$

The joint cdf of $\mathbf{Z} = (Z_1, \dots, Z_n)$ at the point $\mathbf{z} = (z_1, \dots, z_n)$ is equal to the probability of the region of $\mathbf{x}$ where $g_k(\mathbf{x}) \le z_k$ for $k = 1, \dots, n$:

$$F_{Z_1,\dots,Z_n}(z_1, \dots, z_n) = P[g_1(\mathbf{X}) \le z_1, \dots, g_n(\mathbf{X}) \le z_n]. \tag{6.19a}$$

If $X_1, \dots, X_n$ have a joint pdf, then

$$F_{Z_1,\dots,Z_n}(z_1, \dots, z_n) = \underset{x': g_k(x') \le z_k}{\int \dots \int} f_{X_1,\dots,X_n}(x_1', \dots, x_n')\, dx_1' \dots dx'. \tag{6.19b}$$

Example 6.12

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

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

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

$$F_{Z_1,Z_2,\ldots,Z_n}(z_1, z_2, \ldots, z_n) = P\left[X_1 \leq \frac{z_1 - b_1}{a_1}, X_2 \leq \frac{z_2 - b_2}{a_2}, \ldots, X_n \leq \frac{z_n - b_n}{a_n}\right]$$

$$= F_{X_1,X_2,\ldots,X_n}\left(\frac{z_1 - b_1}{a_1}, \frac{z_2 - b_2}{a_2}, \ldots, \frac{z_n - b_n}{a_n}\right)$$

$$f_{Z_1,Z_2,\ldots,Z_n}(z_1, z_2, \ldots, z_n) = \frac{\partial^n}{\partial z_1 \ldots \partial z_n} F_{Z_1,Z_2,\ldots,Z_n}(z_1, z_2, \ldots, z_n)$$

$$= \frac{1}{a_1 \ldots a_n} f_{X_1,X_2,\ldots,X_n}\left(\frac{z_1 - b_1}{a_1}, \frac{z_2 - b_2}{a_2}, \ldots, \frac{z_n - b_n}{a_n}\right).$$

---

*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,

Así cualquiera escribe un libro de probabilidad

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

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

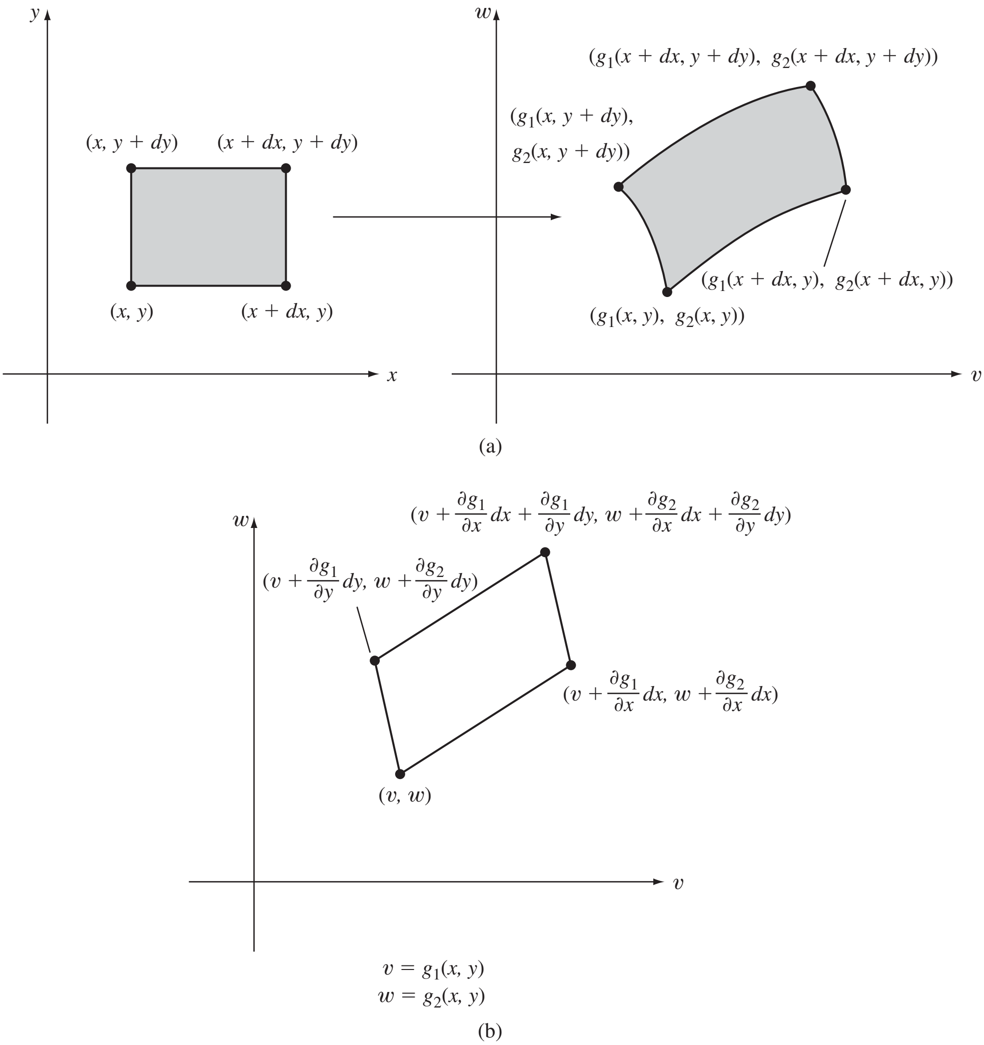

FIGURE 6.1
(a) Image of an infinitesimal rectangle under general transformation. (b) Approximation of image by a parallelogram.

"stretch factor" at the point (v, w) is given by the determinant of a matrix of partial derivatives:

$$J(x, y) = \det\begin{bmatrix} \dfrac{\partial v}{\partial x} & \dfrac{\partial v}{\partial y} \\[2ex] \dfrac{\partial w}{\partial x} & \dfrac{\partial w}{\partial y} \end{bmatrix}.$$

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

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

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.

Example 6.13

Server 1 receives m Web page requests and server 2 receives k Web page requests. Web page transmission times are exponential random variables with mean $1/\mu$. Let X be the total time to transmit files from server 1 and let Y be the total time for server 2. Find the joint pdf for T, the total transmission time, and W, the proportion of the total transmission time contributed by server 1:

$$T = X + Y \quad \text{and} \quad W = \frac{X}{X + Y}.$$

From Chapter 4, the sum of j independent exponential random variables is an Erlang random variable with parameters j and μ. Therefore X and Y are independent Erlang random variables with parameters m and μ, and k and μ, respectively:

$$f_X(x) = \frac{\mu e^{-\mu x}(\mu x)^{m-1}}{(m-1)!} \quad \text{and} \quad f_Y(y) = \frac{\mu e^{-\mu y}(\mu y)^{k-1}}{(k-1)!}.$$

We solve for X and Y in terms of T and W:

$$X = TW \quad \text{and} \quad Y = T(1 - W).$$

The Jacobian of the transformation is:

$$J(x, y) = \det\begin{bmatrix} 1 & 1 \\ \frac{y}{(x+y)^2} & \frac{-x}{(x+y)^2} \end{bmatrix}$$

$$= \frac{-x}{(x+y)^2} - \frac{y}{(x+y)^2} = \frac{-1}{x+y} = \frac{-1}{t}.$$

The joint pdf of T and W is then:

$$f_{T,W}(t,w) = \frac{1}{|J(x,y)|}\left[\frac{\mu e^{-\mu x}(\mu x)^{m-1}}{(m-1)!}\frac{\mu e^{-\mu y}(\mu y)^{k-1}}{(k-1)!}\right]_{\substack{x=tw \\ y=t(1-w)}}$$

$$= t\frac{\mu e^{-\mu tw}(\mu tw)^{m-1}}{(m-1)!}\frac{\mu e^{-\mu t(1-w)}(\mu t(1-w))^{k-1}}{(k-1)!}$$

$$= \frac{\mu e^{-\mu t}(\mu t)^{m+k-1}}{(m+k-1)!}\frac{(m+k-1)!}{(m-1)!(k-1)!}(w)^{m-1}(1-w)^{k-1}.$$

We see that T and W are independent random variables. As expected, T is Erlang with parameters $m + k$ and μ, since it is the sum of $m + k$ independent Erlang random variables. W is the beta random variable introduced in Chapter 3.

The method developed above can be used even if we are interested in only one function of a random variable. By defining an "auxiliary" variable, we can use the transformation method to find the joint pdf of both random variables, and then we can find the marginal pdf involving the random variable of interest. The following example demonstrates the method.

Example 6.14 Student's *t*-distribution

Let X be a zero-mean, unit-variance Gaussian random variable and let Y be a chi-square random variable with n degrees of freedom. Assume that X and Y are independent. Find the pdf of $V = X/\sqrt{Y/n}$.

Define the auxiliary function of $W = Y$. The variables X and Y are then related to V and W by

$$X = V\sqrt{W/n} \qquad \text{and} \qquad Y = W.$$

The Jacobian of the inverse transformation is

$$|J(v,w)| = \begin{vmatrix} \sqrt{w/n} & (v/2)\sqrt{wn} \\ 0 & 1 \end{vmatrix} = \sqrt{w/n}.$$

Since $f_{X,Y}(x,y) = f_X(x)f_Y(y)$, the joint pdf of V and W is thus

$$f_{V,W}(v,w) = \frac{e^{-x^2/2}}{\sqrt{2\pi}}\frac{(y/2)^{n/2-1}e^{-y/2}}{2\Gamma(n/2)}|J(v,w)|\Bigg|_{\substack{x = v\sqrt{w/n} \\ y = w}}$$

$$= \frac{(w/2)^{(n-1)/2}e^{-[(w/2)(1+v2/n)]}}{2\sqrt{n\pi}\Gamma(n/2)}.$$

The pdf of V is found by integrating the joint pdf over w:

$$f_V(v) = \frac{1}{2\sqrt{n\pi}\Gamma(n/2)}\int_0^\infty (w/2)^{(n-1)/2}e^{-[(w/2)(1+v2/n)]}\,dw.$$

If we let $w' = (w/2)(v^2/n + 1)$, the integral becomes

$$f_V(v) = \frac{(1+v^2/n)^{-(n+1)/2}}{\sqrt{n\pi}\Gamma(n/2)}\int_0^\infty (w')^{(n-1)/2}e^{-w'}\,dw'.$$

By noting that the above integral is the gamma function evaluated at $(n + 1)/2$, we finally obtain the *Student's t-distribution*:

$$f_V(v) = \frac{(1 + v^2/n)^{-(n+1)/2}\Gamma((n + 1)/2)}{\sqrt{n\pi}\Gamma(n/2)}.$$

This pdf is used extensively in statistical calculations. (See Chapter 8.)

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:

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

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:

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

Example 6.15 Sum of Random Variables

Given a random vector $\mathbf{X} = (X_1, X_2, X_3)$, find the joint pdf of the sum:

$$Z = X_1 + X_2 + X_3.$$

We will use the transformation by introducing auxiliary variables as follows:

$$Z_1 = X_1, Z_2 = X_1 + X_2, Z_3 = X_1 + X_2 + X_3.$$

The inverse transformation is given by:

$$X_1 = Z_1, X_2 = Z_2 - Z_1, X_3 = Z_3 - Z_2.$$

The Jacobian matrix is:

$$J(x_1, x_2, x_3) = \det \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} = 1.$$

Therefore the joint pdf of $\mathbf{Z}$ is

$$f_{\mathbf{Z}}(z_1, z_2, z_3) = f_{\mathbf{X}}(z_1, z_2 - z_1, z_3 - z_2).$$

The pdf of Z_3 is obtained by integrating with respect to z_1 and z_2:

$$f_{Z_3}(z) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{\mathbf{X}}(z_1, z_2 - z_1, z - z_2)\, dz_1 dz_2.$$

This expression can be simplified further if X_1, X_2, and X_3 are independent random variables.

---

[1]Appendix C provides a summary of definitions and useful results from linear algebra.

6.3 EXPECTED VALUES OF VECTOR RANDOM VARIABLES

In this section we are interested in the characterization of a vector random variable through the expected values of its components and of functions of its components. We focus on the characterization of a vector random variable through its mean vector and its covariance matrix. We then introduce the joint characteristic function for a vector random variable.

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

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

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

The mean, variance, and covariance provide useful information about the distribution of a random variable and are easy to estimate, so we are frequently interested in characterizing multiple random variables in terms of their first and second moments. We now introduce the mean vector and the covariance matrix. We then investigate the mean vector and the covariance matrix of a linear transformation of a random vector.

