Chapter 9

Discrete N-Dimensional Random Variables

9.1 Introduction

In this chapter we extend the results of Chapters 5–8 to N-dimensional random variables, which are represented as an $N \times 1$ random vector. Hence, our discussions will apply to the 2×1 random vector previously studied. In fact, most of the concepts introduced earlier are trivially extended so that we do not dwell on the conceptualization. The only exception is the introduction of the covariance *matrix*, which we have not seen before. We will introduce more general notation in combination with vector/matrix representations to allow the convenient manipulation of $N \times 1$ random vectors. This representation allows many results to be easily derived and is useful for the more advanced theory of probability that the reader may encounter later. Also, it lends itself to straightforward computer implementations, particularly if one uses MATLAB, which is a vector-based programming language. Since many of the methods and subsequent properties rely on linear and matrix algebra, a brief summary of relevant concepts is given in Appendix C.

9.2 Summary

The N-dimensional joint PMF is given by (9.1) and satisfies the usual properties of (9.3) and (9.4). The joint PMF of any subset of the N random variables is obtained by summing the joint PMF over the undesired ones. If the joint PMF factors as in (9.7), the random variables are independent and vice versa. The joint PMF of a transformed random vector is given by (9.9). In particular, if the transformed random variable is the sum of N independent random variables with the same PMF, then the PMF is most easily found from (9.14). The expected value of a random vector is defined by (9.15) and the expected value of a scalar function of a random

vector is found via (9.16). As usual, the expectation operator is linear with a special case given by (9.17). The variance of a sum of N random variables is given by (9.20) or (9.21). If the random variables are uncorrelated, then this variance is the sum of the variances as per (9.22). The covariance matrix of a random vector is defined by (9.25). It has many important properties that are summarized in Properties 9.1–5. Particularly useful results are the covariance matrix of a linearly transformed random vector given by (9.27) and the ability to decorrelate the elements of a random vector using a linear transformation as explained in the proof of Property 9.5. An example of this procedure is given in Example 9.4. The joint moments and characteristic function of an N-dimensional PMF are defined by (9.32) and (9.34), respectively. The joint moments are obtainable from the characteristic function by using (9.36). An important relationship is the factorization of the joint PMF into a product of conditional PMFs as given by (9.39). When the random variables exhibit the Markov property, then this factorization simplifies even further into the product of first-order conditional PMFs as given by (9.41). The estimates of the mean vector and the covariance matrix of a random vector are given by (9.44) and (9.46), respectively. Some MATLAB code for implementing these estimates is listed in Section 9.8. Finally, a real-world example of the use of transform coding to store/transmit image data is described in Section 9.9. It is based on decorrelation of random vectors and so makes direct use of the properties of the covariance matrix.

9.3 Random Vectors and Probability Mass Functions

Previously, we denoted a two-dimensional random vector by either of the equivalent notations (X, Y) or $[X\,Y]^T$. Since we now wish to extend our results to an $N \times 1$ random vector, we shall use $(X_1, X_2, \ldots, X_N)$ or $\mathbf{X} = [X_1\, X_2 \ldots X_N]^T$. Note that a boldface character will always denote a vector or a matrix, in contrast to a scalar variable. Also, all vectors are assumed to be *column* vectors. A random vector is defined as a mapping from the original sample space $\mathcal{S}$ of the experiment to a numerical sample space, which we term $\mathcal{S}_{X_1,X_2,\ldots,X_N}$. The latter is normally referred to as R^N, which is the N-dimensional Euclidean space. Hence, $\mathbf{X}$ takes on values in R^N so that

$$\mathbf{X}(s) = \begin{bmatrix} X_1(s) \\ X_2(s) \\ \vdots \\ X_N(s) \end{bmatrix}$$

will have values

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix}$$

where $\mathbf{x}$ is a point in the N-dimensional Euclidean space R^N. A simple example is $\mathcal{S} = \{\text{all lottery tickets}\}$ with $\mathbf{X}(s)$ representing the number printed on the ticket. Then, $X_1(s)$ is the first digit of the number, $X_2(s)$ is the second digit of the number, ..., and $X_N(s)$ is the Nth digit of the number.

We are, as usual, interested in the probability that $\mathbf{X}$ takes on its possible values. This probability is $P[X_1 = x_1, X_2 = x_2, \ldots, X_N = x_N]$ and it is defined as the joint PMF. The joint PMF is therefore defined as

$$p_{X_1,X_2,\ldots,X_N}[x_1, x_2, \ldots, x_N] = P[X_1 = x_1, X_2 = x_2, \ldots, X_N = x_N] \tag{9.1}$$

or more succinctly using vector notation as

$$p_{\mathbf{X}}[\mathbf{x}] = P[\mathbf{X} = \mathbf{x}]. \tag{9.2}$$

When $\mathbf{x}$ consists of integer values only, we will replace x_i by k_i. Then, the joint PMF will be $p_{X_1,X_2,\ldots,X_N}[k_1, k_2, \ldots, k_N]$ or more succintly as $p_{\mathbf{x}}[\mathbf{k}]$, where $\mathbf{k} = [k_1\, k_2 \ldots k_N]^T$. An example of an N-dimensional joint PMF, which is of considerable importance, is the multinomial PMF (see (4.19)). In our new notation the joint PMF is

$$p_{X_1,X_2,\ldots,X_N}[k_1, k_2, \ldots, k_N] = \binom{M}{k_1, k_2, \ldots, k_N} p_1^{k_1} p_2^{k_2} \ldots p_N^{k_N}$$

where $k_i \geq 0$ with $\sum_{i=1}^N k_i = M$, and $0 \leq p_i \leq 1$ for all i with $\sum_{i=1}^N p_i = 1$. That this is a valid joint PMF follows from its adherence to the usual properties

$$0 \leq p_{X_1,X_2,\ldots,X_N}[k_1, k_2, \ldots, k_N] \leq 1 \tag{9.3}$$

$$\sum_{k_1}\sum_{k_2}\cdots\sum_{k_N} p_{X_1,X_2,\ldots,X_N}[k_1, k_2, \ldots, k_N] = 1. \tag{9.4}$$

To prove (9.4) we need only use the multinomial expansion, which is (see Problem 9.3)

$$(a_1 + a_2 + \cdots + a_N)^M = \sum_{k_1}\sum_{k_2}\cdots\sum_{k_N} \binom{M}{k_1, k_2, \ldots, k_N} a_1^{k_1} a_2^{k_2} \ldots a_N^{k_N} \tag{9.5}$$

where $\sum_{i=1}^N k_i = M$.

The marginal PMFs are obtained from the joint PMF by summing over the *other variables*. For example, if $p_{X_1}[x_1]$ is desired, then

$$p_{X_1}[x_1] = \sum_{\{x_2 : x_2 \in \mathcal{S}_{X_2}\}} \sum_{\{x_3 : x_3 \in \mathcal{S}_{X_3}\}} \cdots \sum_{\{x_N : x_N \in \mathcal{S}_{X_N}\}} p_{X_1,X_2,\ldots,X_N}[x_1, x_2, \ldots, x_N] \tag{9.6}$$

and similarly for the other $N - 1$ marginals. This is because the right-hand side of (9.6) is

$$P[X_1 = x_1, X_2 \in \mathcal{S}_{X_2}, X_3 \in \mathcal{S}_{X_3}, \ldots, X_N \in \mathcal{S}_{X_N}] = P[X_1 = x_1].$$

When the random vector is composed of more than two random variables, we can also obtain the joint PMF of any subset of the random variables. We do this by summing over the variables that we wish to eliminate. If, say, we wish to determine the joint PMF of X_1 and X_N, we have

$$p_{X_1,X_N}[x_1,x_N] = \sum_{x_2}\sum_{x_3}\cdots\sum_{x_{N-1}} p_{X_1,X_2,\ldots,X_N}[x_1,x_2,\ldots,x_N].$$

As in the case of $N = 2$ the marginal PMFs do not determine the joint PMF, unless of course the random variables are independent. In the N-dimensional case the random variables are defined to be independent if the joint PMF factors or if

$$p_{X_1,X_2,\ldots,X_N}[x_1,x_2,\ldots,x_N] = p_{X_1}[x_1]p_{X_2}[x_2]\ldots p_{X_N}[x_N]. \tag{9.7}$$

Hence, if (9.7) holds, the random variables are independent, and if the random variables are independent (9.7) holds. Unlike the case of $N = 2$, it is possible that the joint PMF may factor into two or more joint PMFs. Then, the subsets of random variables are said to be independent of each other. For example, if $N = 4$ and the joint PMF factors as $p_{X_1,X_2,X_3,X_4}[x_1,x_2,x_3,x_4] = p_{X_1,X_2}[x_1,x_2]p_{X_3,X_4}[x_3,x_4]$, then the random variables (X_1,X_2) are independent of the random variables (X_3,X_4). An example of the determination of a joint PMF follows.

Example 9.1 – Joint PMF for independent Bernoulli trials

Consider an experiment in which we toss a coin with a probability of heads p, N times in succession. We let $X_i = 1$ if the ith outcome is a head and $X_i = 0$ if it is a tail. Furthermore, *assume* that the trials are *independent.* As defined in Chapter 4, this means that the probability of the outcome on any trial is not affected by the outcomes of any of the other trials. Thus, the experiment is a sequence of independent Bernoulli trials. The sample space is N-dimensional and is given by $\mathcal{S}_{X_1,X_2,\ldots,X_N} = \{(k_1,k_2,\ldots,k_N) : k_i = 0,1\ ; i = 1,2,\ldots,N\}$, and since $p_{X_i}[k_i] = p^{k_i}(1-p)^{1-k_i}$, we have the joint PMF from (9.7)

$$\begin{aligned} p_{X_1,X_2,\ldots,X_N}[k_1,k_2,\ldots,k_N] &= \prod_{i=1}^{N} p_{X_i}[k_i] \\ &= \prod_{i=1}^{N} p^{k_i}(1-p)^{1-k_i} \\ &= p^{\sum_{i=1}^{N} k_i}(1-p)^{N-\sum_{i=1}^{N} k_i}. \end{aligned} \tag{9.8}$$

$\Diamond$

A joint cumulative distribution function (CDF) can be defined in the N-dimensional case as

$$F_{X_1,X_2,\ldots,X_N}(x_1,x_2,\ldots,x_N) = P[X_1 \le x_1, X_2 \le x_2, \ldots, X_N \le x_N].$$

It has the usual properties of being between 0 and 1, being monotonically increasing as any of the variables increases, and being "right continuous". Also,

$$\begin{aligned} F_{X_1,X_2,\ldots,X_N}(-\infty,-\infty,\ldots,-\infty) &= 0 \\ F_{X_1,X_2,\ldots,X_N}(+\infty,+\infty,\ldots,+\infty) &= 1. \end{aligned}$$