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:

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

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 $\mathbf{X}^T$, a $1 \times n$ matrix, we obtain the following $n \times n$ matrix:

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

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:

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

The covariance matrix is then:

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

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 $\mathbf{Y}$:

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

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:

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

$$= \mathbf{A}E[\mathbf{X}] = \mathbf{A}\mathbf{m_X}. \tag{6.30a}$$

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:

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

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

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

Example 6.17 Transformation of Uncorrelated Random Vector

Suppose that the components of $\mathbf{X}$ are uncorrelated and have unit variance, then $\mathbf{K_X} = \mathbf{I}$, the identity matrix. The covariance matrix for $\mathbf{Y} = \mathbf{AX}$ is

$$\mathbf{K_Y} = \mathbf{AK_XA}^{\mathrm{T}} = \mathbf{AIA}^{\mathrm{T}} = \mathbf{AA}^{\mathrm{T}}. \tag{6.31}$$

In general $\mathbf{K_Y} = \mathbf{AA}^{\mathrm{T}}$ is not a diagonal matrix and so the components of $\mathbf{Y}$ are correlated. In Section 6.6 we discuss how to find a matrix $\mathbf{A}$ so that Eq. (6.31) holds for a given $\mathbf{K_Y}$. We can then generate a random vector $\mathbf{Y}$ with any desired covariance matrix $\mathbf{K_Y}$.

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 $\mathbf{A}$.

Example 6.18 Transformation to Uncorrelated Random Vector

Suppose the random vector X_1, X_2, and X_3 in Example 6.16 is transformed using the matrix:

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

Find the $E[\mathbf{Y}]$ and $\mathbf{K_Y}$.

Since $\mathbf{m_X} = \mathbf{0}$, then $E[\mathbf{Y}] = \mathbf{Am_X} = \mathbf{0}$. The covariance matrix of $\mathbf{Y}$ is:

$$\mathbf{K_Y} = \mathbf{AK_XA}^{\mathrm{T}} = \frac{1}{2}\begin{bmatrix} 1 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & -\frac{1}{\sqrt{2}} & 0 \\ -\frac{1}{\sqrt{2}} & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$= \frac{1}{2}\begin{bmatrix} 1 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 - \frac{1}{\sqrt{2}} & 1 + \frac{1}{\sqrt{2}} & 0 \\ 1 - \frac{1}{\sqrt{2}} & -\left(1 + \frac{1}{\sqrt{2}}\right) & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 - \frac{1}{\sqrt{2}} & 0 & 0 \\ 0 & 1 + \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

The linear transformation has produced a vector of random variables $\mathbf{Y} = (Y_1, Y_2, Y_3)$ with components that are uncorrelated.

*6.3.3 Joint Characteristic Function

The **joint characteristic function** of n random variables is defined as

$$\Phi_{X_1,X_2,\ldots,X_n}(\omega_1, \omega_2, \ldots, \omega_n) = E[e^{j(\omega_1 X_1 + \omega_2 X_2 + \cdots + \omega_n X_n)}]. \tag{6.32a}$$

In this section we develop the properties of the joint characteristic function of two random variables. These properties generalize in straightforward fashion to the case of n random variables. Therefore consider

$$\Phi_{X,Y}(\omega_1, \omega_2) = E[e^{j(\omega_1 X + \omega_2 Y)}]. \tag{6.32b}$$

If X and Y are jointly continuous random variables, then

$$\Phi_{X,Y}(\omega_1, \omega_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x, y)e^{j(\omega_1 x + \omega_2 y)}\, dx\, dy. \tag{6.32c}$$

Equation (6.32c) shows that the joint characteristic function is the two-dimensional Fourier transform of the joint pdf of X and Y. The inversion formula for the Fourier transform implies that the joint pdf is given by

$$f_{X,Y}(x, y) = \frac{1}{4\pi^2} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \Phi_{X,Y}(\omega_1, \omega_2)e^{-j(\omega_1 x + \omega_2 y)}\, d\omega_1\, d\omega_2. \tag{6.33}$$

Note in Eq. (6.32b) that the marginal characteristic functions can be obtained from joint characteristic function:

$$\Phi_X(\omega) = \Phi_{X,Y}(\omega, 0) \quad \Phi_Y(\omega) = \Phi_{X,Y}(0, \omega). \tag{6.34}$$

If X and Y are independent random variables, then the joint characteristic function is the product of the marginal characteristic functions since

$$\begin{aligned} \Phi_{X,Y}(\omega_1, \omega_2) &= E[e^{j(\omega_1 X + \omega_2 Y)}] = E[e^{j\omega_1 X}e^{j\omega_2 Y}] \\ &= E[e^{j\omega_1 X}]E[e^{j\omega_2 Y}] = \Phi_X(\omega_1)\Phi_Y(\omega_2), \end{aligned} \tag{6.35}$$

where the third equality follows from Eq. (6.27).

The characteristic function of the sum $Z = aX + bY$ can be obtained from the joint characteristic function of X and Y as follows:

$$\Phi_Z(\omega) = E[e^{j\omega(aX+bY)}] = E[e^{j(\omega aX+\omega bY)}] = \Phi_{X,Y}(a\omega, b\omega). \tag{6.36a}$$

If X and Y are independent random variables, the characteristic function of $Z = aX + bY$ is then

$$\Phi_Z(\omega) = \Phi_{X,Y}(a\omega, b\omega) = \Phi_X(a\omega)\Phi_Y(b\omega). \tag{6.36b}$$

In Section 8.1 we will use the above result in dealing with sums of random variables.

The joint moments of X and Y (if they exist) can be obtained by taking the derivatives of the joint characteristic function. To show this we rewrite Eq. (6.32b) as the expected value of a product of exponentials and we expand the exponentials in a power series:

$$\begin{aligned}\Phi_{X,Y}(\omega_1, \omega_2) &= E[e^{j\omega_1 X}e^{j\omega_2 Y}]\\ &= E\left[\sum_{i=0}^{\infty}\frac{(j\omega_1 X)^i}{i!}\sum_{k=0}^{\infty}\frac{(j\omega_2 Y)^k}{k!}\right]\\ &= \sum_{i=0}^{\infty}\sum_{k=0}^{\infty}E[X^iY^k]\frac{(j\omega_1)^i}{i!}\frac{(j\omega_2)^k}{k!}.\end{aligned}$$

It then follows that the moments can be obtained by taking an appropriate set of derivatives:

$$E[X^iY^k] = \frac{1}{j^{i+k}}\frac{\partial^i\partial^k}{\partial\omega_1^i\partial\omega_2^k}\Phi_{X,Y}(\omega_1, \omega_2)\big|_{\omega_1=0,\omega_2=0}. \tag{6.37}$$

Example 6.19

Suppose U and V are independent zero-mean, unit-variance Gaussian random variables, and let

$$X = U + V \qquad Y = 2U + V.$$

Find the joint characteristic function of X and Y, and find $E[XY]$.

The joint characteristic function of X and Y is

$$\begin{aligned}\Phi_{X,Y}(\omega_1, \omega_2) &= E[e^{j(\omega_1 X+\omega_2 Y)}] = E[e^{j\omega_1(U+V)}e^{j\omega_2(2U+V)}]\\ &= E[e^{j((\omega_1+2\omega_2)U+(\omega_1+\omega_2)V)}].\end{aligned}$$

Since U and V are independent random variables, the joint characteristic function of U and V is equal to the product of the marginal characteristic functions:

$$\begin{aligned}\Phi_{X,Y}(\omega_1, \omega_2) &= E[e^{j((\omega_1+2\omega_2)U)}]E[e^{j((\omega_1+\omega_2)V)}]\\ &= \Phi_U(\omega_1 + 2\omega_2)\Phi_V(\omega_1 + \omega_2)\\ &= e^{-\frac{1}{2}(\omega_1+2\omega_2)^2}e^{-\frac{1}{2}(\omega_1+\omega_2)^2}\\ &= e^{\{-\frac{1}{2}(2\omega_1^2+6\omega_1\omega_2+5\omega_2^2)\}}.\end{aligned}$$

where marginal characteristic functions were obtained from Table 4.1.

The correlation $E[XY]$ is found from Eq. (6.37) with $i = 1$ and $k = 1$:

$$
\begin{aligned}
E[XY] &= \frac{1}{j^2}\frac{\partial^2}{\partial\omega_1\partial\omega_2}\Phi_{X,Y}(\omega_1,\omega_2)\,|_{\omega_1=0,\omega_2=0} \\
&= -\exp\{-\tfrac{1}{2}(2\omega_1^2+6\omega_1\omega_2+5\omega_2^2)\}[6\omega_1+10\omega_2]\left(\frac{1}{4}\right)[4\omega_1+6\omega_2] \\
&\quad + \frac{1}{2}\exp\{-\tfrac{1}{2}(2\omega_1^2+6\omega_1\omega_2+5\omega_2^2)\}[6]\,|_{\omega_1=0,\omega_2=0} \\
&= 3.
\end{aligned}
$$

You should verify this answer by evaluating $E[XY] = E[(U+V)(2U+V)]$ directly.

*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:

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

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:

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

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:

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

From Eq. (6.38b) we have:

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

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:

$$\mathbf{P}^{\mathrm{T}}\mathbf{K_X P} = \mathbf{P}^{\mathrm{T}}\mathbf{P\Lambda} = \mathbf{\Lambda}. \tag{6.39b}$$

[2]See Appendix C.

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

$$\mathbf{Y} = \mathbf{A}\mathbf{X} = \mathbf{P}^{\mathrm{T}}\mathbf{X}, \tag{6.40a}$$

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

$$\mathbf{K}_{\mathbf{Y}} = \mathbf{P}^{\mathrm{T}}\mathbf{K}_{\mathbf{X}}\mathbf{P} = \Lambda. \tag{6.40b}$$

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

Equation (6.40b) provides insight into the invertibility of $\mathbf{K}_{\mathbf{X}}$ and $\mathbf{K}_{\mathbf{Y}}$. From linear algebra we know that the determinant of a product of $n \times n$ matrices is the product of the determinants, so:

$$\det \mathbf{K}_{\mathbf{Y}} = \det \mathbf{P}^{\mathrm{T}} \det \mathbf{K}_{\mathbf{X}} \det \mathbf{P} = \det \Lambda = \lambda_1 \lambda_2 \ldots \lambda_n,$$

where we used the fact that $\det \mathbf{P}^{\mathrm{T}} \det \mathbf{P} = \det \mathbf{I} = 1$. Recall that a matrix is invertible if and only if its determinant is nonzero. Therefore $\mathbf{K}_{\mathbf{Y}}$ is not invertible if and only if one or more of the eigenvalues of $\mathbf{K}_{\mathbf{X}}$ is zero.

Now suppose that one of the eigenvalues is zero, say $\lambda_k = 0$. Since $\text{VAR}[Y_k] = \lambda_k = 0$, then $Y_k = 0$. But Y_k is defined as a linear combination, so

$$0 = Y_k = a_{k1}X_1 + a_{k2}X_2 + \cdots + a_{kn}X_n.$$

We conclude that the components of $\mathbf{X}$ are linearly dependent. Therefore, one or more of the components in $\mathbf{X}$ are redundant and can be expressed as a linear combination of the other components.

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 $\mathbf{P}\mathbf{P}^{\mathrm{T}} = \mathbf{I}$:

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

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 $\mathbf{K}_{\mathbf{X}}$, 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

En esta sección está mucho de lo q vimos en la clase del 31/3
Después lo veo, ahora necesito una medialuna

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

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

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

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

The $(.)^{\text{T}}$ 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 6.20

Verify that the two-dimensional Gaussian pdf given in Eq. (5.61a) has the form of Eq. (6.42a).

The covariance matrix for the two-dimensional case is given by

$$K = \begin{bmatrix} \sigma_1^2 & \rho_{X,Y}\sigma_1\sigma_2 \\ \rho_{X,Y}\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix},$$

where we have used the fact the $\text{COV}(X_1, X_2) = \rho_{X,Y}\sigma_1\sigma_2$. The determinant of K is $\sigma_1^2 \sigma_2^2(1 - \rho^2_{X,Y})$ so the denominator of the pdf has the correct form. The inverse of the covariance matrix is also a real symmetric matrix:

$$K^{-1} = \frac{1}{\sigma_1^2\sigma_2^2(1 - \rho_{X,Y}^2)} \begin{bmatrix} \sigma_2^2 & -\rho_{X,Y}\sigma_1\sigma_2 \\ -\rho_{X,Y}\sigma_1\sigma_2 & \sigma_1^2 \end{bmatrix}.$$

The term in the exponent is therefore

$$\frac{1}{\sigma_1^2\sigma_2^2(1 - \rho_{X,Y}^2)}(x - m_1, y - m_2) \begin{bmatrix} \sigma_2^2 & -\rho_{X,Y}\sigma_1\sigma_2 \\ -\rho_{X,Y}\sigma_1\sigma_2 & \sigma_1^2 \end{bmatrix} \begin{bmatrix} x - m_1 \\ y - m_2 \end{bmatrix}$$

$$= \frac{1}{\sigma_1^2\sigma_2^2(1 - \rho_{X,Y}^2)}(x - m_1, y - m_2) \begin{bmatrix} \sigma_2^2(x - m_1) - \rho_{X,Y}\sigma_1\sigma_2(y - m_2) \\ -\rho_{X,Y}\sigma_1\sigma_2(x - m_1) + \sigma_1^2(y - m_2) \end{bmatrix}$$

$$= \frac{((x - m_1)/\sigma_1)^2 - 2\rho_{X,Y}((x - m_1)/\sigma_1)((y - m_2)/\sigma_2) + ((y - m_2)/\sigma_2)^2}{(1 - \rho_{X,Y}^2)}.$$

Thus the two-dimensional pdf has the form of Eq. (6.42a).

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).

Example 6.22 Independence of Uncorrelated Jointly Gaussian Random Variables

Suppose $X_1, X_2, \ldots, X_n$ are jointly Gaussian random variables with $\text{COV}(X_i, X_j) = 0$ for $i \neq j$. Show that $X_1, X_2, \ldots, X_n$ are independent random variables.

From Eq. (6.42b) we see that the covariance matrix is a diagonal matrix:

$$K = \text{diag}[\text{VAR}(X_i)] = \text{diag}[\sigma_i^2]$$

Therefore

$$K^{-1} = \text{diag}\left[\frac{1}{\sigma_i^2}\right]$$

forma cuadrática?

and

$$(\mathbf{x} - \mathbf{m})^{\text{T}} K^{-1} (\mathbf{x} - \mathbf{m}) = \sum_{i=1}^{n} \left(\frac{x_i - m_i}{\sigma_i}\right)^2.$$

Thus from Eq. (6.42a)

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{\exp\left\{-\frac{1}{2}\sum_{i=1}^{n}[(x_i - m_i)/\sigma_i]^2\right\}}{(2\pi)^{n/2}|K|^{1/2}} = \prod_{i=1}^{n} \frac{\exp\left\{-\frac{1}{2}[(x_i - m_i)/\sigma_i]^2\right\}}{\sqrt{2\pi\sigma_i^2}} = \prod_{i=1}^{n} f_{X_i}(x_i).$$

Thus $X_1, X_2, \ldots, X_n$ are independent Gaussian random variables.

Example 6.23 Conditional pdf of Gaussian Random Variable

Find the conditional pdf of X_n given $X_1, X_2, \ldots, X_{n-1}$.

Let $\mathbf{K}_n$ be the covariance matrix for $\mathbf{X}_n = (X_1, X_2, \ldots, X_n)$ and $\mathbf{K}_{n-1}$ be the covariance matrix for $\mathbf{X}_{n-1} = (X_1, X_2, \ldots, X_{n-1})$. Let $\mathbf{Q}_n = \mathbf{K}_n^{-1}$ and $\mathbf{Q}_{n-1} = \mathbf{K}_{n-1}^{-1}$, then the latter matrices are submatrices of the former matrices as shown below:

$$\mathbf{K}_n = \begin{bmatrix} & \mathbf{K}_{n-1} & & K_{1n} \\ & & & K_{2n} \\ & & & \vdots \\ K_{1n} & K_{2n} & \dots & K_{nn} \end{bmatrix} \qquad \mathbf{Q}_n = \begin{bmatrix} & \mathbf{Q}_{n-1} & & Q_{1n} \\ & & & Q_{2n} \\ & & & \vdots \\ Q_{1n} & Q_{2n} & \dots & Q_{nn} \end{bmatrix}$$

Below we will use the subscript n or $n - 1$ to distinguish between the two random vectors and their parameters. The marginal pdf of X_n given $X_1, X_2, \dots, X_{n-1}$ is given by:

$$\begin{aligned} f_{X_n}(x_n \mid x_1, \dots, x_{n-1}) &= \frac{f_{\mathbf{X}_n}(\mathbf{x}_n)}{f_{\mathbf{X}_{n-1}}(\mathbf{x}_{n-1})} \\ &= \frac{\exp\{-\frac{1}{2}(\mathbf{x}_n - \mathbf{m}_n)^{\mathrm{T}}\mathbf{Q}_n(\mathbf{x}_n - \mathbf{m}_n)\}}{(2\pi)^{n/2}|\mathbf{K}_n|^{1/2}} \frac{(2\pi)^{(n-1)1/2}|\mathbf{K}_{n-1}|^{1/2}}{\exp\{-\frac{1}{2}(\mathbf{x}_{n-1} - \mathbf{m}_{n-1})^{\mathrm{T}}\mathbf{Q}_{n-1}(\mathbf{x}_{n-1} - \mathbf{m}_{n-1})\}} \\ &= \frac{\exp\{-\frac{1}{2}(\mathbf{x}_n - \mathbf{m}_n)^{\mathrm{T}}\mathbf{Q}_n(\mathbf{x}_n - \mathbf{m}_n) + \frac{1}{2}(\mathbf{x}_{n-1} - \mathbf{m}_{n-1})^{\mathrm{T}}\mathbf{Q}_{n-1}(\mathbf{x}_{n-1} - \mathbf{m}_{n-1})\}}{\sqrt{2\pi}|\mathbf{K}_n|^{1/2}/|\mathbf{K}_{n-1}|^{1/2}}. \end{aligned}$$

In Problem 6.60 we show that the terms in the above expression are given by:

$$\begin{aligned} &\tfrac{1}{2}(\mathbf{x}_n - \mathbf{m}_n)^{\mathrm{T}}\mathbf{Q}_n(\mathbf{x}_n - \mathbf{m}_n) - \tfrac{1}{2}(\mathbf{x}_{n-1} - \mathbf{m}_{n-1})^{\mathrm{T}}\mathbf{Q}_{n-1}(\mathbf{x}_{n-1} - \mathbf{m}_{n-1}) \\ &\qquad = Q_{nn}\{(x_n - m_n) + B\}^2 - Q_{nn}B^2 \end{aligned} \tag{6.43}$$

where $B = \dfrac{1}{Q_{nn}} \displaystyle\sum_{j=1}^{n-1} Q_{jn}(x_j - m_j)$ and $|\mathbf{K}_n|/|\mathbf{K}_{n-1}| = 1/Q_{nn}$.

This implies that X_n has mean $m_n - B$, and variance $1/Q_{nn}$. The term $Q_{nn}B^2$ is part of the normalization constant. We therefore conclude that:

$$f_{X_n}(x_n \mid x_1, \dots, x_{n-1}) = \frac{\exp\left\{-\dfrac{Q_{nn}}{2}\left(x - m_n + \dfrac{1}{Q_{nn}}\displaystyle\sum_{j=1}^{n-1} Q_{jn}(x_j - m_j)\right)^2\right\}}{\sqrt{2\pi/Q_{nn}}}$$

We see that the conditional mean of X_n is a linear function of the "observations" $x_1, x_2, \dots, x_{n-1}$.

---

*6.4.1 Linear Transformation of Gaussian Random Variables

A very important property of jointly Gaussian random variables is that *the linear transformation of any n jointly Gaussian random variables results in n random variables that are also jointly Gaussian*. This is easy to show using the matrix notation in Eq. (6.42a). Let $\mathbf{X} = (X_1, \dots, X_n)$ be jointly Gaussian with covariance matrix K_X and mean vector $\mathbf{m}_\mathbf{X}$ and define $\mathbf{Y} = (Y_1, \dots, Y_n)$ by

$$\mathbf{Y} = A\mathbf{X},$$

Prop importante: $(A^T)^{-1} = (A^{-1})^T = A^{-1T}$

$\rightarrow A^{-1T} K^{-1} A^{-1} = (A K A^T)^{-1} = C_Y^{-1} \rightarrow C_Y = A K A^T \rightarrow$ Debe ser

simétrica $\Rightarrow C_Y^T = (A K A^T)^T = (A^T)^T K^T A^T = A K A^T = C_Y$ ✓

where A is an invertible $n \times n$ matrix. From Eq. (5.60) we know that the pdf of $\mathbf{Y}$ is given by

$$
\begin{aligned}
f_{\mathbf{Y}}(\mathbf{y}) &= \frac{f_{\mathbf{X}}(A^{-1}\mathbf{y})}{|A|} \\
&= \frac{\exp\{-\frac{1}{2}(A^{-1}\mathbf{y} - \mathbf{mX})^{\mathrm{T}} K_X^{-1}(A^{-1}\mathbf{y} - \mathbf{mX})\}}{(2\pi)^{n/2}|A||K_X|^{1/2}}.
\end{aligned}
\tag{6.44}
$$

From elementary properties of matrices we have that

$$(A^{-1}\mathbf{y} - \mathbf{m_X}) = A^{-1}(\mathbf{y} - A\mathbf{m_X})$$

and

$$(A^{-1}\mathbf{y} - \mathbf{m_X})^{\mathrm{T}} = (\mathbf{y} - A\mathbf{m_X})^{\mathrm{T}} A^{-1\mathrm{T}}.$$

The argument in the exponential is therefore equal to

$$(\mathbf{y} - A\mathbf{m_X})^{\mathrm{T}} A^{-1\mathrm{T}} K_X^{-1} A^{-1}(\mathbf{y} - A\mathbf{m_X}) = (\mathbf{y} - A\mathbf{m_X})^{\mathrm{T}}(AK_XA^{\mathrm{T}})^{-1}(\mathbf{y} - A\mathbf{m_X})$$

since $A^{-1\mathrm{T}}K_X^{-1} = (AK_XA^{\mathrm{T}})^{-1}$. Letting $K_Y = AK_XA^{\mathrm{T}}$ and $\mathbf{m_Y} = A\mathbf{m_X}$ and noting that $\det(K_Y) = \det(AK_XA^{\mathrm{T}}) = \det(A)\det(K_X)\det(A^{\mathrm{T}}) = \det(A)^2\det(K_X)$, we finally have that the pdf of $\mathbf{Y}$ is

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{e^{-(1/2)(\mathbf{y}-\mathbf{m_Y})^{\mathrm{T}}K_Y^{-1}(\mathbf{y}-\mathbf{m_Y})}}{(2\pi)^{n/2}|K_Y|^{1/2}}. \tag{6.45}$$

$|A|\sqrt{|K_X|} = \sqrt{|A|^2 \cdot |K_X|} = \sqrt{|K_Y|}$

Thus the pdf of $\mathbf{Y}$ has the form of Eq. (6.42a) and therefore $Y_1, \ldots, Y_n$ are jointly Gaussian random variables with mean vector and covariance matrix:

Demostración piola ←

$$\mathbf{m_Y} = A\mathbf{m_X} \quad \text{and} \quad K_Y = AK_XA^{\mathrm{T}}.$$

This result is consistent with the mean vector and covariance matrix we obtained before in Eqs. (6.30a) and (6.30b).

In many problems we wish to transform $\mathbf{X}$ to a vector $\mathbf{Y}$ of independent Gaussian random variables. Since K_X is a symmetric matrix, it is always possible to find a matrix A such that $AK_XA^{\mathrm{T}} = \Lambda$ is a diagonal matrix. (See Section 6.6.) For such a matrix A, the pdf of $\mathbf{Y}$ will be

DIAGONALIZACIÓN

$$
\begin{aligned}
f_{\mathbf{Y}}(\mathbf{y}) &= \frac{e^{-(1/2)(\mathbf{y}-\mathbf{n})^{\mathrm{T}}\Lambda^{-1}(\mathbf{y}-\mathbf{n})}}{(2\pi)^{n/2}|\Lambda|^{1/2}} \\
&= \frac{\exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(y_i - n_i)^2/\lambda_i\right\}}{[(2\pi\lambda_1)(2\pi\lambda_2)\ldots(2\pi\lambda_n)]^{1/2}},
\end{aligned}
\tag{6.46}
$$

where $\lambda_1, \ldots, \lambda_n$ are the diagonal components of Λ. We assume that these values are all nonzero. The above pdf implies that $Y_1, \ldots, Y_n$ are *independent* random variables

Tener en cuenta que: 1) $(AB)^T = B^T A^T$ 2) $(AB)^{-1} = B^{-1} A^{-1}$

Doy vuelta y transpongo termino a termino
o doy vuelta e invierto termino a termino

Dusising hermano

with means n_i and variance λ_i. In conclusion, *it is possible to linearly transform a vector of jointly Gaussian random variables into a vector of independent Gaussian random variables*.

It is always possible to select the matrix A that diagonalizes K so that $\det(A) = 1$. The transformation $A\mathbf{X}$ then corresponds to a rotation of the coordinate system so that the principal axes of the ellipsoid corresponding to the pdf are aligned to the axes of the system. Example 5.48 provides an $n = 2$ example of rotation.

In computer simulation models we frequently need to generate jointly Gaussian random vectors with specified covariance matrix and mean vector. Suppose that $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ has components that are zero-mean, unit-variance Gaussian random variables, so its mean vector is $\mathbf{0}$ and its covariance matrix is the identity matrix $\mathbf{I}$. Let $\mathbf{K}$ denote the desired covariance matrix. Using the methods discussed in Section 6.3, it is possible to find a matrix $\mathbf{A}$ so that $\mathbf{A}^T\mathbf{A} = \mathbf{K}$. Therefore $\mathbf{Y} = \mathbf{A}^T\mathbf{U}$ has zero mean vector and covariance $\mathbf{K}$. From Eq. (6.46) we have that $\mathbf{Y}$ is also a jointly Gaussian random vector with zero mean vector and covariance $\mathbf{K}$. If we require a nonzero mean vector $\mathbf{m}$, we use $\mathbf{Y} + \mathbf{m}$.

Example 6.24 Sum of Jointly Gaussian Random Variables

Let $X_1, X_2, \ldots, X_n$ be jointly Gaussian random variables with joint pdf given by Eq. (6.42a). Let

$$Z = a_1X_1 + a_2X_2 + \cdots + a_nX_n.$$

We will show that Z is always a Gaussian random variable.