The marginal CDFs are easily found by letting the undesired variables be evaluated at $+\infty$. For example, to determine the marginal CDF for X_1, we have

$$F_{X_1}[x_1] = F_{X_1,X_2,\ldots,X_N}(x_1,+\infty,+\infty,\ldots,+\infty).$$

9.4 Transformations

Since $\mathbf{X}$ is an $N \times 1$ random vector, a transformation or mapping to a random vector $\mathbf{Y}$ can yield another $N \times 1$ random vector or an $M \times 1$ random vector with $M < N$. In the former case the formula for the joint PMF of $\mathbf{Y}$ is an extension of the usual one (see (7.12)). If the transformation is given as $\mathbf{y} = \mathbf{g}(\mathbf{x})$, where $\mathbf{g}$ represents an N-dimensional function or more explicitly

$$\begin{aligned} y_1 &= g_1(x_1,x_2,\ldots,x_N) \\ y_2 &= g_2(x_1,x_2,\ldots,x_N) \\ &\vdots \\ y_N &= g_N(x_1,x_2,\ldots,x_N) \end{aligned}$$

then

$$p_{Y_1,Y_2,\ldots,Y_N}[y_1,y_2,\ldots,y_N] = \sum_{\substack{\{(x_1,\ldots,x_N):\\ g_1(x_1,\ldots,x_N)=y_1,\ldots,\\ g_N(x_1,\ldots,x_N)=y_N\}}} \sum \cdots \sum p_{X_1,X_2,\ldots,X_N}[x_1,x_2,\ldots,x_N]. \quad (9.9)$$

In the case where the transformation is one-to-one, there is only one solution for $\mathbf{x}$ in the equation $\mathbf{y} = \mathbf{g}(\mathbf{x})$, which we denote symbolically by $\mathbf{x} = \mathbf{g}^{-1}(\mathbf{y})$. The transformed joint PMF becomes from (9.9) $p_{\mathbf{Y}}[\mathbf{y}] = p_{\mathbf{X}}[\mathbf{g}^{-1}(\mathbf{y})]$, using vector notation. A simple example of this is when the transformation is linear and so can be represented by $\mathbf{y} = \mathbf{A}\mathbf{x}$, where $\mathbf{A}$ is an $N \times N$ nonsingular matrix. Then, the solution is $\mathbf{x} = \mathbf{A}^{-1}\mathbf{y}$ and the transformed joint PMF becomes

$$p_{\mathbf{Y}}[\mathbf{y}] = p_{\mathbf{X}}[\mathbf{A}^{-1}\mathbf{y}]. \quad (9.10)$$

The other case, in which $\mathbf{Y}$ has dimension less than N, can be solved using the technique of auxiliary random variables. We add enough random variables to make the dimension of the transformed random vector equal to N, find the joint PMF via

(9.9), and finally sum the N-dimensional PMF over the auxiliary random variables. More specifically, if $\mathbf{Y}$ is $M \times 1$ with $M < N$, we define a new $N \times 1$ random vector

$$\mathbf{Z} = [Y_1\ Y_2 \ldots Y_M\ Z_{M+1} = X_{M+1}\ Z_{M+2} = X_{M+2} \ldots Z_N = X_N]^T$$

so that the transformation becomes one-to-one, if possible. Once the joint PMF of $\mathbf{Z}$ is found, we can determine the joint PMF of $\mathbf{Y}$ as

$$p_{Y_1,Y_2,\ldots,Y_M}[y_1, y_2, \ldots, y_M] = \sum_{z_{M+1}} \sum_{z_{M+2}} \cdots \sum_{z_N} p_{Z_1,Z_2,\ldots,Z_N}[z_1, z_2, \ldots, z_N].$$

The determination of the PMF of a transformed random vector is in general *not an easy task*. Even to determine the possible values of $\mathbf{Y}$ can be quite difficult. An example follows that illustrates the work involved.

Example 9.2 – PMF for one-to-one transformation of N-dimensional random vector

In Example 9.1 $\mathbf{X}$ has the joint PMF given by (9.8). We define a transformed random vector as

$$\begin{aligned} Y_1 &= X_1 \\ Y_2 &= X_1 + X_2 \\ Y_3 &= X_1 + X_2 + X_3. \end{aligned}$$

This is a linear transformation that maps a 3×1 random vector $\mathbf{X}$ into another 3×1 random vector $\mathbf{Y}$. It can be represented by the 3×3 matrix

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}.$$

Note that the transformed random variables are the sums of the outcomes of the first Bernoulli trial, the first and second Bernoulli trials, and finally the sum of the first three Bernoulli trials. As such the values of the transformed random variables must take on certain values. In particular, $y_1 \leq y_2 \leq y_3$ or the outcomes must increase as the index i increases. This is sometimes called a *counting process* and will be studied in more detail when we discuss random processes. Some typical realizations of the random vector $\mathbf{Y}$ are shown in Figure 9.1. To determine the sample space for $\mathbf{Y}$ we enumerate the possible values, making sure that the values in the vector increase or stay the same and that the increase is at most one unit from y_i to y_{i+1}. The sample space is composed of integer 3-tuples (l_1, l_2, l_3), which is given by

$$\mathcal{S}_{Y_1,Y_2,Y_3} = \{(0,0,0), (0,0,1), (0,1,1), (1,1,1), (0,1,2), (1,1,2), (1,2,2), (1,2,3)\}. \tag{9.11}$$

These are the values of $\mathbf{y}$ for which p_{Y_1,Y_2,Y_3} is nonzero and are seen to be integer-valued. Next, we need to solve for $\mathbf{x}$ according to (9.10). It is easily shown that the

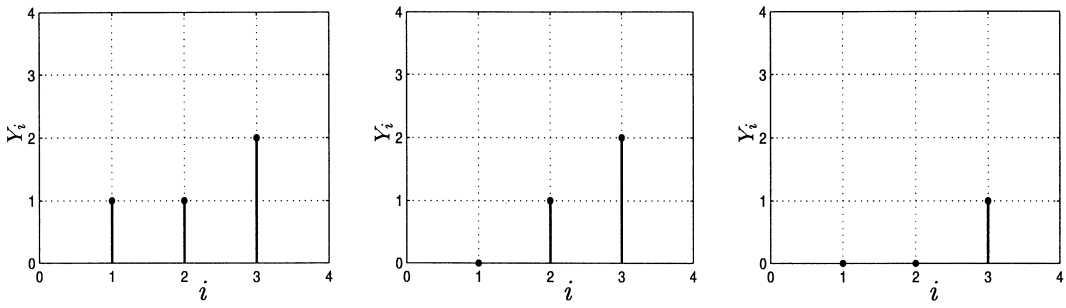


Figure 9.1: Typical realizations for sum of outcomes of independent Bernoulli trials.

linear transformation is one-to-one since $\mathbf{A}$ has an inverse (note that the determinant of $\mathbf{A}$ is nonzero since $\det(\mathbf{A}) = 1$, and so $\mathbf{A}$ has an inverse), which is

$$\mathbf{A}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix}.$$

This says that $\mathbf{x} = \mathbf{A}^{-1}\mathbf{y}$ or $x_1 = y_1, x_2 = y_2 - y_1, x_3 = y_3 - y_2$. Thus, we can use (9.10) and then (9.8) to find the joint PMF of $\mathbf{Y}$, which becomes from (9.10)

$$p_{Y_1,Y_2,Y_3}[l_1, l_2, l_3] = p_{X_1,X_2,X_3}[l_1, l_2 - l_1, l_3 - l_2]$$

and since from (9.8)

$$p_{X_1,X_2,X_3}[k_1, k_2, k_3] = p^{k_1+k_2+k_3}(1-p)^{3-(k_1+k_2+k_3)}$$

we have that

$$p_{Y_1,Y_2,Y_3}[l_1, l_2, l_3] = p^{l_3}(1-p)^{3-l_3}. \tag{9.12}$$

Note that the joint PMF is nonzero only over the sample space $\mathcal{S}_{Y_1,Y_2,Y_3}$ given in (9.11).

$\diamondsuit$

Always make sure PMF values sum to one.

The result of the previous example looks strange in that the joint PMF of $\mathbf{Y}$ does not depend on l_1 and l_2. A simple check that should always be made when working these types of problems is to verify that the PMF values sum to one. If not, then

there is an error in the calculation. If they do sum to one, then there could still be an error but it is not likely. For the previous example, we have from (9.11) 1 outcome for which $l_3 = 0$, 3 outcomes for which $l_3 = 1$, 3 outcomes for which $l_3 = 2$, and 1 outcome for which $l_3 = 3$. If we sum the probabilities of these outcomes we have from (9.12)

$$1(1-p)^3 + 3p(1-p)^2 + 3p^2(1-p) + p^3 = 1$$

and hence we can assert with some confidence that the result is correct.

A transformation that is not one-to-one but that frequently is of interest is the sum of N independent discrete random variables. It is given by

$$Y = \sum_{i=1}^{N} X_i \tag{9.13}$$

where the X_i's are independent random variables with integer values. For the case of $N = 2$ and integer-valued discrete random variables we saw in Section 7.6 that $p_Y = p_{X_1} \star p_{X_2}$, where $\star$ denotes discrete convolution. This is most easily evaluated using the characteristic functions and the inverse Fourier transform to yield

$$p_Y[k] = \int_{-\pi}^{\pi} \phi_{X_1}(\omega)\phi_{X_2}(\omega)\exp(-j\omega k)\frac{d\omega}{2\pi}.$$

For a sum of N independent random variables we have the similar result

$$p_Y[k] = \int_{-\pi}^{\pi} \prod_{i=1}^{N} \phi_{X_i}(\omega)\exp(-j\omega k)\frac{d\omega}{2\pi}$$

and if all the X_i's have the same PMF and hence the same characteristic function, this becomes

$$p_Y[k] = \int_{-\pi}^{\pi} \phi_X^N(\omega)\exp(-j\omega k)\frac{d\omega}{2\pi} \tag{9.14}$$

where $\phi_X(\omega)$ is the common characteristic function. An example follows (see also Problem 9.9).