We find the pdf of Z by introducing auxiliary random variables. Let

$$Z_2 = X_2, \quad Z_3 = X_3, \ldots, \quad Z_n = X_n.$$

If we define $\mathbf{Z} = (Z_1, Z_2, \ldots, Z_n)$, then

$$\mathbf{Z} = A\mathbf{X}$$

where

$$A = \begin{bmatrix} a_1 & a_2 & \ldots & \cdot & a_n \\ 0 & 1 & \ldots & \cdot & 0 \\ \cdot & \cdot & \ldots & \cdot & \cdot \\ 0 & \cdot & \ldots & 0 & 1 \end{bmatrix}.$$

From Eq. (6.45) we have that $\mathbf{Z}$ is jointly Gaussian with mean $\mathbf{n} = A\mathbf{m}$, and covariance matrix $C = AKA^T$. Furthermore, it then follows that the marginal pdf of Z is a Gaussian pdf with mean given by the first component of $\mathbf{n}$ and variance given by the 1-1 component of the covariance matrix C. By carrying out the above matrix multiplications, we find that

$$E[Z] = \sum_{i=1}^{n} a_i E[X_i] \tag{6.47a}$$

$$\text{VAR}[Z] = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \,\text{COV}(X_i, X_j). \tag{6.47b}$$

FALOPA

*6.4.2 Joint Characteristic Function of a Gaussian Random Variable

The joint characteristic function is very useful in developing the properties of jointly Gaussian random variables. We now show that **the joint characteristic function of n jointly Gaussian random variables $X_1, X_2, \ldots, X_n$** is given by

$$\Phi_{X_1,X_2,\ldots,X_n}(\omega_1, \omega_2, \ldots, \omega_n) = e^{j\sum_{i=1}^{n}\omega_i m_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{k=1}^{n}\omega_i\omega_k \,\mathrm{COV}(X_i,X_k)}, \tag{6.48a}$$

which can be written more compactly as follows:

$$\Phi_{\mathbf{X}}(\boldsymbol{\omega}) \triangleq \Phi_{X_1,X_2,\ldots,X_n}(\omega_1, \omega_2, \ldots, \omega_n) = e^{j\boldsymbol{\omega}^{\mathrm{T}}\mathbf{m} - \frac{1}{2}\boldsymbol{\omega}^{\mathrm{T}}K\boldsymbol{\omega}}, \tag{6.48b}$$

where $\mathbf{m}$ is the vector of means and K is the covariance matrix defined in Eq. (6.42b).

Equation (6.48) can be verified by direct integration (see Problem 6.65). We use the approach in [Papoulis] to develop Eq. (6.48) by using the result from Example 6.24 that a linear combination of jointly Gaussian random variables is always Gaussian. Consider the sum

$$Z = a_1X_1 + a_2X_2 + \cdots + a_nX_n.$$

The characteristic function of Z is given by

$$\begin{aligned}\Phi_Z(\omega) &= E[e^{j\omega Z}] = E[e^{j(\omega a_1X_1 + \omega a_2X_2 + \cdots + \omega a_nX_n)}] \\ &= \Phi_{X_1,\ldots,X_n}(a_1\omega, a_2\omega, \ldots, a_n\omega).\end{aligned}$$

On the other hand, since Z is a Gaussian random variable with mean and variance given Eq. (6.47), we have

$$\begin{aligned}\Phi_Z(\omega) &= e^{j\omega E[Z] - \frac{1}{2}\mathrm{VAR}[Z]\omega^2} \\ &= e^{j\omega\sum_{i=1}^{n}a_i m_i - \frac{1}{2}\omega^2\sum_{i=1}^{n}\sum_{k=1}^{n}a_i a_k \,\mathrm{COV}(X_i,X_k)}.\end{aligned} \tag{6.49}$$

By equating both expressions for $\Phi_Z(\omega)$ with $\omega = 1$, we finally obtain

$$\begin{aligned}\Phi_{X_1,X_2,\ldots,X_n}(a_1, a_2, \ldots, a_n) &= e^{j\sum_{i=1}^{n} a_i m_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{k=1}^{n}a_i a_k \,\mathrm{COV}(X_i,X_k)} \\ &= e^{j\mathbf{a}^{\mathrm{T}}\mathbf{m} - \frac{1}{2}\mathbf{a}^{\mathrm{T}}K\mathbf{a}}.\end{aligned} \tag{6.50}$$

By replacing the a_i's with ω_i's we obtain Eq. (6.48).

The marginal characteristic function of any subset of the random variables $X_1, X_2, \ldots, X_n$ can be obtained by setting appropriate ω_i's to zero. Thus, for example, the marginal characteristic function of $X_1, X_2, \ldots, X_m$ for $m < n$ is obtained by setting $\omega_{m+1} = \omega_{m+2} = \cdots = \omega_n = 0$. Note that the resulting characteristic function again corresponds to that of jointly Gaussian random variables with mean and covariance terms corresponding the reduced set $X_1, X_2, \ldots, X_m$.

The derivation leading to Eq. (6.50) suggests an alternative definition for jointly Gaussian random vectors:

Definition: $\mathbf{X}$ is a jointly Gaussian random vector if and only every linear combination $Z = \boldsymbol{a}^{\mathrm{T}}\mathbf{X}$ is a Gaussian random variable.

In Example 6.24 we showed that if $\mathbf{X}$ is a jointly Gaussian random vector then the linear combination $Z = \boldsymbol{a}^{\mathrm{T}}\mathbf{X}$ is a Gaussian random variable. Suppose that we do not know the joint pdf of $\mathbf{X}$ but we are given that $Z = \boldsymbol{a}^{\mathrm{T}}\mathbf{X}$ is a Gaussian random variable for any choice of coefficients $\boldsymbol{a}^{\mathrm{T}} = (a_1, a_2, \ldots, a_n)$. This implies that Eqs. (6.48) and (6.49) hold, which together imply Eq. (6.50) which states that $\mathbf{X}$ has the characteristic function of a jointly Gaussian random vector.

The above definition is slightly broader than the definition using the pdf in Eq. (6.44). The definition based on the pdf requires that the covariance in the exponent be invertible. The above definition leads to the characteristic function of Eq. (6.50) which does not require that the covariance be invertible. Thus the above definition allows for cases where the covariance matrix is not invertible.

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.

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:

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

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:

$$\max_x f_X(Y = y \mid X = x).$$

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 5.32, the conditional pdf of X given Y is given by:

$$f_X(x \mid y) = e^{-(x-y)} \quad \text{for} \quad y \le x$$

which decreases as x increases beyond y. Therefore the MAP estimator is $\hat{X}_{\text{MAP}} = y$. On the other hand, the conditional pdf of Y given X is:

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

As x increases beyond y, the denominator becomes larger so the conditional pdf decreases. Therefore the ML estimator is $\hat{X}_{\text{ML}} = y$. 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.

The conditional pdf of X given Y is given by:

$$f_X(x|y) = \frac{\exp\left\{-\dfrac{1}{2(1-\rho^2)\sigma_X^2}\left(x - \rho\dfrac{\sigma_X}{\sigma_Y}(y - m_Y) - m_X\right)^2\right\}}{\sqrt{2\pi\sigma_X^2\,(1-\rho^2)}}$$

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

$$\hat{X}_{\text{MAP}} = \rho\frac{\sigma_X}{\sigma_Y}(y - m_Y) + m_X.$$

The conditional pdf of Y given X is:

$$f_Y(y|x) = \frac{\exp\left\{-\dfrac{1}{2(1-\rho^2)\sigma_Y^2}\left(y - \rho\dfrac{\sigma_Y}{\sigma_X}(x - m_X) - m_Y\right)^2\right\}}{\sqrt{2\pi\sigma_Y^2\,(1-\rho^2)}}.$$

which is also maximized for the value of x for which the exponent is zero:

$$0 = y - \rho\frac{\sigma_Y}{\sigma_X}(x - m_X) - m_Y.$$

The ML estimator for X given $Y = y$ is then:

$$\hat{X}_{\mathrm{ML}} = \frac{\sigma_X}{\rho\sigma_Y}(y - m_Y) + m_X.$$

Therefore we conclude that $\hat{X}_{\mathrm{ML}} \neq \hat{X}_{\mathrm{MAP}}$. In other words, knowledge of the a priori probabilities of X will affect the estimator.

6.5.2 Minimum MSE Linear Estimator

The estimate for X is given by a function of the observation $\hat{X} = g(Y)$. In general, the *estimation error*, $X - \hat{X} = X - g(Y)$, is nonzero, and there is a *cost* associated with the error, $c(X - g(Y))$. We are usually interested in finding the function $g(Y)$ that minimizes the expected value of the cost, $E[c(X - g(Y))]$. 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:

$$e = E[(X - g(Y))^2].$$

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:

$$\min_a E[(X - a)^2] = E[X^2] - 2aE[X] + a^2. \tag{6.51}$$

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

$$a^* = E[X], \tag{6.52}$$

which makes sense since the expected value of X is the center of mass of the pdf. The mean square error for this estimator is equal to $E[(X - a^*)^2] = \mathrm{VAR}(X)$.

Now consider estimating X by a *linear* function $g(Y) = aY + b$:

$$\min_{a,b} E[(X - aY - b)^2]. \tag{6.53a}$$

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

$$b^* = E[X - aY] = E[X] - aE[Y]. \tag{6.53b}$$

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

$$\min_a E[\{(X - E[X]) - a(Y - E[Y])\}^2].$$

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

$$0 = \frac{d}{da}E[(X - E[X]) - a(Y - E[Y])^2]$$

$$= -2E[\{(X - E[X]) - a(Y - E[Y])\}(Y - E[Y])]$$
$$= -2(\text{COV}(X, Y) - a\text{VAR}(Y)). \tag{6.54}$$

The best coefficient a is found to be

$$a^* = \frac{\text{COV}(X, Y)}{\text{VAR}(Y)} = \rho_{X,Y}\frac{\sigma_X}{\sigma_Y},$$

where $\sigma_Y = \sqrt{\text{VAR}(Y)}$ and $\sigma_X = \sqrt{\text{VAR}(X)}$. Therefore, ***the minimum mean square error (mmse) linear estimator*** for X in terms of Y is

$$\hat{X} = a^*Y + b^*$$
$$= \rho_{X,Y}\sigma_X\frac{Y - E[Y]}{\sigma_Y} + E[X]. \tag{6.55}$$

The term $(Y - E[Y])/\sigma_Y$ is simply a zero-mean, unit-variance version of Y. Thus $\sigma_X(Y - E[Y])/\sigma_Y$ is a rescaled version of Y that has the variance of the random variable that is being estimated, namely σ_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: *$\rho_{X,Y}$ specifies the sign and extent of the estimate of Y relative to $\sigma_X(Y - E[Y])/\sigma_Y$*. If X and Y are uncorrelated (i.e., $\rho_{X,Y} = 0$) 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 $\pm\sigma_X(Y - E[Y])/\sigma_Y + E[X]$.

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

$$E[\{(X - E[X]) - a^*(Y - E[Y])\}(Y - E[Y])] = 0. \tag{6.56}$$

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.

The mean square error of the best *linear* estimator is

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

where the second equality follows from the orthogonality condition. Note that when $|\rho_{X,Y}| = 1$, the mean square error is zero. This implies that $P[|X - a^*Y - b^*| = 0] = P[X = a^*Y + b^*] = 1$, so that X is essentially a linear function of Y.

6.5.3 Minimum MSE Estimator

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

$$\underset{g(.)}{\text{minimize}}\ E[(X - g(Y))^2].$$

The problem can be solved by using **conditional expectation**:

$$\begin{aligned} E[(X - g(Y))^2] &= E[E[(X - g(Y))^2 \mid Y]] \\ &= \int_{-\infty}^{\infty} E[(X - g(Y))^2 \mid Y = y] f_Y(y) dy. \end{aligned}$$

The integrand above is positive for all y; therefore, the integral is minimized by minimizing $E[(X - g(Y))^2 \mid Y = y]$ 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 $E[(X - g(y))^2 \mid Y = y]$ is

$$g^*(y) = E[X \mid Y = y]. \tag{6.58}$$

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

$$\begin{aligned} e^* = E[(X - g^*(Y))^2] &= \int_R E[(X - E[X \mid y])^2 \mid Y = y] f_Y(y)\, dy \\ &= \int_{R^n} \text{VAR}[X \mid Y = y] f_Y(y)\, dy. \end{aligned}$$

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

Example 6.27 Comparison of Linear and Minimum MSE Estimators

Let X and Y be the random pair in Example 5.16. Find the best linear and nonlinear estimators for X in terms of Y, and of Y in terms of X.

Example 5.28 provides the parameters needed for the linear estimator: $E[X] = 3/2$, $E[Y] = 1/2$, $\text{VAR}[X] = 5/4$, $\text{VAR}[Y] = 1/4$, and $\rho_{X,Y} = 1/\sqrt{5}$. Example 5.32 provides the conditional pdf's needed to find the nonlinear estimator. The best linear and nonlinear estimators for X in terms of Y are:

$$\hat{X} = \frac{1}{\sqrt{5}} \frac{\sqrt{5}}{2} \frac{Y - 1/2}{1/2} + \frac{3}{2} = Y + 1$$

$$E[X \mid y] = \int_y^{\infty} x e^{-(x-y)}\, dx = y + 1 \text{ and so } E[X \mid Y] = Y + 1.$$

Thus the optimum linear and nonlinear estimators are the same.

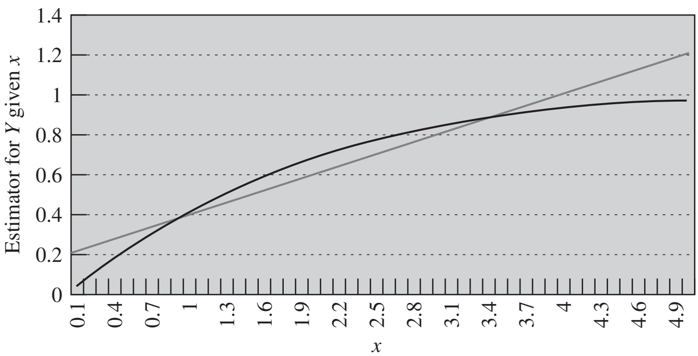

FIGURE 6.2
Comparison of linear and nonlinear estimators.

The best linear and nonlinear estimators for Y in terms of X are:

$$\hat{Y} = \frac{1}{\sqrt{5}}\frac{1}{2}\frac{X - 3/2}{\sqrt{5/2}} + \frac{1}{2} = (X + 1)/5.$$

$$E[Y \mid x] = \int_0^x y\frac{e^{-y}}{1 - e^{-x}}\,dy = \frac{1 - e^{-x} - xe^{-x}}{1 - e^{-x}} = 1 - \frac{xe^{-x}}{1 - e^{-x}}.$$

The optimum linear and nonlinear estimators are not the same in this case. Figure 6.2 compares the two estimators. It can be seen that the linear estimator is close to $E[Y \mid x]$ for lower values of x, where the joint pdf of X and Y are concentrated and that it diverges from $E[Y \mid x]$ for larger values of x.

Example 6.28

Let X be uniformly distributed in the interval $(-1, 1)$ and let $Y = X^2$. Find the best linear estimator for Y in terms of X. Compare its performance to the best estimator.

The mean of X is zero, and its correlation with Y is

$$E[XY] = E[XX^2] = \int_{-\frac{1}{2}}^{1} x^3/2\,dx = 0.$$

Therefore $\text{COV}(X, Y) = 0$ and the best linear estimator for Y is $E[Y]$ by Eq. (6.55). The mean square error of this estimator is the $\text{VAR}(Y)$ by Eq. (6.57).

The best estimator is given by Eq. (6.58):

$$E[Y \mid X = x] = E[X^2 \mid X = x] = x^2.$$

The mean square error of this estimator is

$$E[(Y - g(X))^2] = E[(X^2 - X^2)^2] = 0.$$

Thus in this problem, the best linear estimator performs poorly while the nonlinear estimator gives the smallest possible mean square error, zero.

Example 6.29 Jointly Gaussian Random Variables

Find the minimum mean square error estimator of X in terms of Y when X and Y are jointly Gaussian random variables.

The minimum mean square error estimator is given by the conditional expectation of X given Y. From Eq. (5.63), we see that the conditional expectation of X given $Y = y$ is given by

$$E[X \mid Y = y] = E[X] + \rho_{X,Y}\frac{\sigma_X}{\sigma_Y}(Y - E[Y]).$$

This is identical to the best linear estimator. Thus *for jointly Gaussian random variables the minimum mean square error estimator is linear*.

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 $g(\mathbf{Y})$ of a random vector of observations $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)^{\mathrm{T}}$ so that the **mean square error** is minimized:

$$\underset{g(.)}{\text{minimize}}\; E[(X - g(\mathbf{Y}))^2].$$

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

$$g^*(\boldsymbol{y}) = E[X \mid \mathbf{Y} = \boldsymbol{y}]. \tag{6.59}$$

The minimum mean square error is then:

$$E[(X - g^*(Y))^2] = \int_{R^n} E[(X - E[X \mid \mathbf{Y}])^2 \mid \mathbf{Y} = \mathbf{y}]f_{\mathbf{Y}}(\mathbf{y})d\mathbf{y}$$

$$= \int_{R^n} \mathrm{VAR}[X \mid \mathbf{Y} = \mathbf{y}]f_{\mathbf{Y}}(\mathbf{y})d\mathbf{y}.$$

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:

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

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

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

Now suppose that X has mean m_X and $\mathbf{Y}$ has mean vector $\mathbf{m}_{\mathbf{Y}}$, so our estimator now has the form:

$$\hat{X} = g(\mathbf{Y}) = \sum_{k=1}^{n} a_k Y_k + b = \boldsymbol{a}^{\mathrm{T}}\mathbf{Y} + b. \tag{6.62}$$

The same argument that led to Eq. (6.53b) implies that the optimum choice for b is:

$$b = E[X] - \boldsymbol{a}^{\mathrm{T}}\mathbf{m}_{\mathbf{Y}}.$$

Therefore the optimum linear estimator has the form:

$$\hat{X} = g(\mathbf{Y}) = \boldsymbol{a}^{\mathrm{T}}(\mathbf{Y} - \mathbf{m}_{\mathbf{Y}}) + m_X = \boldsymbol{a}^{\mathrm{T}}\mathbf{Z} + m_X$$

where $\mathbf{Z} = \mathbf{Y} - \mathbf{m}_{\mathbf{Y}}$ is a random vector with zero mean vector. The mean square error for this estimator is:

$$E[(X - g(\mathbf{Y}))^2] = E[(X - \boldsymbol{a}^{\mathrm{T}}\mathbf{Z} - m_X)^2] = E[(W - \boldsymbol{a}^{\mathrm{T}}\mathbf{Z})^2]$$

where $W = X - m_X$ has zero mean. We have reduced the general estimation problem to one with zero mean random variables, i.e., W and $\mathbf{Z}$, which has solution given by Eq. (6.61a). Therefore the optimum set of linear predictors is given by:

$$\boldsymbol{a} = \mathbf{R}_{\mathbf{z}}^{-1}E[W\mathbf{Z}] = \mathbf{K}_{\mathbf{Y}}^{-1}E[(X - m_X)(\mathbf{Y} - \mathbf{m}_{\mathbf{Y}})]. \tag{6.63a}$$

The mean square error is:

$$\begin{aligned} E[(X - \boldsymbol{a}^{\mathrm{T}}\mathbf{Y} - b)^2] &= E[(W - \boldsymbol{a}^{\mathrm{T}}\mathbf{Z}\, W] = \mathrm{VAR}(W) - \boldsymbol{a}^{\mathrm{T}}E[W\mathbf{Z}] \\ &= \mathrm{VAR}(X) - \boldsymbol{a}^{\mathrm{T}}E[(X - \mathbf{m}_{\mathbf{X}})(\mathbf{Y} - \mathbf{m}_{\mathbf{Y}})]. \end{aligned} \tag{6.63b}$$

This result is of particular importance in the case where X and $\mathbf{Y}$ are jointly Gaussian random variables. In Example 6.23 we saw that the conditional expected value of X given Y is a linear function of Y of the form in Eq. (6.62). Therefore in this case the optimum minimum mean square estimator corresponds to the optimum linear estimator.

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:

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

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

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

and the optimum estimator is:

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

The mean square error for the two antenna estimator is:

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

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 $aX_{n-2} + bX_{n-1}$.

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

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

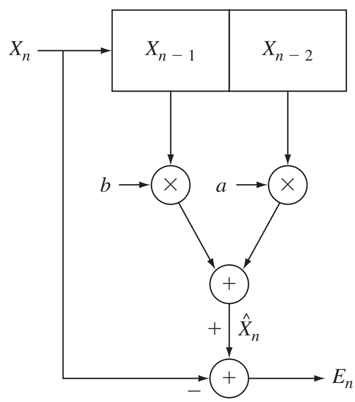

FIGURE 6.3
A two-tap linear predictor for processing speech.

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 $\mathbf{K_X} = \mathbf{I}$, $\mathbf{m_Y} = \mathbf{A}\mathbf{m_X} = \mathbf{0}$, 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 Λ be the diagonal matrix of eigenvalues, then from Eq. (6.39b) we have:

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

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:

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

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

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

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 $\mathbf{K_Y}$.

Example 6.32

Let $\mathbf{X} = (X_1, X_2)$ consist of two zero-mean, unit-variance, uncorrelated random variables. Find the matrix A such that $\mathbf{Y} = A\mathbf{X}$ has covariance matrix

$$K = \begin{bmatrix} 4 & 2 \\ 2 & 4 \end{bmatrix}.$$

First we need to find the eigenvalues of K which are determined from the following equation:

$$\det(\mathbf{K} - \lambda\mathbf{I}) = 0 = \det\begin{bmatrix} 4-\lambda & 2 \\ 2 & 4-\lambda \end{bmatrix} = (4-\lambda)^2 - 4 = \lambda^2 - 8\lambda + 12$$
$$= (\lambda - 6)(\lambda - 2).$$

We find the eigenvalues to be $\lambda_1 = 2$ and $\lambda_2 = 6$. Next we need to find the eigenvectors corresponding to each eigenvalue:

$$\begin{bmatrix} 4 & 2 \\ 2 & 4 \end{bmatrix}\begin{bmatrix} e_1 \\ e_2 \end{bmatrix} = \lambda_1\begin{bmatrix} e_1 \\ e_2 \end{bmatrix} = 2\begin{bmatrix} e_1 \\ e_2 \end{bmatrix}$$

which implies that $2e_1 + 2e_2 = 0$. Thus any vector of the form $[1, -1]^{\mathrm{T}}$ is an eigenvector. We choose the normalized eigenvector corresponding to $\lambda_1 = 2$ as $\mathbf{e}_1 = [1/\sqrt{2}, -1/\sqrt{2}]^{\mathrm{T}}$. We similarly find the eigenvector corresponding to $\lambda_2 = 6$ as $\mathbf{e}_2 = [1/\sqrt{2}, 1/\sqrt{2}]^{\mathrm{T}}$.

The method developed in Section 6.3 requires that we form the matrix P whose columns consist of the eigenvectors of K:

$$P = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}.$$

Next it requires that we form the diagonal matrix with elements equal to the square root of the eigenvalues:

$$\mathbf{\Lambda}^{1/2} = \begin{bmatrix} \sqrt{2} & 0 \\ 0 & \sqrt{6} \end{bmatrix}.$$

The desired matrix is then

$$\mathbf{A} = \mathbf{P\Lambda}^{1/2} = \begin{bmatrix} 1 & \sqrt{3} \\ -1 & \sqrt{3} \end{bmatrix}.$$

You should verify that $K = AA^{\mathrm{T}}$.

Example 6.33

Use Octave to find the eigenvalues and eigenvectors calculated in the previous example.

After entering the matrix K, we use the `eig(K)` function to find the matrix of eigenvectors P and eigenvalues Λ. We then find A and its transpose A^{T}. Finally we confirm that $A^{\mathrm{T}}A$ gives the desired covariance matrix.

```
> K=[4, 2; 2, 4];
> [P,D] =eig (K)
P =
 −0.70711  0.70711
  0.70711  0.70711
D =
 2  0
 0  6
> A=(P*sqrt(D))'
A =
 −1.0000  1.0000
  1.7321  1.7321
> A'
ans =
 −1.0000  1.7321
  1.0000  1.7321
> A'*A
ans =
 4.0000  2.0000
 2.0000  4.0000
```

The above steps can be used to find the transformation A^{T} for any desired covariance matrix K. The only check required is to ascertain that K is a valid covariance matrix: (1) K is symmetric (trivial); (2) K has positive eigenvalues (easy to check numerically).

6.6.2 Generating Vectors of Jointly Gaussian Random Variables

In Section 6.4 we found that if $\mathbf{X}$ is a vector of jointly Gaussian random variables with covariance K_X, then $\mathbf{Y} = A\mathbf{X}$ is also jointly Gaussian with covariance matrix $K_Y = AK_XA^{\mathrm{T}}$. If we assume that $\mathbf{X}$ consists of unit-variance, uncorrelated random variables, then $K_X = I$, the identity matrix, and therefore $K_Y = AA^{\mathrm{T}}$.

We can use the method from the first part of this section to find A for any desired covariance matrix K_Y. We generate jointly Gaussian random vectors $\mathbf{Y}$ with arbitrary covariance matrix K_Y and mean vector $\mathbf{m_Y}$ as follows:

1. Find a matrix A such that $K_Y = AA^{\mathrm{T}}$.
2. Use the method from Section 5.10 to generate $\mathbf{X}$ consisting of n independent, zero-mean, Gaussian random variables.
3. Let $\mathbf{Y} = A\mathbf{X} + \mathbf{m_Y}$.

Example 6.34

The Octave commands below show necessary steps for generating the Gaussian random variables with the covariance matrix from Example 6.30.

```
> U1=rand(1000, 1);          % Create a 1000-element vector U1.
> U2=rand(1000, 1);          % Create a 1000-element vector U2.
> R2=-2*log(U1);             % Find R^2.
> TH=2*pi*U2;                % Find Θ.

> X1=sqrt(R2).*sin(TH);      % Generate X1.
> X2=sqrt(R2).*cos(TH);      % Generate X2.
> Y1=X1+sqrt(3)*X2           % Generate Y1.
> Y2=-X1+sqrt(3)*X2          % Generate Y2.
> plot(Y1,Y2,'+')            % Plot scattergram.
```

We plotted the Y_1 values vs. the Y_2 values for 1000 pairs of generated random variables in a scattergram as shown in Fig. 6.4. Good agreement with the elliptical symmetry of the desired jointly Gaussian pdf is observed.