Example 9.3 – Binomial PMF derived as PMF of sum of independent Bernoulli random variables

We had previously derived the binomial PMF by examining the number of successes in N independent Bernoulli trials (see Section 4.6.2). We can rederive this result by using (9.14) with $X_i = 1$ for a success and $X_i = 0$ for a failure and determining the

PMF of $Y = \sum_{i=1}^{N} X_i$. The random variable Y will be the number of successes in N trials. The characteristic function of X is for a single Bernoulli trial

$$\begin{aligned}\phi_X(\omega) &= E_X[\exp(j\omega X)] \\ &= \exp(j\omega(1))p + \exp(j\omega(0))(1-p) \\ &= p\exp(j\omega) + (1-p).\end{aligned}$$

Now using (9.14) we have

$$\begin{aligned}p_Y[k] &= \int_{-\pi}^{\pi} [p\exp(j\omega) + (1-p)]^N \exp(-j\omega k)\frac{d\omega}{2\pi} \\ &= \int_{-\pi}^{\pi} \sum_{i=0}^{N} \binom{N}{i} [p\exp(j\omega)]^i (1-p)^{N-i} \exp(-j\omega k)\frac{d\omega}{2\pi} \\ &\qquad\qquad \text{(use binomial theorem)} \\ &= \sum_{i=0}^{N} \binom{N}{i} p^i (1-p)^{N-i} \int_{-\pi}^{\pi} \exp[j\omega(i-k)]\frac{d\omega}{2\pi}.\end{aligned}$$

But the integral can be shown to be 0 if $i \neq k$ and 1 if $i = k$ (see Problem 9.8). Using this result we have as the only term in the sum being nonzero the one for which $i = k$, and therefore

$$p_Y[k] = \binom{N}{k} p^k (1-p)^{N-k} \qquad k = 0, 1, \ldots, N.$$

The sum of N independent Bernoulli random variables has the PMF bin(N, p) in accordance with our earlier results.

♢

9.5 Expected Values

The expected value of a random vector is defined as the vector of the expected values of the elements of the random vector. This is to say that we define

$$E_{\mathbf{X}}[\mathbf{X}] = E_{X_1,X_2,\ldots,X_N}\left[\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_N \end{bmatrix}\right] = \begin{bmatrix} E_{X_1}[X_1] \\ E_{X_2}[X_2] \\ \vdots \\ E_{X_N}[X_N] \end{bmatrix}. \tag{9.15}$$

We can view this definition as "passing" the expectation "through" the left bracket of the vector since $E_{X_1,X_2,\ldots,X_N}[X_i] = E_{X_i}[X_i]$.

A particular expectation of interest is that of a scalar function of $X_1, X_2, \ldots, X_N$, say $g(X_1, X_2, \ldots, X_N)$. Similar to previous results (see Section 7.7) this is determined by using

$$E_{X_1,X_2,\ldots,X_N}[g(X_1,X_2,\ldots,X_N)]$$
$$= \sum_{x_1}\sum_{x_2}\cdots\sum_{x_N} g(x_1,x_2,\ldots,x_N)p_{X_1,X_2,\ldots,X_N}[x_1,x_2,\ldots,x_N]. \tag{9.16}$$

As an example, if $g(X_1,X_2,\ldots,X_N)=\sum_{i=1}^N X_i$, then

$$E_{X_1,X_2,\ldots,X_N}\left[\sum_{i=1}^N X_i\right]$$

$$\begin{aligned}
&= \sum_{x_1}\sum_{x_2}\cdots\sum_{x_N}(x_1+x_2+\cdots+x_N)p_{X_1,X_2,\ldots,X_N}[x_1,x_2,\ldots,x_N]\\
&= \sum_{x_1}\sum_{x_2}\cdots\sum_{x_N} x_1 p_{X_1,X_2,\ldots,X_N}[x_1,x_2,\ldots,x_N]\\
&\quad+\sum_{x_1}\sum_{x_2}\cdots\sum_{x_N} x_2 p_{X_1,X_2,\ldots,X_N}[x_1,x_2,\ldots,x_N]\\
&\quad+\cdots+\sum_{x_1}\sum_{x_2}\cdots\sum_{x_N} x_N p_{X_1,X_2,\ldots,X_N}[x_1,x_2,\ldots,x_N]\\
&= E_{X_1}[X_1]+E_{X_2}[X_2]+\cdots+E_{X_N}[X_N].
\end{aligned}$$

By a slight modification we can also show that

$$E_{X_1,X_2,\ldots,X_N}\left[\sum_{i=1}^N a_i X_i\right] = \sum_{i=1}^N a_i E_{X_i}[X_i] \tag{9.17}$$

which says that the expectation is a linear operator. It is also possible to write (9.17) more succinctly by defining the $N\times 1$ vector $\mathbf{a}=[a_1\,a_2\ldots a_N]^T$ to yield

$$E_{\mathbf{X}}[\mathbf{a}^T\mathbf{X}] = \mathbf{a}^T E_{\mathbf{X}}[\mathbf{X}]. \tag{9.18}$$

We next determine the variance of a sum of random variables. Previously it was shown that

$$\mathrm{var}(X_1+X_2) = \mathrm{var}(X_1)+\mathrm{var}(X_2)+2\mathrm{cov}(X_1,X_2). \tag{9.19}$$

Our goal is to extend this to $\mathrm{var}(\sum_{i=1}^N X_i)$ for any N. To do so we proceed as follows.

$$\begin{aligned}
\mathrm{var}\left(\sum_{i=1}^N X_i\right) &= E_{\mathbf{X}}\left[\left(\sum_{i=1}^N X_i - E_{\mathbf{X}}\left[\sum_{i=1}^N X_i\right]\right)^2\right]\\
&= E_{\mathbf{X}}\left[\left(\sum_{i=1}^N (X_i - E_{X_i}[X_i])\right)^2\right] \qquad (\text{since } E_{\mathbf{X}}[X_i]=E_{X_i}[X_i])
\end{aligned}$$

and by letting $U_i = X_i - E_{X_i}[X_i]$ we have

$$\begin{aligned}\operatorname{var}\left(\sum_{i=1}^{N} X_i\right) &= E_{\mathbf{X}}\left[\left(\sum_{i=1}^{N} U_i\right)^2\right] \\ &= E_{\mathbf{X}}\left[\sum_{i=1}^{N}\sum_{j=1}^{N} U_i U_j\right] \\ &= \sum_{i=1}^{N}\sum_{j=1}^{N} E_{\mathbf{X}}[U_i U_j].\end{aligned}$$

But

$$\begin{aligned}E_{\mathbf{X}}[U_i U_j] &= E_{\mathbf{X}}[(X_i - E_{X_i}[X_i])(X_j - E_{X_j}[X_j])] \\ &= E_{X_i X_j}[(X_i - E_{X_i}[X_i])(X_j - E_{X_j}[X_j])] \\ &= \operatorname{cov}(X_i, X_j)\end{aligned}$$

so that we have as our final result

$$\operatorname{var}\left(\sum_{i=1}^{N} X_i\right) = \sum_{i=1}^{N}\sum_{j=1}^{N} \operatorname{cov}(X_i, X_j). \tag{9.20}$$

Noting that since $\operatorname{cov}(X_i, X_i) = \operatorname{var}(X_i)$ and $\operatorname{cov}(X_j, X_i) = \operatorname{cov}(X_i, X_j)$, we have for $N = 2$ our previous result (9.19). Also, we can write (9.20) in the alternative form

$$\operatorname{var}\left(\sum_{i=1}^{N} X_i\right) = \sum_{i=1}^{N} \operatorname{var}(X_i) + \sum_{\substack{i=1 \\ \{(i,j):i\neq j\}}}^{N}\sum_{j=1}^{N} \operatorname{cov}(X_i, X_j). \tag{9.21}$$

As an immediate and important consequence, we see that if all the random variables are uncorrelated so that $\operatorname{cov}(X_i, X_j) = 0$ for $i \neq j$, then

$$\operatorname{var}\left(\sum_{i=1}^{N} X_i\right) = \sum_{i=1}^{N} \operatorname{var}(X_i) \tag{9.22}$$

which says that *the variance of a sum of uncorrelated random variables is the sum of the variances.*

We wish to explore (9.20) further since it embodies some important concepts that we have not yet touched upon. For clarity let $N = 2$. Then (9.20) becomes

$$\operatorname{var}(X_1 + X_2) = \sum_{i=1}^{2}\sum_{j=1}^{2} \operatorname{cov}(X_i, X_j). \tag{9.23}$$

If we define a 2×2 matrix $\mathbf{C}_X$ as

$$\mathbf{C}_X = \begin{bmatrix} \text{var}(X_1) & \text{cov}(X_1, X_2) \\ \text{cov}(X_2, X_1) & \text{var}(X_2) \end{bmatrix}$$

then we can rewrite (9.23) as

$$\text{var}(X_1 + X_2) = \begin{bmatrix} 1 & 1 \end{bmatrix} \mathbf{C}_X \begin{bmatrix} 1 \\ 1 \end{bmatrix} \tag{9.24}$$

as is easily verified. The matrix $\mathbf{C}_X$ is called the *covariance matrix*. It is a matrix with the variances along the main diagonal and the covariances off the main diagonal. For $N = 3$ it is given by

$$\mathbf{C}_X = \begin{bmatrix} \text{var}(X_1) & \text{cov}(X_1, X_2) & \text{cov}(X_1, X_3) \\ \text{cov}(X_2, X_1) & \text{var}(X_2) & \text{cov}(X_2, X_3) \\ \text{cov}(X_3, X_1) & \text{cov}(X_3, X_2) & \text{var}(X_3) \end{bmatrix}$$

and in general it becomes

$$\mathbf{C}_X = \begin{bmatrix} \text{var}(X_1) & \text{cov}(X_1, X_2) & \dots & \text{cov}(X_1, X_N) \\ \text{cov}(X_2, X_1) & \text{var}(X_2) & \dots & \text{cov}(X_2, X_N) \\ \vdots & \vdots & \ddots & \vdots \\ \text{cov}(X_N, X_1) & \text{cov}(X_N, X_2) & \dots & \text{cov}(X_N, X_N) \end{bmatrix}. \tag{9.25}$$

The covariance matrix has many important properties, which are discussed next.

Property 9.1 – Covariance matrix is symmetric, i.e., $\mathbf{C}_X^T = \mathbf{C}_X$.

Proof:

$$\text{cov}(X_j, X_i) = \text{cov}(X_i, X_j) \qquad \text{(Why?)}$$

□

Property 9.2 – Covariance matrix is positive semidefinite.

Being positive semidefinite means that if $\mathbf{a}$ is the $N \times 1$ column vector $\mathbf{a} = [a_1\, a_2 \dots a_N]^T$, then $\mathbf{a}^T \mathbf{C}_X \mathbf{a} \geq 0$ for all $\mathbf{a}$. Note that $\mathbf{a}^T \mathbf{C}_X \mathbf{a}$ is a scalar and is referred to as a *quadratic form* (see Appendix C).