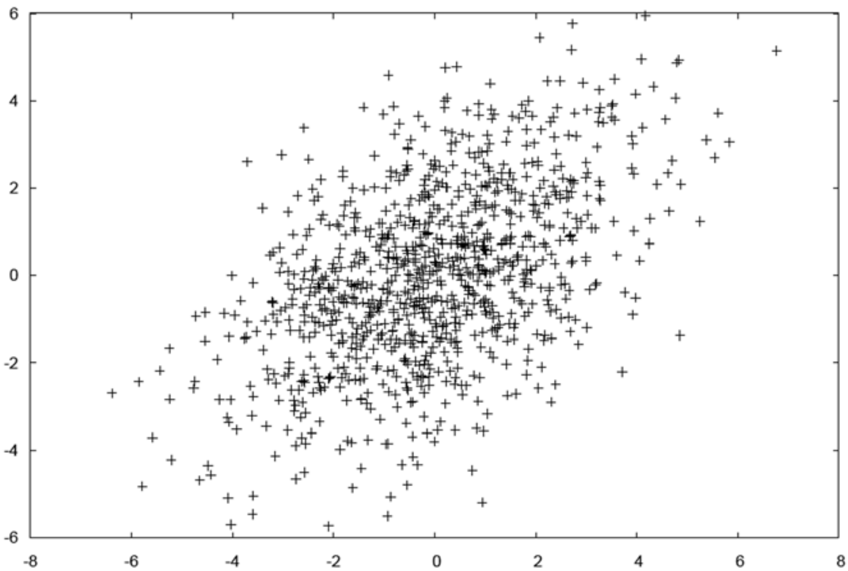

FIGURE 6.4
Scattergram of jointly Gaussian random variables.

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

ANNOTATED REFERENCES

Reference [3] provides excellent coverage on linear transformation and jointly Gaussian random variables. Reference [5] provides excellent coverage of vector random variables. The book by Anton [6] provides an accessible introduction to linear algebra.

1. A. Papoulis and S. Pillai, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, New York, 2002.
2. N. Johnson et al., *Continuous Multivariate Distributions*, Wiley, New York, 2000.
3. H. Cramer, *Mathematical Methods of Statistics*, Princeton Press, 1999.
4. R. Gray and L.D. Davisson, *An Introduction to Statistical Signal Processing*, Cambridge Univ. Press, Cambridge, UK, 2005.
5. H. Stark and J. W. Woods, *Probability, Random Processes, and Estimation Theory for Engineers*, Prentice Hall, Englewood Cliffs, N.J., 1986.
6. H. Anton, *Elementary Linear Algebra*, 9th ed., Wiley, New York, 2005.
7. C. H. Edwards, Jr., and D. E. Penney, *Calculus and Analytic Geometry*, 4th ed., Prentice Hall, Englewood Cliffs, N.J., 1984.

PROBLEMS

Section 6.1: Vector Random Variables

6.1. The point $\mathbf{X} = (X, Y, Z)$ is uniformly distributed inside a sphere of radius 1 about the origin. Find the probability of the following events:

(a) $\mathbf{X}$ is inside a sphere of radius r, $r > 0$.

(b) $\mathbf{X}$ is inside a cube of length $2/\sqrt{3}$ centered about the origin.

(c) All components of $\mathbf{X}$ are positive.

(d) Z is negative.

6.2. A random sinusoid signal is given by $X(t) = A\sin(t)$ where A is a uniform random variable in the interval $[0, 1]$. Let $\mathbf{X} = (X(t_1), X(t_2), X(t_3))$ be samples of the signal taken at times t_1, t_2, and t_3.

(a) Find the joint cdf of $\mathbf{X}$ in terms of the cdf of A if $t_1 = 0$, $t_2 = \pi/2$, and $t_3 = \pi$. Are $X(t_1), X(t_2), X(t_3)$ independent random variables?

(b) Find the joint cdf of $\mathbf{X}$ for t_1, $t_2 = t_1 + \pi/2$, and $t_3 = t_1 + \pi$. Let $t_1 = \pi/6$.

6.3. Let the random variables X, Y, and Z be independent random variables. Find the following probabilities in terms of $F_X(x)$, $F_Y(y)$, and $F_Z(z)$.

(a) $P[|X| < 5, Y < 4, Z^3 > 8]$.

(b) $P[X = 5, Y < 0, Z > 1]$.

(c) $P[\min(X, Y, Z) < 2]$.

(d) $P[\max(X, Y, Z) > 6]$.

6.4. A radio transmitter sends a signal $s > 0$ to a receiver using three paths. The signals that arrive at the receiver along each path are:

$$X_1 = s + N_1, X_2 = s + N_2, \text{ and } X_3 = s + N_3,$$

where N_1, N_2, and N_3 are independent Gaussian random variables with zero mean and unit variance.

(a) Find the joint pdf of $\mathbf{X} = (X_1, X_2, X_3)$. Are X_1, X_2, and X_3 independent random variables?

(b) Find the probability that the minimum of all three signals is positive.

(c) Find the probability that a majority of the signals are positive.

6.5. An urn contains one black ball and two white balls. Three balls are drawn from the urn. Let $I_k = 1$ if the outcome of the kth draw is the black ball and let $I_k = 0$ otherwise. Define the following three random variables:

$$X = I_1 + I_2 + I_3,$$
$$Y = \min\{I_1, I_2, I_3\},$$
$$Z = \max\{I_1, I_2, I_3\}.$$

(a) Specify the range of values of the triplet (X, Y, Z) if each ball is put back into the urn after each draw; find the joint pmf for (X, Y, Z).

(b) In part a, are X, Y, and Z independent? Are X and Y independent?

(c) Repeat part a if each ball is not put back into the urn after each draw.

6.6. Consider the packet switch in Example 6.1. Suppose that each input has one packet with probability p and no packets with probability $1 - p$. Packets are equally likely to be destined to each of the outputs. Let $X_{1,} X_2$ and X_3 be the number of packet arrivals destined for output 1, 2, and 3, respectively.

(a) Find the joint pmf of X_1, X_2, and X_3 Hint: Imagine that every input has a *packet* go to a fictional port 4 with probality $1 - p$.

(b) Find the joint pmf of X_1 and X_2.

(c) Find the pmf of X_2.

(d) Are X_1, X_2, and X_3 independent random variables?

(e) Suppose that each output will accept at most one packet and discard all additional packets destined to it. Find the average number of packets discarded by the module in each T-second period.

6.7. Let X, Y, Z have joint pdf

$$f_{X,Y,Z}(x, y, z) = k(x + y + z) \quad \text{for} \quad 0 \le x \le 1, 0 \le y \le 1, 0 \le z \le 1.$$

(a) Find k.

(b) Find $f_X(x \mid y, z)$ and $f_Z(z \mid x, y)$.

(c) Find $f_X(x), f_Y(y)$, and $f_Z(z)$.

6.8. A point $\mathbf{X} = (X, Y, Z)$ is selected at random inside the unit sphere.

(a) Find the marginal joint pdf of Y and Z.

(b) Find the marginal pdf of Y.

(c) Find the conditional joint pdf of X and Y given Z.

(d) Are X, Y, and Z independent random variables?

(e) Find the joint pdf of $\mathbf{X}$ given that the distance from $\mathbf{X}$ to the origin is greater than 1/2 and all the components of $\mathbf{X}$ are positive.

6.9. Show that $p_{X_1,X_2,X_3}(x_1, x_2, x_3) = p_{X_3}(x_3 \mid x_1, x_2)p_{X_2}(x_2 \mid x_1)p_{X_1}(x_1)$.

6.10. Let $X_1, X_2, \ldots, X_n$ be binary random variables taking on values 0 or 1 to denote whether a speaker is silent (0) or active (1). A silent speaker remains idle at the next time slot with probability 3/4, and an active speaker remains active with probability 1/2. Find the joint pmf for X_1, X_2, X_3, and the marginal pmf of X_3. Assume that the speaker begins in the silent state.

6.11. Show that $f_{X,Y,Z}(x, y, z) = f_Z(z \mid x, y)f_Y(y \mid x)f_X(x)$.

6.12. Let U_1, U_2, and U_3 be independent random variables and let $X = U_1, Y = U_1 + U_2$, and $Z = U_1 + U_2 + U_3$.

(a) Use the result in Problem 6.11 to find the joint pdf of X, Y, and Z.

(b) Let the U_i be independent uniform random variables in the interval [0, 1]. Find the marginal joint pdf of Y and Z. Find the marginal pdf of Z.

(c) Let the U_i be independent zero-mean, unit-variance Gaussian random variables. Find the marginal pdf of Y and Z. Find the marginal pdf of Z.

6.13. Let X_1, X_2, and X_3 be the multiplicative sequence in Example 6.7.

(a) Find, plot, and compare the marginal pdfs of X_1, X_2, and X_3.

(b) Find the conditional pdf of X_3 given $X_1 = x$.

(c) Find the conditional pdf of X_1 given $X_3 = z$.

6.14. Requests at an online music site are categorized as follows: Requests for most popular title with $p_1 = 1/2$; second most popular title with $p_2 = 1/4$; third most popular title with $p_3 = 1/8$; and other $p_4 = 1 - p_1 - p_2 - p_3 = 1/8$. Suppose there are a total number of

n requests in T seconds. Let X_k be the number of times category k occurs.

(a) Find the joint pmf of (X_1, X_2, X_3).

(b) Find the marginal pmf of (X_1, X_2). *Hint*: Use the binomial theorem.

(c) Find the marginal pmf of X_1.

(d) Find the conditional joint pmf of (X_2, X_3) given $X_1 = m$, where $0 \le m \le n$.

6.15. The number N of requests at the online music site in Problem 6.14 is a Poisson random variable with mean α customers per second. Let X_k be the number of type k requests in T seconds. Find the joint pmf of (X_1, X_2, X_3, X_4).

6.16. A random experiment has four possible outcomes. Suppose that the experiment is repeated n independent times and let X_k be the number of times outcome k occurs. The joint pmf of (X_1, X_2, X_3) is given by

$$p(k_1, k_2, k_3) = \frac{n!\,3!}{(n+3)!} = \binom{n+3}{3}^{-1} \quad \text{for} \quad 0 \le k_i \text{ and } k_1 + k_2 + k_3 \le n.$$

(a) Find the marginal pmf of (X_1, X_2).

(b) Find the marginal pmf of X_1.

(c) Find the conditional joint pmf of (X_2, X_3) given $X_1 = m$, where $0 \le m \le n$.

6.17. The number of requests of types 1, 2, and 3, respectively, arriving at a service station in t seconds are independent Poisson random variables with means $\lambda_1 t$, $\lambda_2 t$, and $\lambda_3 t$. Let N_1, N_2, and N_3 be the number of requests that arrive during an exponentially distributed time T with mean αt.

(a) Find the joint pmf of N_1, N_2, and N_3.

(b) Find the marginal pmf of N_1.

(c) Find the conditional pmf of N_1 and N_2, given N_3.

Section 6.2: Functions of Several Random Variables

6.18. N devices are installed at the same time. Let Y be the time until the first device fails.

(a) Find the pdf of Y if the lifetimes of the devices are independent and have the same Pareto distribution.

(b) Repeat part a if the device lifetimes have a Weibull distribution.

6.19. In Problem 6.18 let $I_k(t)$ be the indicator function for the event "kth device is still working at time t." Let $N(t)$ be the number of devices still working at time t: $N(t) = I_1(t) + I_2(t) + \cdots + I_N(t)$. Find the pmf of $N(t)$ as well as its mean and variance.

6.20. A diversity receiver receives N independent versions of a signal. Each signal version has an amplitude X_k that is Rayleigh distributed. The receiver selects that signal with the largest amplitude X_k^2. A signal is not useful if the squared amplitude falls below a threshold γ. Find the probability that all N signals are below the threshold.

6.21. (Haykin) A receiver in a multiuser communication system accepts K binary signals from K independent transmitters: $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_K)$, where Y_k is the received signal from the kth transmitter. In an ideal system the received vector is given by:

$$\mathbf{Y} = \mathbf{Ab} + \mathbf{N}$$

where $\mathbf{A} = [\alpha_k]$ is a diagonal matrix of positive channel gains, $\mathbf{b} = (b_1, b_2, \ldots, b_K)$ is the vector of bits from each of the transmitters where $b_k = \pm 1$, and $\mathbf{N}$ is a vector of $\mathbf{K}$

independent zero-mean, unit-variance Gaussian random variables.

(a) Find the joint pdf of $\mathbf{Y}$.

(b) Suppose $\mathbf{b} = (1, 1, \ldots, 1)$, find the probability that all components of $\mathbf{Y}$ are positive.

6.22. **(a)** Find the joint pdf of $U = X_1, V = X_1 + X_2$, and $W = X_1 + X_2 + X_3$.

(b) Evaluate the joint pdf of (U, V, W) if the X_i are independent zero-mean, unit variance Gaussian random variables.

(c) Find the marginal pdf of V and of W.

6.23. **(a)** Find the joint pdf of the sample mean and variance of two random variables:

$$M = \frac{X_1 + X_2}{2} \qquad V = \frac{(X_1 - M)^2 + (X_2 - M)^2}{2}$$

in terms of the joint pdf of X_1 and X_2.

(b) Evaluate the joint pdf if X_1 and X_2 are independent Gaussian random variables with the same mean 1 and variance 1.

(c) Evaluate the joint pdf if X_1 and X_2 are independent exponential random variables with the same parameter 1.

6.24. **(a)** Use the auxiliary variable method to find the pdf of

$$Z = \frac{X}{X + Y}.$$

(b) Find the pdf of Z if X and Y are independent exponential random variables with the parameter 1.

(c) Repeat part b if X and Y are independent Pareto random variables with parameters $k = 2$ and $x_m = 1$.

6.25. Repeat Problem 6.24 parts a and b for $Z = X/Y$.

6.26. Let X and Y be zero-mean, unit-variance Gaussian random variables with correlation coefficient 1/2. Find the joint pdf of $U = X^2$ and $V = Y^4$.

6.27. Use auxilliary variables to find the pdf of $Z = X_1X_2X_3$ where the X_i are independent random variables that are uniformly distributed in $[0, 1]$.

6.28. Let X, Y, and Z be independent zero-mean, unit-variance Gaussian random variables.

(a) Find the pdf of $R = (X^2 + Y^2 + Z^2)^{1/2}$.

(b) Find the pdf of $R^2 = X^2 + Y^2 + Z^2$.

6.29. Let X_1, X_2, X_3, X_4 be processed as follows:

$$Y_1 = X_1, Y_2 = X_1 + X_2, Y_3 = X_2 + X_3, Y_4 = X_3 + X_4.$$

(a) Find an expression for the joint pdf of $\mathbf{Y} = (Y_1, Y_2, Y_3, Y_4)$ in terms of the joint pdf of $\mathbf{X} = (X_1, X_2, X_3, X_4)$.

(b) Find the joint pdf of $\mathbf{Y}$ if X_1, X_2, X_3, X_4 are independent zero-mean, unit-variance Gaussian random variables.

Section 6.3: Expected Values of Vector Random Variables

6.30. Find $E[M]$, $E[V]$, and $E[MV]$ in Problem 6.23c.

6.31. Compute $E[Z]$ in Problem 6.27 in two ways:

(a) by integrating over $f_Z(z)$;

(b) by integrating over the joint pdf of (X_1, X_2, X_3).

6.32. Find the mean vector and covariance matrix for three multipath signals $\mathbf{X} = (X_1, X_2, X_3)$ in Problem 6.4.

6.33. Find the mean vector and covariance matrix for the samples of the sinusoidal signals $\mathbf{X} = (X(t_1), X(t_2), X(t_3))$ in Problem 6.2.

6.34. **(a)** Find the mean vector and covariance matrix for (X, Y, Z) in Problem 6.5a.

(b) Repeat part a for Problem 6.5c.

6.35. Find the mean vector and covariance matrix for (X, Y, Z) in Problem 6.7.

6.36. Find the mean vector and covariance matrix for the point (X, Y, Z) inside the unit sphere in Problem 6.8.

6.37. **(a)** Use the results of Problem 6.6c to find the mean vector for the packet arrivals X_1, X_2, and X_3 in Example 6.5.

(b) Use the results of Problem 6.6b to find the covariance matrix.

(c) Explain why X_1, X_2, and X_3 are correlated.

6.38. Find the mean vector and covariance matrix for the joint number of packet arrivals in a random time N_1, N_2, and N_3 in Problem 6.17. *Hint:* Use conditional expectation.

6.39. **(a)** Find the mean vector and covariance matrix (U, V, W) in terms of (X_1, X_2, X_3) in Problem 6.22b.

(b) Find the cross-covariance matrix between (U, V, W) and (X_1, X_2, X_3).

6.40. **(a)** Find the mean vector and covariance matrix of $\mathbf{Y} = (Y_1, Y_2, Y_3, Y_4)$ in terms of those of $\mathbf{X} = (X_1, X_2, X_3, X_4)$ in Problem 6.29.

(b) Find the cross-covariance matrix between $\mathbf{Y}$ and $\mathbf{X}$.

(c) Evaluate the mean vector, covariance, and cross-covariance matrices if X_1, X_2, X_3, X_4 are independent random variables.

(d) Generalize the results in part c to $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_{n-1}, Y_n)$.

6.41. Let $\mathbf{X} = (X_1, X_2, X_3, X_4)$ consist of equal mean, independent, unit-variance random variables. Find the mean vector, covariance, and cross-covariance matrices of $\mathbf{Y} = \mathbf{AX}$:

(a) $$\mathbf{A} = \begin{bmatrix} 1 & 1/2 & 1/4 & 1/8 \\ 0 & 1 & 1/2 & 1/4 \\ 0 & 0 & 1 & 1/2 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

(b) $$\mathbf{A} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}.$$

6.42. Let $W = aX + bY + c$, where X and Y are random variables.

(a) Find the characteristic function of W in terms of the joint characteristic function of X and Y.

(b) Find the characteristic function of W if X and Y are the random variables discussed in Example 6.19. Find the pdf of W.

6.43. **(a)** Find the joint characteristic function of the jointly Gaussian random variables X and Y introduced in Example 5.45. *Hint*: Consider X and Y as a transformation of the independent Gaussian random variables V and W.

(b) Find $E[X^2Y]$.

(c) Find the joint characteristic function of $X' = X + a$ and $Y' = Y + b$.

6.44. Let $X = aU + bV$ and $y = cU + dV$, where $|ad - bc| \neq 0$.

(a) Find the joint characteristic function of X and Y in terms of the joint characteristic function of U and V.

(b) Find an expression for $E[XY]$ in terms of joint moments of U and V.

6.45. Let X and Y be nonnegative, integer-valued random variables. The joint probability generating function is defined by

$$G_{X,Y}(z_1, z_2) = E[z_1^X z_2^Y] = \sum_{j=0}^{\infty}\sum_{k=0}^{\infty} z_1^j z_2^k P[X = j, Y = k].$$

(a) Find the joint pgf for two independent Poisson random variables with parameters α_1 and α_2.

(b) Find the joint pgf for two independent binomial random variables with parameters (n, p) and (m, p).

6.46. Suppose that X and Y have joint pgf

$$G_{X,Y}(z_1, z_2) = e^{\alpha_1(z_1-1)+\alpha_2(z_2-1)+\beta(z_1z_2-1)}.$$

(a) Use the marginal pgf's to show that X and Y are Poisson random variables.

(b) Find the pgf of $Z = X + Y$. Is Z a Poisson random variable?

6.47. Let X and Y be trinomial random variables with joint pmf

$$P[X = j, Y = k] = \frac{n!\, p_1^j p_2^k (1 - p_1 - p_2)^{n-j-k}}{j!\, k!(n - j - k)!} \quad \text{for} \quad 0 \le j, k \text{ and } j + k \le n.$$

(a) Find the joint pgf of X and Y.

(b) Find the correlation and covariance of X and Y.

6.48. Find the mean vector and covariance matrix for (X, Y) in Problem 6.46.

6.49. Find the mean vector and covariance matrix for (X, Y) in Problem 6.47.

6.50. Let $\mathbf{X} = (X_1, X_2)$ have covariance matrix:

$$\mathbf{K_X} = \begin{bmatrix} 1 & 1/4 \\ 1/4 & 1 \end{bmatrix}.$$

(a) Find the eigenvalues and eigenvectors of $\mathbf{K_X}$.

(b) Find the orthogonal matrix $\mathbf{P}$ that diagonalizes $\mathbf{K_X}$. Verify that $\mathbf{P}$ is orthogonal and that $\mathbf{P}^{\mathrm{T}}\mathbf{K_X}\mathbf{P} = \mathbf{\Lambda}$.

(c) Express $\mathbf{X}$ in terms of the eigenvectors of $\mathbf{K_X}$ using the Karhunen-Loeve expansion.

6.51. Repeat Problem 6.50 for $\mathbf{X} = (X_1, X_2, X_3)$ with covariance matrix:

$$\mathbf{K_X} = \begin{bmatrix} 1 & -1/2 & -1/2 \\ -1/2 & 1 & -1/2 \\ -1/2 & -1/2 & 1 \end{bmatrix}.$$

6.52. A square matrix $\mathbf{A}$ is said to be nonnegative definite if for any vector $\boldsymbol{a} = (a_1, a_2, \ldots, a_n)^{\mathrm{T}}$: $\boldsymbol{a}^{\mathrm{T}}\mathbf{A}\,\boldsymbol{a} \geq 0$. Show that the covariance matrix is nonnegative definite. *Hint*: Use the fact that $E[(\boldsymbol{a}^{\mathrm{T}}(\mathbf{X} - \mathbf{m_X}))^2] \geq 0$.

6.53. $\mathbf{A}$ is positive definite if for any nonzero vector $\boldsymbol{a} = (a_1, a_2, \ldots, a_n)^{\mathrm{T}}$: $\boldsymbol{a}^{\mathrm{T}}\mathbf{A}\,\boldsymbol{a} > 0$.

(a) Show that if all the eigenvalues are positive, then $\mathbf{K_X}$ is positive definite. *Hint*: Let $\mathbf{b} = \mathbf{P}^{\mathrm{T}}\boldsymbol{a}$.

(b) Show that if $\mathbf{K_X}$ is positive definite, then all the eigenvalues are positive. *Hint*: Let $\boldsymbol{a}$ be an eigenvector of $\mathbf{K_X}$.

Section 6.4: Jointly Gaussian Random Vectors

6.54. Let $\mathbf{X} = (X_1, X_2)$ be the jointly Gaussian random variables with mean vector and covariance matrix given by:

$$\mathbf{m_X} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \qquad \mathbf{K_X} = \begin{bmatrix} 3/2 & -1/2 \\ -1/2 & 3/2 \end{bmatrix}.$$

(a) Find the pdf of $\mathbf{X}$ in matrix notation.

(b) Find the pdf of $\mathbf{X}$ using the quadratic expression in the exponent.

(c) Find the marginal pdfs of X_1 and X_2.

(d) Find a transformation A such that the vector $\mathbf{Y} = A\mathbf{X}$ consists of independent Gaussian random variables.

(e) Find the joint pdf of $\mathbf{Y}$.

6.55. Let $\mathbf{X} = (X_1, X_2, X_3)$ be the jointly Gaussian random variables with mean vector and covariance matrix given by:

$$\mathbf{m_X} = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} \qquad \mathbf{K_X} = \begin{bmatrix} 3/2 & 0 & 1/2 \\ 0 & 1 & 0 \\ 1/2 & 0 & 3/2 \end{bmatrix}.$$

(a) Find the pdf of $\mathbf{X}$ in matrix notation.

(b) Find the pdf of $\mathbf{X}$ using the quadratic expression in the exponent.

(c) Find the marginal pdfs of X_1, X_2, and X_3.

(d) Find a transformation A such that the vector $\mathbf{Y} = A\mathbf{X}$ consists of independent Gaussian random variables.

(e) Find the joint pdf of $\mathbf{Y}$.

6.56. Let U_1, U_2, and U_3 be independent zero-mean, unit-variance Gaussian random variables and let $X = U_1$, $Y = U_1 + U_2$, and $Z = U_1 + U_2 + U_3$.

(a) Find the covariance matrix of (X, Y, Z).

(b) Find the joint pdf of (X, Y, Z).

(c) Find the conditional pdf of Y and Z given X.

(d) Find the conditional pdf of Z given X and Y.

6.57. Let X_1, X_2, X_3, X_4 be independent zero-mean, unit-variance Gaussian random variables that are processed as follows:

$$Y_1 = X_1 + X_2, Y_2 = X_2 + X_3, Y_3 = X_3 + X_4.$$

(a) Find the covariance matrix of $\mathbf{Y} = (Y_1, Y_2, Y_3)$.

(b) Find the joint pdf of $\mathbf{Y}$.

(c) Find the joint pdf of Y_1 and Y_2; Y_1 and Y_3.

(d) Find a transformation A such that the vector $\mathbf{Z} = A\mathbf{Y}$ consists of independent Gaussian random variables.

6.58. A more realistic model of the receiver in the multiuser communication system in Problem 6.21 has the K received signals $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_K)$ given by:

$$\mathbf{Y} = \mathbf{ARb} + \mathbf{N}$$

where $\mathbf{A} = [\alpha_k]$ is a diagonal matrix of positive channel gains, $\mathbf{R}$ is a symmetric matrix that accounts for the interference between users, and $\mathbf{b} = (b_1, b_2, \ldots, b_K)$ is the vector of bits from each of the transmitters. $\mathbf{N}$ is the vector of K independent zero-mean, unit-variance Gaussian noise random variables.

(a) Find the joint pdf of $\mathbf{Y}$.

(b) Suppose that in order to recover $\mathbf{b}$, the receiver computes $\mathbf{Z} = (\mathbf{AR})^{-1}\mathbf{Y}$. Find the joint pdf of $\mathbf{Z}$.