Proof: Consider the case of $N = 2$ since the extension is immediate. Let $U_i = X_i - E_{X_i}[X_i]$, which is zero mean, and therefore we have

$$\text{var}(a_1 X_1 + a_2 X_2)$$

$$
\begin{aligned}
&= \text{var}(a_1U_1 + a_2U_2) \qquad (\text{since } a_1X_1 + a_2X_2 = a_1U_1 + a_2U_2 + c \text{ for } c \text{ a constant}) \\
&= E_{\mathbf{X}}[(a_1U_1 + a_2U_2)^2] \quad (E_{\mathbf{X}}[U_1] = E_{\mathbf{X}}[U_2] = 0) \\
&= a_1^2 E_{\mathbf{X}}[U_1^2] + a_2^2 E_{\mathbf{X}}[U_2^2] + a_1a_2E_{\mathbf{X}}[U_1U_2] + a_2a_1E_{\mathbf{X}}[U_2U_1] \quad (\text{linearity of } E_{\mathbf{X}}) \\
&= a_1^2\text{var}(X_1) + a_2^2\text{var}(X_2) + a_1a_2\text{cov}(X_1, X_2) + a_2a_1\text{cov}(X_2, X_1) \\
&= \begin{bmatrix} a_1 & a_2 \end{bmatrix} \begin{bmatrix} \text{var}(X_1) & \text{cov}(X_1, X_2) \\ \text{cov}(X_2, X_1) & \text{var}(X_2) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \\
&= \mathbf{a}^T\mathbf{C}_X\mathbf{a}.
\end{aligned}
$$

Since $\text{var}(a_1X_1 + a_2X_2) \geq 0$ for all a_1 and a_2, it follows that $\mathbf{C}_X$ is positive semidefinite.

□

Also, note that the covariance matrix of random variables that are not perfectly predictable by a linear predictor is *positive definite*. A positive definite covariance matrix is one for which $\mathbf{a}^T\mathbf{C}_X\mathbf{a} > 0$ for all $\mathbf{a} \neq \mathbf{0}$. If, however, perfect prediction is possible, as would be the case if for $N = 2$ we had $a_1X_1 + a_2X_2 + c = 0$, for c a constant and for some a_1 and a_2, or equivalently if $X_2 = -(a_1/a_2)X_1 - (c/a_2)$, then the covariance matrix is only positive *semi*definite. This is because $\text{var}(a_1X_1 + a_2X_2) = \mathbf{a}^T\mathbf{C}_X\mathbf{a} = 0$ in this case.

Finally, with the general result that (see Problem 9.14)

$$\text{var}\left(\sum_{i=1}^{N} a_iX_i\right) = \mathbf{a}^T\mathbf{C}_X\mathbf{a} \tag{9.26}$$

we have upon letting $\mathbf{a} = \mathbf{1} = [1\,1\ldots1]^T$ be an $N \times 1$ vector of ones that

$$\text{var}\left(\sum_{i=1}^{N} X_i\right) = \mathbf{1}^T\mathbf{C}_X\mathbf{1}$$

which is another way of writing (9.20) (the effect of premultiplying a matrix by $\mathbf{1}^T$ and postmultiplying by $\mathbf{1}$ is to sum all the elements in the matrix).

The fact that the covariance matrix is a symmetric positive semidefinite matrix is important in that it must exhibit all the properties of that type of matrix. For example, if a matrix is symmetric positive semidefinite, then it can be shown that its determinant is nonnegative. As a result, it follows that the correlation coefficient must have a magnitude less than or equal to one (see Problem 9.18). Some other properties of a covariance matrix follow.

Property 9.3 – Covariance matrix for uncorrelated random variables is a diagonal matrix.

Note that a diagonal matrix is one for which all the off-diagonal elements are zero.
Proof: Let $\text{cov}(X_i, X_j) = 0$ for $i \neq j$ in (9.25).

□

Before listing the next property a new definition is needed. Similar to the definition that the expected value of a random vector is the vector of expected values of the elements, we define the expectation of a random matrix as the matrix of expected values of its elements. As an example, if $N = 2$ the definition is

$$E_{\mathbf{X}}\begin{bmatrix} g_{11}(\mathbf{X}) & g_{12}(\mathbf{X}) \\ g_{21}(\mathbf{X}) & g_{22}(\mathbf{X}) \end{bmatrix} = \begin{bmatrix} E_{\mathbf{X}}[g_{11}(\mathbf{X})] & E_{\mathbf{X}}[g_{12}(\mathbf{X})] \\ E_{\mathbf{X}}[g_{21}(\mathbf{X})] & E_{\mathbf{X}}[g_{22}(\mathbf{X})] \end{bmatrix}.$$

Property 9.4 – Covariance matrix of $\mathbf{Y} = \mathbf{AX}$, where $\mathbf{A}$ is an $M \times N$ matrix (with $M \leq N$), is easily determined.
The covariance matrix of $\mathbf{Y}$ is

$$\mathbf{C}_Y = \mathbf{A}\mathbf{C}_X\mathbf{A}^T. \tag{9.27}$$

Proof:
To prove this result without having to explicitly write out each element of the various matrices requires the use of matrix algebra. We therefore only sketch the proof and leave some details to the problems. The covariance matrix of $\mathbf{Y}$ can alternatively be defined by (see Problem 9.21)

$$\mathbf{C}_Y = E_{\mathbf{Y}}\left[(\mathbf{Y} - E_{\mathbf{Y}}[\mathbf{Y}])(\mathbf{Y} - E_{\mathbf{Y}}[\mathbf{Y}])^T\right].$$

Therefore,

$$\begin{aligned} \mathbf{C}_Y &= E_{\mathbf{X}}\left[(\mathbf{AX} - E_{\mathbf{X}}[\mathbf{AX}])(\mathbf{AX} - E_{\mathbf{X}}[\mathbf{AX}])^T\right] \\ &= E_{\mathbf{X}}\left[\mathbf{A}(\mathbf{X} - E_{\mathbf{X}}[\mathbf{X}])(\mathbf{A}(\mathbf{X} - E_{\mathbf{X}}[\mathbf{X}]))^T\right] && \text{(see Problem 9.22)} \\ &= \mathbf{A}E_{\mathbf{X}}\left[(\mathbf{X} - E_{\mathbf{X}}[\mathbf{X}])(\mathbf{X} - E_{\mathbf{X}}[\mathbf{X}])^T\right]\mathbf{A}^T && \text{(see Problem 9.23)} \\ &= \mathbf{A}\mathbf{C}_X\mathbf{A}^T. \end{aligned}$$

□

This result subsumes many of our previous ones (try $\mathbf{A} = \mathbf{1}^T = [1\,1\ldots1]$ and note that $\mathbf{C}_Y = \text{var}(Y)$ if $M = 1$, for example!).

Property 9.5 – Covariance matrix can always be diagonalized.
The importance of this property is that a diagonalized covariance matrix implies that the random variables are uncorrelated. Hence, by transforming a random vector of correlated random variable elements to one whose covariance matrix is diagonal, we can *decorrelate* the random variables. It is exceedingly fortunate that this transformation is a linear one and is easily found. In summary, if $\mathbf{X}$ has a covariance matrix $\mathbf{C}_X$, then we can find an $N \times N$ matrix $\mathbf{A}$ so that $\mathbf{Y} = \mathbf{AX}$ has the covariance matrix

$$\mathbf{C}_Y = \begin{bmatrix} \text{var}(Y_1) & 0 & \ldots & 0 \\ 0 & \text{var}(Y_2) & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \text{var}(Y_N) \end{bmatrix}.$$

The matrix $\mathbf{A}$ is not unique (see Problem 7.35 for a particular method). One possible determination of $\mathbf{A}$ is contained within the proof given next.

Proof:

We only sketch the proof of this result since it relies heavily on linear and matrix algebra (see also Appendix C). More details are available in [Noble and Daniel 1977]. Since $\mathbf{C}_X$ is a symmetric matrix, it has a set of N orthonormal eigenvectors with corresponding real eigenvalues. Since $\mathbf{C}_X$ is also positive semidefinite, the eigenvalues are nonnegative. Hence, we can find $N \times 1$ eigenvectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_N\}$ so that

$$\mathbf{C}_X \mathbf{v}_i = \lambda_i \mathbf{v}_i \qquad i = 1, 2, \ldots, N$$

where $\mathbf{v}_i^T \mathbf{v}_j = 0$ for $i \neq j$ (orthogonality), $\mathbf{v}_i^T \mathbf{v}_i = 1$ (normalized to unit length), and $\lambda_i \geq 0$. We can arrange the $N \times 1$ column vectors $\mathbf{C}_X \mathbf{v}_i$ and also $\lambda_i \mathbf{v}_i$ into $N \times N$ matrices so that

$$\begin{bmatrix} \mathbf{C}_X\mathbf{v}_1 & \mathbf{C}_X\mathbf{v}_2 & \ldots & \mathbf{C}_X\mathbf{v}_N \end{bmatrix} = \begin{bmatrix} \lambda_1\mathbf{v}_1 & \lambda_2\mathbf{v}_2 & \ldots & \lambda_N\mathbf{v}_N \end{bmatrix}. \tag{9.28}$$

But it may be shown that for an $N \times N$ matrix $\mathbf{A}$ and $N \times 1$ vectors $\mathbf{b}_1, \mathbf{b}_2, \mathbf{d}_1, \mathbf{d}_2$, using $N = 2$ for simplicity (see Problem 9.24),

$$\begin{bmatrix} \mathbf{A}\mathbf{b}_1 & \mathbf{A}\mathbf{b}_2 \end{bmatrix} = \mathbf{A} \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 \end{bmatrix} \tag{9.29}$$

$$\begin{bmatrix} c_1\mathbf{d}_1 & c_2\mathbf{d}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{d}_1 & \mathbf{d}_2 \end{bmatrix} \begin{bmatrix} c_1 & 0 \\ 0 & c_2 \end{bmatrix}. \tag{9.30}$$

Using these relationships (9.28) becomes

$$\mathbf{C}_X \underbrace{\begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \ldots & \mathbf{v}_N \end{bmatrix}}_{\mathbf{V}} = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \ldots & \mathbf{v}_N \end{bmatrix} \underbrace{\begin{bmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \lambda_N \end{bmatrix}}_{\mathbf{\Lambda}}$$

or

$$\mathbf{C}_X \mathbf{V} = \mathbf{V}\mathbf{\Lambda}.$$

(The matrix $\mathbf{V}$ is known as the *modal* matrix and is invertible.) Premultiplying both sides by $\mathbf{V}^{-1}$ produces

$$\mathbf{V}^{-1}\mathbf{C}_X\mathbf{V} = \mathbf{\Lambda}.$$

Next we use the property that the eigenvectors are orthonormal to assert that $\mathbf{V}^{-1} = \mathbf{V}^T$ (a property of orthogonal matrices), and therefore

$$\mathbf{V}^T\mathbf{C}_X\mathbf{V} = \mathbf{\Lambda} \tag{9.31}$$

Now recall from Property 9.4 that if $\mathbf{Y} = \mathbf{AX}$, then $\mathbf{C}_Y = \mathbf{A}\mathbf{C}_X\mathbf{A}^T$. Thus, if we let $\mathbf{Y} = \mathbf{AX} = \mathbf{V}^T\mathbf{X}$, we will have

$$\begin{aligned} \mathbf{C}_Y &= \mathbf{V}^T\mathbf{C}_X\mathbf{V} && \text{(from Property 9.4)} \\ &= \mathbf{\Lambda} && \text{(from (9.31))} \end{aligned}$$

and the covariance matrix of $\mathbf{Y}$ will be diagonal with ith diagonal element $\text{var}(Y_i) = \lambda_i \geq 0$.

□

This important result is used extensively in many disciplines. Later we will see that for some types of *continuous* random vectors, the use of this linear transformation will make the random variables not only uncorrelated but independent as well (see Example 12.14). An example follows.

Example 9.4 – Decorrelation of random variables

We consider a two-dimensional example whose joint PMF is given in Table 9.1. We

	$x_2 = -8$	$x_2 = 0$	$x_2 = 2$	$x_2 = 6$	$p_{X_1}[x_1]$
$x_1 = -8$	0	$\frac{1}{4}$	0	0	$\frac{1}{4}$
$x_1 = 0$	$\frac{1}{4}$	0	0	0	$\frac{1}{4}$
$x_1 = 2$	0	0	0	$\frac{1}{4}$	$\frac{1}{4}$
$x_1 = 6$	0	0	$\frac{1}{4}$	0	$\frac{1}{4}$
$p_{X_2}[x_2]$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	

Table 9.1: Joint PMF values.

first determine the covariance matrix $\mathbf{C}_X$ and then $\mathbf{A}$ so that $\mathbf{Y} = \mathbf{AX}$ consists of uncorrelated random variables. From Table 9.1 we have that

$$\begin{aligned} E_{X_1}[X_1] &= E_{X_2}[X_2] = 0 \\ E_{X_1}[X_1^2] &= E_{X_2}[X_2^2] = 26 \\ E_{X_1X_2}[X_1X_2] &= 6 \end{aligned}$$

and therefore we have that

$$\begin{aligned} \text{var}(X_1) &= \text{var}(X_2) = 26 \\ \text{cov}(X_1, X_2) &= 6 \end{aligned}$$

yielding a covariance matrix

$$\mathbf{C}_X = \begin{bmatrix} 26 & 6 \\ 6 & 26 \end{bmatrix}.$$

To find the eigenvectors we need to first find the eigenvalues and then solve $(\mathbf{C}_X - \lambda\mathbf{I})\mathbf{v} = \mathbf{0}$ for each eigenvector $\mathbf{v}$. To determine the eigenvalues we need to solve for λ in the equation $\det(\mathbf{C}_X - \lambda\mathbf{I}) = 0$. This is

$$\det\left(\begin{bmatrix} 26-\lambda & 6 \\ 6 & 26-\lambda \end{bmatrix}\right) = 0$$

or

$$(26-\lambda)(26-\lambda) - 36 = 0$$

and has solutions $\lambda_1 = 20$ and $\lambda_2 = 32$. Then, solving for the corresponding eigenvectors yields

$$(\mathbf{C}_X - \lambda_1\mathbf{I})\mathbf{v}_1 = \begin{bmatrix} 6 & 6 \\ 6 & 6 \end{bmatrix}\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

which yields after normalizing the eigenvector to have unit length

$$\mathbf{v}_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix}.$$

Similarly,

$$(\mathbf{C}_X - \lambda_2\mathbf{I})\mathbf{v}_2 = \begin{bmatrix} -6 & 6 \\ 6 & -6 \end{bmatrix}\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

which yields after normalizing the eigenvector to have unit length

$$\mathbf{v}_2 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}.$$

The modal matrix becomes

$$\mathbf{V} = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

and therefore

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

Hence, the transformed random vector $\mathbf{Y} = \mathbf{AX}$ is explicitly

$$\begin{aligned} Y_1 &= \frac{1}{\sqrt{2}}X_1 - \frac{1}{\sqrt{2}}X_2 \\ Y_2 &= \frac{1}{\sqrt{2}}X_1 + \frac{1}{\sqrt{2}}X_2 \end{aligned}$$

and Y_1 and Y_2 are uncorrelated random variables with

$$\begin{aligned} E_{\mathbf{Y}}[\mathbf{Y}] &= E_{\mathbf{Y}}[\mathbf{AX}] = \mathbf{A}E_{\mathbf{X}}[\mathbf{X}] = \mathbf{0} \\ \mathbf{C}_Y &= \mathbf{A}\mathbf{C}_X\mathbf{A}^T = \mathbf{V}^T\mathbf{C}_X\mathbf{V} = \mathbf{\Lambda} = \begin{bmatrix} 20 & 0 \\ 0 & 32 \end{bmatrix}. \end{aligned}$$

It is interesting to note in this example, and in general, that $\mathbf{A}$ is a *rotation* matrix or

$$\mathbf{A} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

where $\theta = \pi/4$. The effect of multiplying a 2×1 vector by this matrix is to rotate the vector 45° in the counterclockwise direction (see Problem 9.27). As seen in Figure 9.2 the values of $\mathbf{X}$, indicated by the small circles and also given in Table 9.1, become the values of $\mathbf{Y}$, indicated by the large circles. One can easily verify the rotation.

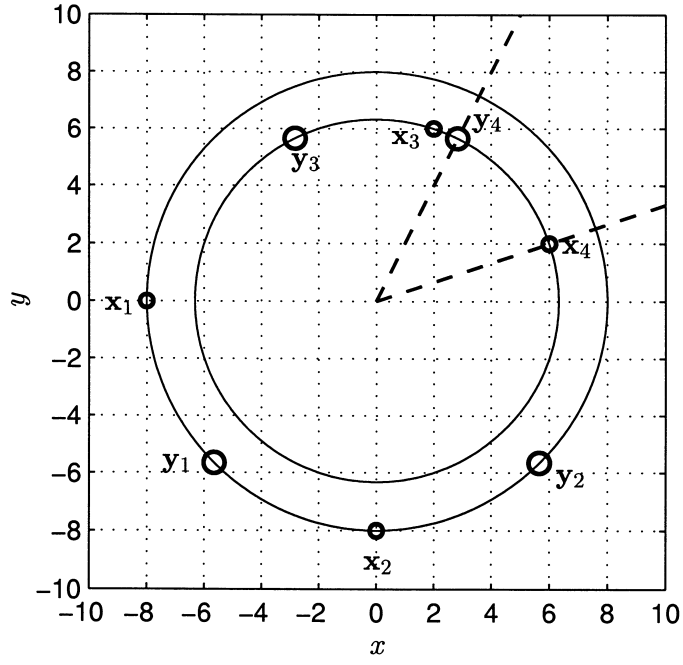


Figure 9.2: Sample points for $\mathbf{X}$ (small circles) and $\mathbf{Y}$ (large circles). The dashed lines indicate a 45° rotation.

9.6 Joint Moments and the Characteristic Function

The joint moments corresponding to an N-dimensional PMF are defined as

$$E_{X_1,X_2,\ldots,X_N}[X_1^{l_1}X_2^{l_2}\ldots X_N^{l_N}] = \sum_{x_1}\sum_{x_2}\cdots\sum_{x_N} x_1^{l_1}x_2^{l_2}\ldots x_N^{l_N} p_{X_1,X_2,\ldots,X_N}[x_1,x_2,\ldots,x_N]. \tag{9.32}$$

As usual if the random variables are independent, the joint PMF factors and therefore

$$E_{X_1,X_2,\ldots,X_N}[X_1^{l_1}X_2^{l_2}\ldots X_N^{l_N}] = E_{X_1}[X_1^{l_1}]E_{X_2}[X_2^{l_2}]\ldots E_{X_N}[X_N^{l_N}]. \tag{9.33}$$

The joint characteristic function is defined as

$$\phi_{X_1,X_2,\ldots,X_N}(\omega_1,\omega_2,\ldots,\omega_N) = E_{X_1,X_2,\ldots,X_N}\left[\exp[j(\omega_1X_1+\omega_2X_2+\cdots+\omega_NX_N)]\right] \tag{9.34}$$

and is evaluated as

$$\phi_{X_1,X_2,\ldots,X_N}(\omega_1,\omega_2,\ldots,\omega_N)$$
$$= \sum_{x_1}\sum_{x_2}\cdots\sum_{x_N}\exp[j(\omega_1x_1+\omega_2x_2+\cdots+\omega_Nx_N)]p_{X_1,X_2,\ldots,X_N}[x_1,x_2,\ldots,x_N].$$

In particular, for independent random variables, we have (see Problem 9.28)

$$\phi_{X_1,X_2,\ldots,X_N}(\omega_1,\omega_2,\ldots,\omega_N) = \phi_{X_1}(\omega_1)\phi_{X_1}(\omega_2)\ldots\phi_{X_1}(\omega_N).$$

Also, if $\mathbf{X}$ takes on integer values, the joint PMF can be found from the joint characteristic function using the inverse Fourier transform or

$$p_{X_1,X_2,\ldots,X_N}[k_1,k_2,\ldots,k_N]$$
$$= \int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\cdots\int_{-\pi}^{\pi}\phi_{X_1,X_2,\ldots,X_N}(\omega_1,\omega_2,\ldots,\omega_N)$$
$$\cdot\exp[-j(\omega_1k_1+\omega_2k_2+\cdots+\omega_Nk_N)]\frac{d\omega_1}{2\pi}\frac{d\omega_2}{2\pi}\ldots\frac{d\omega_N}{2\pi}. \tag{9.35}$$

All the properties of the 2-dimensional characteristic function extend to the general case. Note that once $\phi_{X_1,X_2,\ldots,X_N}(\omega_1,\omega_2,\ldots,\omega_N)$ is known, the characteristic function for any subset of the X_i's is found by setting ω_i equal to zero for the ones not in the subset. For example, to find $p_{X_1,X_2}[x_1,x_2]$, we let $\omega_3=\omega_4=\cdots=\omega_N=0$ in the joint characteristic function to yield

$$p_{X_1,X_2}[k_1,k_2] = \int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\underbrace{\phi_{X_1,X_2,\ldots,X_N}(\omega_1,\omega_2,0,0,\ldots,0)}_{\phi_{X_1,X_2}(\omega_1,\omega_2)}\exp[-j(\omega_1k_1+\omega_2k_2)]\frac{d\omega_1}{2\pi}\frac{d\omega_2}{2\pi}.$$