6.59. **(a)** Let $\mathbf{K_3}$ be the covariance matrix in Problem 6.55. Find the corresponding $\mathbf{Q_2}$ and $\mathbf{Q_3}$ in Example 6.23.

(b) Find the conditional pdf of X_3 given X_1 and X_2.

6.60. In Example 6.23, show that:

$$\tfrac{1}{2}(\mathbf{x}_n - \mathbf{m}_n)^{\mathrm{T}}\mathbf{Q}_n(\mathbf{x}_n - \mathbf{m}_n) - \tfrac{1}{2}(\mathbf{x}_{n-1} - \mathbf{m}_{n-1})^{\mathrm{T}}\mathbf{Q}_{n-1}(\mathbf{x}_{n-1} - \mathbf{m}_{n-1})$$
$$= Q_{nn}\{(x_n - m_n) + B\}^2 - Q_{nn}B^2$$
$$\text{where } B = \frac{1}{Q_{nn}}\sum_{j=1}^{n-1} Q_{jk}(x_j - m_j) \quad \text{and} \quad |\mathbf{K}_n|/|\mathbf{K}_{n-1}| = Q_{nn}.$$

6.61. Find the pdf of the sum of Gaussian random variables in the following cases:

(a) $Z = X_1 + X_2 + X_3$ in Problem 6.55.

(b) $Z = X + Y + Z$ in Problem 6.56.

(c) $Z = Y_1 + Y_2 + Y_3$ in Problem 6.57.

6.62. Find the joint characteristic function of the jointly Gaussian random vector $\mathbf{X}$ in Problem 6.54.

6.63. Suppose that a jointly Gaussian random vector $\mathbf{X}$ has zero mean vector and the covariance matrix given in Problem 6.51.

(a) Find the joint characteristic function.

(b) Can you obtain an expression for the joint pdf? Explain your answer.

6.64. Let X and Y be jointly Gaussian random variables. Derive the joint characteristic function for X and Y using conditional expectation.

6.65. Let $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ be jointly Gaussian random variables. Derive the characteristic function for $\mathbf{X}$ by carrying out the integral in Eq. (6.32). *Hint*: You will need to complete the square as follows:

$$(\mathbf{x} - j\mathbf{K}\omega)^{\mathrm{T}}\mathbf{K}^{-1}(\mathbf{x} - j\mathbf{K}\omega) = \mathbf{x}^{\mathrm{T}}\mathbf{K}^{-1}\mathbf{x} - 2j\mathbf{x}^{\mathrm{T}}\omega + j^2\omega^{\mathrm{T}}\mathbf{K}\omega.$$

6.66. Find $E[X^2Y^2]$ for jointly Gaussian random variables from the characteristic function.

6.67. Let $\mathbf{X} = (X_1, X_2, X_3, X_4)$ be zero-mean jointly Gaussian random variables. Show that $E[X_1X_2X_3X_4] = E[X_1X_2]E[X_3X_4] + E[X_1X_3]E[X_2X_4] + E[X_1X_4]E[X_2X_3]$.

Section 6.5: Mean Square Estimation

6.68. Let X and Y be discrete random variables with three possible joint pmf's:

(i)				(ii)				(iii)			
X/Y	-1	0	1	X/Y	-1	0	1	X/Y	-1	0	1
-1	1/6	1/6	0	-1	1/9	1/9	1/9	-1	1/3	0	0
0	0	0	1/3	0	1/9	1/9	1/9	0	0	1/3	0
1	1/6	1/6	0	1	1/9	1/9	1/9	1	0	0	1/3

(a) Find the minimum mean square error linear estimator for Y given X.
(b) Find the minimum mean square error estimator for Y given X.
(c) Find the MAP and ML estimators for Y given X.
(d) Compare the mean square error of the estimators in parts a, b, and c.

6.69. Repeat Problem 6.68 for the continuous random variables X and Y in Problem 5.26.

6.70. Find the ML estimator for the signal s in Problem 6.4.

6.71. Let N_1 be the number of Web page requests arriving at a server in the period (0, 100) ms and let N_2 be the *total* combined number of Web page requests arriving at a server in the period (0, 200) ms. Assume page requests occur every 1-ms interval according to independent Bernoulli trials with probability of success p.
(a) Find the minimum linear mean square estimator for N_2 given N_1 and the associated mean square error.
(b) Find the minimum mean square error estimator for N_2 given N_1 and the associated mean square error.
(c) Find the maximum a posteriori estimator for N_2 given N_1.
(d) Repeat parts a, b, and c for the estimation of N_1 given N_2.

6.72. Let $Y = X + N$ where X and N are independent Gaussian random variables with different variances and N is zero mean.
(a) Plot the correlation coefficient between the "observed signal" Y and the "desired signal" X as a function of the signal-to-noise ratio σ_X/σ_N.
(b) Find the minimum mean square error estimator for X given Y.
(c) Find the MAP and ML estimators for X given Y.
(d) Compare the mean square error of the estimators in parts a, b and c.

6.73. Let X, Y, Z be the random variables in Problem 6.7.
(a) Find the minimum mean square error linear estimator for Y given X and Z.
(b) Find the minimum mean square error estimator for Y given X and Z.
(c) Find the MAP and ML estimators for Y given X and Z.
(d) Compare the mean square error of the estimators in parts b and c.

6.74. **(a)** Repeat Problem 6.73 for the estimator of X_2, given X_1 and X_3 in Problem 6.13.
(b) Repeat Problem 6.73 for the estimator of X_3 given X_1 and X_2.

6.75. Consider the ideal multiuser communication system in Problem 6.21. Assume the transmitted bits b_k are independent and equally likely to be $+1$ or -1.
(a) Find the ML and MAP estimators for $\mathbf{b}$ given the observation $\mathbf{Y}$.
(b) Find the minimum mean square linear estimator for $\mathbf{b}$ given the observation $\mathbf{Y}$. How can this estimator be used in deciding what were the transmitted bits?

6.76. Repeat Problem 6.75 for the multiuser system in Problem 6.58.

6.77. A second-order predictor for samples of an image predicts the sample E as a linear function of sample D to its left and sample B in the previous line, as shown below:

line j	...	A	B	C ...	
line $j+1$	...	D	E		...

Estimate for $E = aD + bB$.

(a) Find a and b if all samples have variance σ^2 and if the correlation coefficient between D and E is ρ, between B and E is ρ, and between D and B is ρ^2.
(b) Find the mean square error of the predictor found in part a, and determine the reduction in the variance of the signal in going from the input to the output of the predictor.

6.78. Show that the mean square error of the two-tap linear predictor is given by Eq. (6.64).

6.79. In "hexagonal sampling" of an image, the samples in consecutive lines are offset relative to each other as shown below:

line j	$\dots$		A		B
line $j+1$	$\dots$	C		D	

The covariance between two samples a and b is given by $\rho^{d(a,b)}$ where $d(a, b)$ is the Euclidean distance between the points. In the above samples, the distance between A and B, A and C, A and D, C and D, and B and D is 1. Suppose we wish to use a two-tap linear predictor to predict the sample D. Which two samples from the set $\{A, B, C\}$ should we use in the predictor? What is the resulting mean square error?

*Section 6.6: Generating Correlated Vector Random Variables

6.80. Find a linear transformation that diagonalizes $\mathbf{K}$.

(a) $\mathbf{K} = \begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix}$.

(b) $\mathbf{K} = \begin{bmatrix} 4 & 1 \\ 1 & 4 \end{bmatrix}$.

6.81. Generate and plot the scattergram of 1000 pairs of random variables $\mathbf{Y}$ with the covariance matrices in Problem 6.80 if:

(a) X_1 and X_2 are independent random variables that are each uniform in the unit interval;

(b) X_1 and X_2 are independent zero-mean, unit-variance Gaussian random variables.

6.82. Let $\mathbf{X} = (X_1, X_2, X_3)$ be the jointly Gaussian random variables in Problem 6.55.

(a) Find a linear transformation that diagonalizes the covariance matrix.

(b) Generate 1000 triplets of $\mathbf{Y} = \mathbf{AX}$ and plot the scattergrams for Y_1 and Y_2, Y_1 and Y_3, and Y_2 and Y_3. Confirm that the scattergrams are what is expected.

6.83. Let $\mathbf{X}$ be a jointly Gaussian random vector with mean $\mathbf{m_X}$ and covariance matrix $\mathbf{K_X}$ and let $\mathbf{A}$ be a matrix that diagonalizes $\mathbf{K_X}$. What is the joint pdf of $\mathbf{A}^{-1}(\mathbf{X} - \mathbf{m_X})$?

6.84. Let $X_1, X_2, \dots, X_n$ be independent zero-mean, unit-variance Gaussian random variables. Let $Y_k = (X_k + X_{k-1})/2$, that is, Y_k is the moving average of pairs of values of X. Assume $X_{-1} = 0 = X_{n+1}$.

(a) Find the covariance matrix of the Y_k's.

(b) Use Octave to generate a sequence of 1000 samples $Y_1, \dots, Y_n$. How would you check whether the Y_k's have the correct covariances?

6.85. Repeat Problem 6.84 with $Y_k = X_k - X_{k-1}$.

6.86. Let $\mathbf{U}$ be an orthogonal matrix. Show that if $\mathbf{A}$ diagonalizes the covariance matrix $\mathbf{K}$, then $\mathbf{B} = \mathbf{UA}$ also diagonalizes $\mathbf{K}$.

6.87. The transformation in Problem 6.56 is said to be "causal" because each output depends only on "past" inputs.

(a) Find the covariance matrix of X, Y, Z in Problem 6.56.

(b) Find a noncausal transformation that diagonalizes the covariance matrix in part a.

6.88. **(a)** Find a causal transformation that diagonalizes the covariance matrix in Problem 6.54.

(b) Repeat for the covariance matrix in Problem 6.55.

Problems Requiring Cumulative Knowledge

6.89. Let $U_0, U_1, \ldots$ be a sequence of independent zero-mean, unit-variance Gaussian random variables. A "low-pass filter" takes the sequence U_i and produces the output sequence $X_n = (U_n + U_{n-1})/2$, and a "high-pass filter" produces the output sequence $Y_n = (U_n - U_{n-1})/2$.

(a) Find the joint pdf of X_{n+1}, X_n, and X_{n-1}; of $X_n, X_{n+m,}and\ X_{n+2m}, m > 1$.

(b) Repeat part a for Y_n.

(c) Find the joint pdf of X_n, X_m, Y_n, and Y_m.

(d) Find the corresponding joint characteristic functions in parts a, b, and c.

6.90. Let $X_1, X_2, \ldots, X_n$ be the samples of a speech waveform in Example 6.31. Suppose we want to interpolate for the value of a sample in terms of the previous and the next samples, that is, we wish to find the best linear estimate for X_2 in terms of X_1 and X_3.

(a) Find the coefficients of the best linear estimator (interpolator).

(b) Find the mean square error of the best linear interpolator and compare it to the mean square error of the two-tap predictor in Example 6.31.

(c) Suppose that the samples are jointly Gaussian. Find the pdf of the interpolation error.

6.91. Let $X_1, X_2, \ldots, X_n$ be samples from some signal. Suppose that the samples are jointly Gaussian random variables with covariance

$$\text{COV}(X_i, X_j) = \begin{cases} \sigma^2 & \text{for } i = j \\ \rho\sigma^2 & \text{for } |i - j| = 1 \\ 0 & \text{otherwise.} \end{cases}$$

Suppose we take blocks of two consecutive samples to form a vector $\mathbf{X}$, which is then linearly transformed to form $\mathbf{Y} = \mathbf{AX}$.

(a) Find the matrix $\mathbf{A}$ so that the components of $\mathbf{Y}$ are independent random variables.

(b) Let $\mathbf{X}_i$ and $\mathbf{X}_{i+1}$ be two consecutive blocks and let $\mathbf{Y}_i$ and $\mathbf{Y}_{i+1}$ be the corresponding transformed variables. Are the components of $\mathbf{Y}_i$ and $\mathbf{Y}_{i+1}$ independent?

6.92. A multiplexer combines N digital television signals into a common communications line. TV signal n generates X_n bits every 33 milliseconds, where X_n is a Gaussian random variable with mean m and variance σ^2. Suppose that the multiplexer accepts a maximum total of T bits from the combined sources every 33 ms, and that any bits in excess of T are discarded. Assume that the N signals are independent.

(a) Find the probability that bits are discarded in a given 33-ms period, if we let $T = m_a + t\sigma$, where m_a is the mean total bits generated by the combined sources, and σ is the standard deviation of the total number of bits produced by the combined sources.

(b) Find the average number of bits discarded per period.

(c) Find the long-term fraction of bits lost by the multiplexer.

(d) Find the average number of bits per source allocated in part a, and find the average number of bits lost per source. What happens as N becomes large?

(e) Suppose we require that t be adjusted with N so that the fraction of bits lost per source is kept constant. Find an equation whose solution yields the desired value of t.

(f) Do the above results change if the signals have pairwise covariance ρ?

6.93. Consider the estimation of T given N_1 and arrivals in Problem 6.17.

(a) Find the ML and MAP estimators for T.

(b) Find the linear mean square estimator for T.

(c) Repeat parts a and b if N_1 and N_2 are given.