As seen previously, the joint moments can be obtained from the characteristic function. The general formula is

$$E_{X_1,X_2,\ldots,X_N}[X_1^{l_1}X_2^{l_2}\ldots X_N^{l_N}]$$
$$= \frac{1}{j^{l_1+l_2+\cdots+l_N}} \frac{\partial^{l_1+l_2+\cdots+l_N}}{\partial\omega_1^{l_1}\partial\omega_2^{l_2}\ldots\partial\omega_N^{l_N}} \phi_{X_1,X_2,\ldots,X_N}(\omega_1,\omega_2,\ldots,\omega_N)\Big|_{\omega_1=\omega_2=\cdots=\omega_N=0}. \tag{9.36}$$

9.7 Conditional Probability Mass Functions

When we have an N-dimensional random vector, many different conditional PMFs can be defined. A straightforward extension of the conditional PMF $p_{Y|X}$ encountered in Chapter 8 is the conditional PMF of a single random variable conditioned on knowledge of the outcomes of all the other random variables. For example, it is of interest to study $p_{X_N|X_1,X_2,\ldots,X_{N-1}}$, whose definition is

$$p_{X_N|X_1,X_2,\ldots,X_{N-1}}[x_N|x_1,x_2,\ldots,x_{N-1}] = \frac{p_{X_1,X_2,\ldots,X_N}[x_1,x_2,\ldots,x_N]}{p_{X_1,X_2,\ldots,X_{N-1}}[x_1,x_2,\ldots,x_{N-1}]}. \tag{9.37}$$

Then by rearranging (9.37) we have upon omitting the arguments

$$p_{X_1,X_2,\ldots,X_N} = p_{X_N|X_1,X_2,\ldots,X_{N-1}}p_{X_1,X_2,\ldots,X_{N-1}}. \tag{9.38}$$

If we replace N by $N-1$ in (9.37), we have

$$p_{X_{N-1}|X_1,X_2,\ldots,X_{N-2}} = \frac{p_{X_1,X_2,\ldots,X_{N-1}}}{p_{X_1,X_2,\ldots,X_{N-2}}}$$

or

$$p_{X_1,X_2,\ldots,X_{N-1}} = p_{X_{N-1}|X_1,X_2,\ldots,X_{N-2}}p_{X_1,X_2,\ldots,X_{N-2}}.$$

Inserting this into (9.38) yields

$$p_{X_1,X_2,\ldots,X_N} = p_{X_N|X_1,X_2,\ldots,X_{N-1}}p_{X_{N-1}|X_1,X_2,\ldots,X_{N-2}}p_{X_1,X_2,\ldots,X_{N-2}}.$$

Continuing this process results in the *general chain rule* for joint PMFs (see also (4.10))

$$p_{X_1,X_2,\ldots,X_N} = p_{X_N|X_1,X_2,\ldots,X_{N-1}}p_{X_{N-1}|X_1,X_2,\ldots,X_{N-2}}\cdots p_{X_2|X_1}p_{X_1}. \tag{9.39}$$

A particularly useful special case of this relationship occurs when the conditional PMFs satisfies

$$p_{X_n|X_1,X_2,\ldots,X_{n-1}} = p_{X_n|X_{n-1}} \qquad \text{for } n = 3,4,\ldots,N \tag{9.40}$$

or X_n is independent of $X_1 \ldots, X_{n-2}$ *if* X_{n-1} *is known* for all $n \geq 3$. If we view n as a time index, then this says that the probability of the current random variable X_n is independent of the past outcomes once the most recent past outcome X_{n-1} is known. This is called the *Markov property*, which was described in Section 4.6.4. When the Markov property holds, we can rewrite (9.39) in the particularly simple form

$$p_{X_1,X_2,\ldots,X_N} = p_{X_N|X_{N-1}} p_{X_{N-1}|X_{N-2}} \cdots p_{X_2|X_1} p_{X_1} \tag{9.41}$$

which is a factorization of the N-dimensional joint PMF into a product of *first-order* conditional PMFs. It can be considered as the logical extension of the factorization of the N-dimensional joint PMF of independent random variables into the product of its marginals. As such it enjoys many useful properties, which are discussed in Chapter 22. A simple example of when (9.40) holds is for a "running" sum of independent random variables or $X_n = \sum_{i=1}^{n} U_i$, where the U_i's are independent. Then, we have

$$\begin{aligned} X_1 &= U_1 \\ X_2 &= U_1 + U_2 = X_1 + U_2 \\ X_3 &= U_1 + U_2 + U_3 = X_2 + U_3 \\ &\vdots = \vdots \\ X_N &= X_{N-1} + U_N. \end{aligned}$$

For example, X_2 is known, the PMF of $X_3 = X_2 + U_3$ depends only on U_3 and not on X_1. Also, it is seen from the definition of the random variables that U_3 and $U_1 = X_1$ are independent. Thus, once X_2 is known, X_3 (a function of U_3) is independent of X_1 (a function of U_1). As a result, $p_{X_3|X_2,X_1} = p_{X_3|X_2}$ and in general

$$p_{X_n|X_1,X_2,\ldots,X_{n-1}} = p_{X_n|X_{n-1}} \qquad \text{for } n = 3, 4, \ldots, N$$

or (9.40) is satisfied. It is said that "the PMF of X_n given the past samples depends only on the most recent past sample". To illustrate this we consider a particular running sum of independent random variables known as a random walk.

Example 9.5 – Random walk

Let U_i for $i = 1, 2, \ldots, N$ be independent random variables with the same PMF

$$p_U[k] = \begin{cases} 1-p & k = -1 \\ p & k = 1 \end{cases}$$

and define

$$X_n = \sum_{i=1}^{n} U_i.$$

At each "time" n the new random variable X_n changes from the old random variable X_{n-1} by ± 1 since $X_n = X_{n-1} + U_n$. The joint PMF is from (9.41)

$$p_{X_1,X_2,\ldots,X_N} = \prod_{n=1}^{N} p_{X_n|X_{n-1}} \tag{9.42}$$

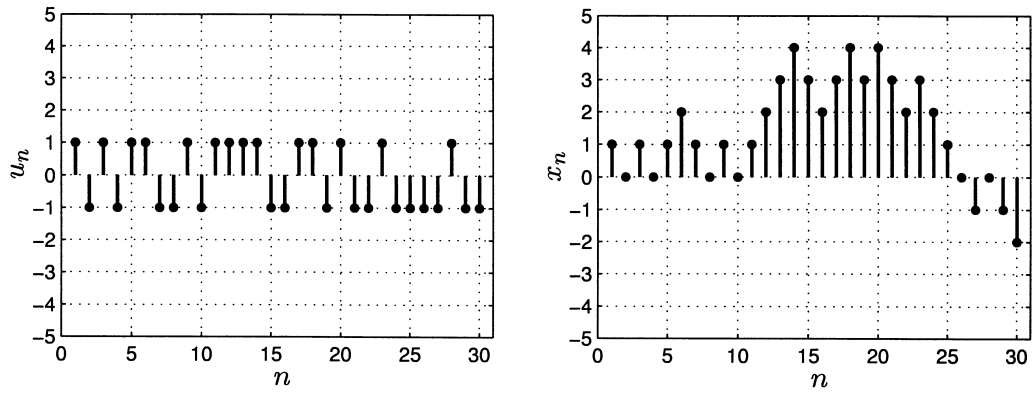
where $p_{X_1|X_0}$ is defined as p_{X_1}. But $p_{X_n|X_{n-1}}$ can be found by noting that $X_n = X_{n-1} + U_n$ and therefore if $X_{n-1} = x_{n-1}$ we have that

$$\begin{aligned} p_{X_n|X_{n-1}}[x_n|x_{n-1}] &= p_{U_n|X_{n-1}}[x_n - x_{n-1}|x_{n-1}] && \text{(step 1 - transform PMF)} \\ &= p_{U_n}[x_n - x_{n-1}] && \text{(step 2 - independence)} \\ &= p_U[x_n - x_{n-1}] && (U_n\text{'s have same PMF}). \end{aligned}$$

Step 1 results from the transformed random variable $Y = X + c$, where c is a constant, having a PMF $p_Y[y_i] = p_X[y_i - c]$. Step 2 results from U_n being independent of $X_{n-1} = \sum_{i=1}^{n-1} U_i$ since all the U_i's are independent. Finally, we have from (9.42)

$$p_{X_1,X_2,\ldots,X_N}[x_1, x_2, \ldots, x_N] = \prod_{n=1}^{N} p_U[x_n - x_{n-1}]. \tag{9.43}$$

A realization of the random variables for $p = 1/2$ is shown in Figure 9.3. As justified by the character of the outcomes in Figure 9.3b, this *random process* is termed a *random walk*. We will say more about this later in Chapter 16. Note that the



(a) Realization of U_n's

(b) Realization of X_n's

Figure 9.3: Typical realization of a random walk.

probability of the realization in Figure 9.3b is from (9.43)

$$p_{X_1,X_2,\ldots,X_{30}}[1, 0, \ldots, -2] = \prod_{n=1}^{30} p_U[x_n - x_{n-1}] = \prod_{n=1}^{30} \frac{1}{2} = \left(\frac{1}{2}\right)^{30}$$

since $p_U[-1] = p_U[1] = 1/2$.

$\Diamond$

9.8 Computer Simulation of Random Vectors

To generate a realization of a random vector we can use the direct method described in Section 7.11 or the conditional approach of Section 8.7. The latter uses the general chain rule (see (9.39)). We will not pursue this further as the extension to an $N \times 1$ random vector is obvious. Instead we concentrate on two important descriptors of a random vector, those being the mean vector given by (9.15) and the covariance matrix given by (9.25). We wish to see how to estimate these quantities. In practice, the N-dimensional PMF is usually quite difficult to estimate and so we settle for the estimation of the means and covariances. The mean vector is easily estimated by estimating each element by its sample mean as we have done in Section 6.8. Here we assume to have M realizations of the $N \times 1$ random vector $\mathbf{X}$, which we denote as $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_M\}$. The mean vector estimate becomes

$$\widehat{E_{\mathbf{X}}[\mathbf{X}]} = \frac{1}{M} \sum_{m=1}^{M} \mathbf{x}_m \tag{9.44}$$

which is the same as estimating the ith component of $E_{\mathbf{X}}[\mathbf{X}]$ by $(1/M) \sum_{m=1}^{M} [\mathbf{x}_m]_i$, where $[\boldsymbol{\xi}]_i$ denotes the ith component of the vector $\boldsymbol{\xi}$. To estimate the $N \times N$ covariance matrix we first recall that the vector/matrix definition is

$$\mathbf{C}_X = E_{\mathbf{X}} \left[(\mathbf{X} - E_{\mathbf{X}}[\mathbf{X}]) (\mathbf{X} - E_{\mathbf{X}}[\mathbf{X}])^T \right].$$

This can also be shown to be equivalent to (see Problem 9.31)

$$\mathbf{C}_X = E_{\mathbf{X}} \left[\mathbf{X}\mathbf{X}^T \right] - (E_{\mathbf{X}}[\mathbf{X}])(E_{\mathbf{X}}[\mathbf{X}])^T. \tag{9.45}$$

We can now replace $E_{\mathbf{X}}[\mathbf{X}]$ by the estimate of (9.44). To estimate the $N \times N$ matrix

$$E_{\mathbf{X}} \left[\mathbf{X}\mathbf{X}^T \right]$$

we replace it by $(1/M) \sum_{m=1}^{M} \mathbf{x}_m \mathbf{x}_m^T$ since it is easily shown that the (i, j) element of $E_{\mathbf{X}} \left[\mathbf{X}\mathbf{X}^T \right]$ is

$$\left[E_{\mathbf{X}}[\mathbf{X}\mathbf{X}^T] \right]_{ij} = E_{\mathbf{X}}[X_i X_j] = E_{X_i X_j}[X_i X_j]$$

and

$$\left[\frac{1}{M} \sum_{m=1}^{M} \mathbf{x}_m \mathbf{x}_m^T \right]_{ij} = \frac{1}{M} \sum_{m=1}^{M} [\mathbf{x}_m]_i [\mathbf{x}_m]_j.$$

Thus we have that

$$\widehat{\mathbf{C}_X} = \frac{1}{M}\sum_{m=1}^{M}\mathbf{x}_m\mathbf{x}_m^T - \left(\frac{1}{M}\sum_{m=1}^{M}\mathbf{x}_m\right)\left(\frac{1}{M}\sum_{m=1}^{M}\mathbf{x}_m\right)^T$$

which can also be written as

$$\widehat{\mathbf{C}_X} = \frac{1}{M}\sum_{m=1}^{M}\left(\mathbf{x}_m - \widehat{E_{\mathbf{X}}[\mathbf{X}]}\right)\left(\mathbf{x}_m - \widehat{E_{\mathbf{X}}[\mathbf{X}]}\right)^T \tag{9.46}$$

where $\widehat{E_{\mathbf{X}}[\mathbf{X}]}$ is given by (9.44). The latter form of the covariance matrix estimate is also more easily implemented. An example follows.

Example 9.6 – Decorrelation of random variables – continued

In Example 9.4 we showed that we could decorrelate the random variable components of a random vector by applying the appropriate linear transformation to the random vector. In particular, if the 2×1 random vector $\mathbf{X}$ whose joint PMF is given in Table 9.1 is transformed to a random vector $\mathbf{Y}$, where

$$\mathbf{Y} = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}\mathbf{X}$$

then the covariance matrix for $\mathbf{X}$

$$\mathbf{C}_X = \begin{bmatrix} 26 & 6 \\ 6 & 26 \end{bmatrix}$$

becomes the diagonal covariance matrix for $\mathbf{Y}$

$$\mathbf{C}_Y = \begin{bmatrix} 20 & 0 \\ 0 & 32 \end{bmatrix}.$$

To check this we generate realizations of $\mathbf{X}$, as explained in Section 7.11 and then use the estimate of the covariance matrix given by (9.46). The results are for $M = 1000$ realizations

$$\begin{aligned} \widehat{\mathbf{C}_X} &= \begin{bmatrix} 25.9080 & 6.1077 \\ 6.1077 & 25.8558 \end{bmatrix} \\ \widehat{\mathbf{C}_Y} &= \begin{bmatrix} 19.7742 & 0.0261 \\ 0.0261 & 31.9896 \end{bmatrix} \end{aligned}$$

and are near to the true covariance matrices. The entire MATLAB program is given next.

```
% covexample.m
clear all % clears out all previous variables from workspace
rand('state',0); % sets random number generator to initial value
M=1000;
for m=1:M % generate realizations of X (see Section 7.11)
   u=rand(1,1);
   if u<=0.25
      x(1,m)=-8;x(2,m)=0;
   elseif u>0.25&u<=0.5
      x(1,m)=0;x(2,m)=-8;
   elseif u>0.5&u<=0.75
      x(1,m)=2;x(2,m)=6;
   else
      x(1,m)=6;x(2,m)=2;
   end
end
meanx=[0 0]'; % estimate mean vector of X
for m=1:M
   meanx=meanx+x(:,m)/M;
end
meanx
CX=zeros(2,2);
for m=1:M % estimate covariance matrix of X
   xbar(:,m)=x(:,m)-meanx;
   CX=CX+xbar(:,m)*xbar(:,m)'/M;
end
CX
A=[1/sqrt(2) -1/sqrt(2);1/sqrt(2) 1/sqrt(2)];
for m=1:M % transform random vector X
   y(:,m)=A*x(:,m);
end
meany=[0 0]'; %estimate mean vector or Y
for m=1:M
   meany=meany+y(:,m)/M;
end
meany
CY=zeros(2,2);
for m=1:M % estimate covariance matrix of Y
   ybar(:,m)=y(:,m)-meany;
   CY=CY+ybar(:,m)*ybar(:,m)'/M;
end
CY
```

9.9 Real-World Example – Image Coding

The methods for digital storage and transmission of images is an important consideration in the modern digital age. One of the standard procedures used to convert an image to its digital representation is the JPEG encoding format [Sayood 1996]. It makes the observation that many images contain portions that do not change significantly in content. Such would be the case for the image of a house in which the color and texture of the siding, whether it be aluminum siding or clapboards, is relatively constant as the image is scanned in the horizontal direction. To store and transmit all this redundant information is costly and time consuming. Hence, it is desirable to reduce the image to its basic set of information. Consider a gray scale image for simplicity. Each pixel, which is a dot of a given intensity level, is modeled as a random variable. For the house image example, note that for the siding pixels, the random variables are heavily correlated. For example, if X_1 and X_2 denote neighboring pixels in the horizontal direction, then we would expect the correlation coefficient $\rho_{X_1,X_2} = 1$. If this is the case, then we know from Section 7.9 that $X_1 = X_2$, assuming zero mean random variables in our model. There is no economy in storing/transmitting the values $X_1 = x_1$ *and* $X_2 = x_2 = x_1$. We should just store/transmit $X_1 = x_1$ and when it is necessary to reconstruct the image let $\hat{X}_2 = X_1 = x_1$. In this case, there is no image degradation in doing so. If, however, $|\rho_{X_1,X_2}| < 1$, then there will be an error in the reconstructed X_2. If the correlation coefficient is close to ± 1, this error will be small. Even if it is not, for many images the errors introduced are perceptually unimportant. Human visual perception can tolerate gross errors before the image becomes unsatisfactory.

To apply this idea to image coding we will consider a simple yet illustrative example. The amount of correlation between random variables is quantified by the covariances. In particular, for multiple random variables this information is embodied in the covariance matrix. For example, if $N = 3$ a covariance matrix of

$$\mathbf{C}_X = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 4 & 3.8 \\ 0 & 3.8 & 4 \end{bmatrix} \tag{9.47}$$

indicates that

$$\rho_{X_1,X_2} = \rho_{X_1,X_3} = 0$$

but

$$\rho_{X_2,X_3} = \frac{3.8}{\sqrt{4 \cdot 4}} = 0.95.$$

Clearly, then (X_1, X_2) or (X_1, X_3) contain most of the information. For more complicated covariance matrices these relationships are not so obvious. For example, if

$$\mathbf{C}_X = \begin{bmatrix} 4 & 1 & 5 \\ 1 & 4 & 5 \\ 5 & 5 & 10 \end{bmatrix} \tag{9.48}$$

it is not obvious that $X_3 = X_1 + X_2$ (assuming zero mean random variables). (This is verified by showing that $E[(X_3 - (X_1 + X_2))^2] = 0$ (see Problem 9.33)).

The technique of *transform coding* [Sayood 1996] used in the JPEG encoding scheme takes advantage of the correlation between random variables. The particular version we describe here can be shown to be an optimal approach [Kramer and Mathews 1956]. It is termed the *Karhunen-Loeve transform* and an approximate version is used in the JPEG encoding. Transform coding operates on a random vector $\mathbf{X}$ and proceeds as follows:

1. Transform the random variables into uncorrelated ones via a linear transformation $\mathbf{Y} = \mathbf{AX}$, where $\mathbf{A}$ is an invertible $N \times N$ matrix.

2. Discard the random variables whose variance is small relative to the others by setting the corresponding elements of $\mathbf{Y}$ equal to zero. This yields a new $N \times 1$ random vector $\hat{\mathbf{Y}}$. This vector would be stored or transmitted. (Of course, the zero vector elements would not require encoding, thereby effecting data compression. Their locations, though, would need to be specified.)

3. Transform back to $\hat{\mathbf{X}} = \mathbf{A}^{-1}\hat{\mathbf{Y}}$ to recover an approximation to the original random variables (if the values $\hat{\mathbf{Y}}$ were stored then this would occur upon retrieval or if they were transmitted, this would occur at the receiver).

By decorrelating the random variables first it becomes obvious which components can be discarded without significantly affecting the reconstructed vector. To accomplish the first step we have already determined that a suitable decorrelation matrix is $\mathbf{V}^T$, where $\mathbf{V}$ is the matrix of eigenvectors of $\mathbf{C}_X$. Thus, we have that

$$\begin{aligned}
\mathbf{C}_Y &= \mathbf{A}\mathbf{C}_X\mathbf{A}^T \\
&= \mathbf{V}^T\mathbf{C}_X\mathbf{V} \\
&= \mathbf{\Lambda} = \begin{bmatrix} \text{var}(Y_1) & 0 & 0 \\ 0 & \text{var}(Y_2) & 0 \\ 0 & 0 & \text{var}(Y_3) \end{bmatrix}.
\end{aligned}$$

We now carry out the transform coding procedure for the covariance matrix of (9.48). This is done numerically using MATLAB. The statement `[V Lambda]=eig(CX)` will produce the matrices $\mathbf{V}$ and $\mathbf{\Lambda}$, as

$$\begin{aligned}
\mathbf{V} &= \begin{bmatrix} 0.4082 & -0.7071 & 0.5774 \\ 0.4082 & -0.7071 & 0.5774 \\ 0.8165 & 0 & -0.5774 \end{bmatrix} \\
\mathbf{\Lambda} &= \begin{bmatrix} 15 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 0 \end{bmatrix}.
\end{aligned}$$

Hence, $\text{var}(Y_3) = \lambda_3 = 0$ so that we discard it by setting $\hat{Y}_3 = 0$ and therefore

$$\hat{\mathbf{Y}} = \begin{bmatrix} Y_1 \\ Y_2 \\ 0 \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}}_{\mathbf{B}} \underbrace{\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix}}_{\mathbf{Y}}.$$

The reconstructed random vector becomes with $\mathbf{A} = \mathbf{V}^T$

$$\begin{aligned} \hat{\mathbf{X}} = \mathbf{A}^{-1}\hat{\mathbf{Y}} &= \mathbf{V}\hat{\mathbf{Y}} \\ &= \mathbf{VBY} \\ &= \mathbf{VBV}^T\mathbf{X} \end{aligned}$$

and since

$$\mathbf{VBV}^T = \begin{bmatrix} \frac{2}{3} & -\frac{1}{3} & \frac{1}{3} \\ -\frac{1}{3} & \frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{2}{3} \end{bmatrix}$$

we have that

$$\begin{aligned} \hat{\mathbf{X}} &= \begin{bmatrix} \frac{2}{3}X_1 - \frac{1}{3}X_2 + \frac{1}{3}X_3 \\ -\frac{1}{3}X_1 + \frac{2}{3}X_2 + \frac{1}{3}X_3 \\ \frac{1}{3}X_1 + \frac{1}{3}X_2 + \frac{2}{3}X_3 \end{bmatrix} \\ &= \begin{bmatrix} X_1 \\ X_2 \\ X_1 + X_2 \end{bmatrix} \qquad (\text{using } X_3 = X_1 + X_2, \text{ see Problem 9.33}) \\ &= \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix}. \end{aligned}$$

Here we see that the reconstructed vector $\hat{\mathbf{X}}$ is identical to the original one. Generally, however, there will be an error. For the covariance matrix of (9.47) there will be an error since X_2 and X_3 are not perfectly correlated. For that covariance matrix the eigenvector and eigenvalue matrices are

$$\begin{aligned} \mathbf{V} &= \begin{bmatrix} 0 & 1 & 0 \\ 0.7071 & 0 & 0.7071 \\ 0.7071 & 0 & -0.7071 \end{bmatrix} \\ \mathbf{\Lambda} &= \begin{bmatrix} 7.8 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 0.2 \end{bmatrix} \end{aligned}$$

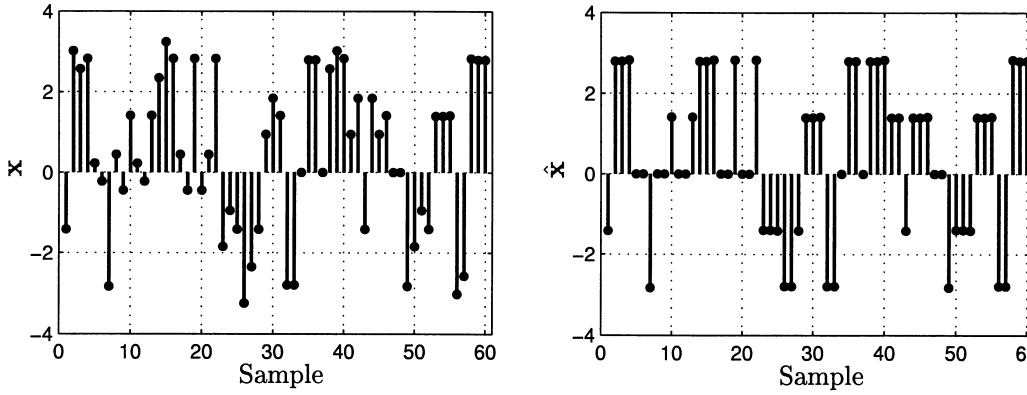
and it is seen that the decorrelated random variables all have a nonzero variance (recall that $\text{var}(Y_i) = \lambda_i$). This indicates that no component of $\mathbf{Y}$ can be discarded without causing an error upon reconstruction. By discarding Y_3, which has the smallest variance, we will incur the least amount of error. Doing so produces the reconstructed random vector

$$\begin{aligned}\hat{\mathbf{X}} &= \mathbf{V}\mathbf{B}\mathbf{V}^T\mathbf{X} \\ &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix} \mathbf{X}\end{aligned}$$

which becomes

$$\hat{\mathbf{X}} = \begin{bmatrix} X_1 \\ \dfrac{X_2 + X_3}{2} \\ \dfrac{X_2 + X_3}{2} \end{bmatrix}.$$

It is seen that the components X_2 and X_3 are replaced by their averages. This is due to the nearly unity correlation coefficient coefficient ($\rho_{X_2,X_3} = 0.95$) between these components. As an example, we generate 20 realizations of $\mathbf{X}$ as shown in Figure 9.4a, where the first realization is displayed in samples $1, 2, 3$; the second realization in samples $4, 5, 6$, etc. The reconstructed realizations are shown in Figure 9.4b.



(a) Original

(b) Reconstruction

Figure 9.4: Realizations of original random vector $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{20}\}$ and reconstructed random vectors $\{\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2, \ldots, \hat{\mathbf{x}}_{20}\}$. The displayed samples shown are components of $\mathbf{x}_1$, followed by components of $\mathbf{x}_2$, etc.

Finally, the error between the two is shown in Figure 9.5. Note that the total average

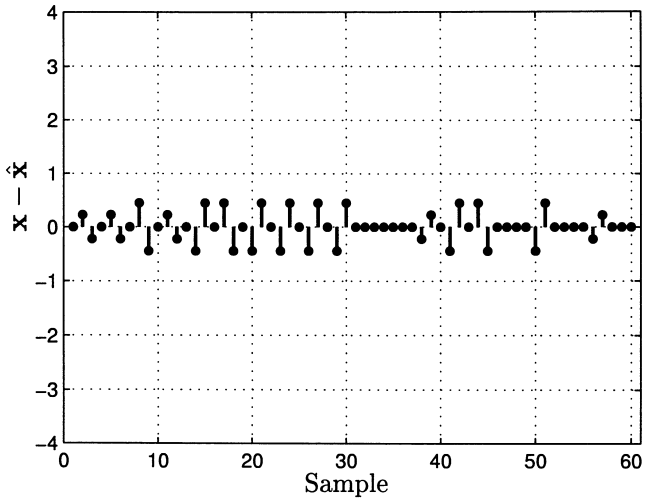


Figure 9.5: Error between original random vector realizations and reconstructed ones shown in Figure 9.4.

squared error or the total mean square error (MSE) is given by $\sum_{i=1}^{3} E_{\mathbf{X}}[(X_i - \hat{X}_i)^2]$ which is

$$\begin{aligned}
\text{Total mse} &= E[(X_1 - \hat{X}_1)^2 + (X_2 - \hat{X}_2)^2 + (X_3 - \hat{X}_3)^2] \\
&= E[(X_2 - (X_2 + X_3)/2)^2] + E[(X_3 - (X_2 + X_3)/2)^2] \\
&= E[((X_2 - X_3)/2)^2] + E[((X_3 - X_2)/2)^2] \\
&= \frac{1}{2}E[(X_2 - X_3)^2] \\
&= \frac{1}{2}[\text{var}(X_2) + \text{var}(X_3) - 2\text{cov}(X_2, X_3)] \\
&= \frac{1}{2}[4 + 4 - 2(3.8)] = 0.2.
\end{aligned}$$

This total MSE is estimated by taking the sum of the squares of the values in Figure 9.5 and dividing by 20, the number of vector realizations. Also, note what the total MSE would have been if $\rho_{X_2,X_3} = 1$.

Finally, to appreciate the error in terms of human vision perception, we can convert the realizations of $\mathbf{X}$ and $\hat{\mathbf{X}}$ into an image. This is shown in Figure 9.6. The grayscale bar shown at the right can be used to convert the various shades of gray into numerical values. Also, note that as expected (see $\mathbf{C}_X$ in (9.47)) X_1 is uncorrelated with X_2 and X_3, while X_2 and X_3 are heavily correlated in the upper image. In the lower image X_2 and X_3 have been replaced by their average.

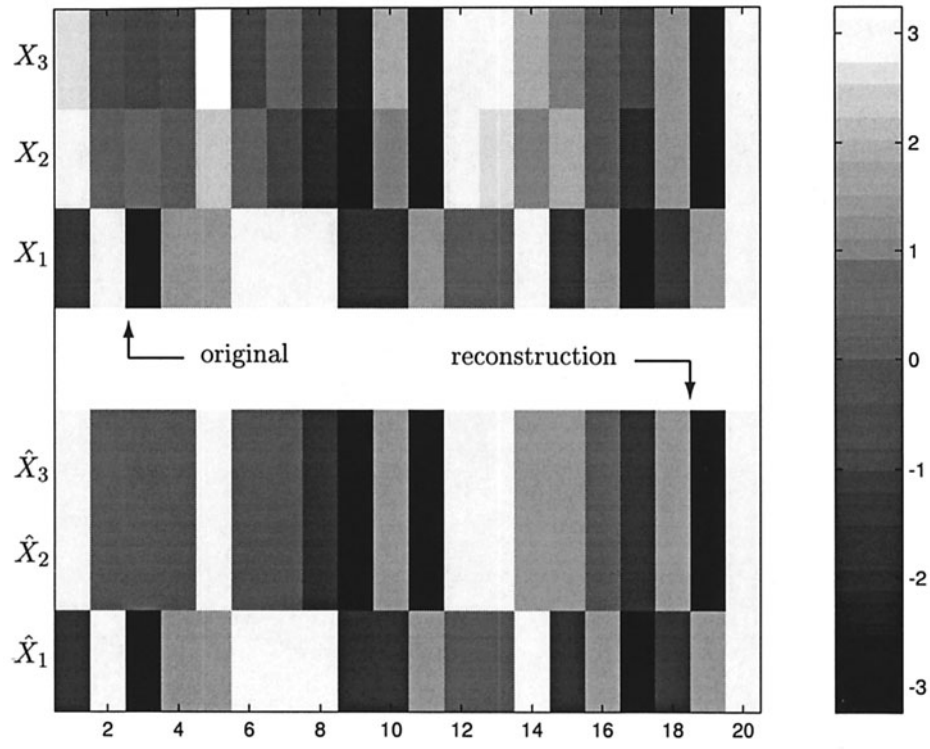


Figure 9.6: Realizations of original random vector and reconstructed random vectors displayed as gray-scale images. The upper image is the original and the lower image is the reconstructed image.

References

Kramer, H.P., Mathews, M.V., "A Linear Coding for Transmitting a Set of Correlated Signals," *IRE Trans. on Information Theory*, Vol. IT-2, pp. 41–46, 1956.

Noble, B., Daniel, J.W., *Applied Linear Algebra*, Prentice-Hall, Englewood Cliffs, NJ, 1977.

Sayood, K. *Introduction to Data Compression*, Morgan Kaufman, San Francisco, 1996.

Problems

9.1 (◡̈) **(w)** A retired person gets up in the morning and decides what to do that day. He will go fishing with probability 0.3, or he will visit his daughter with probability 0.2, or else he will stay home and tend to his garden. If the decision

that he makes each day is independent of the decisions made on the other days, what is the probability that he will go fishing for 3 days, visit his daughter for 2 days, and garden for 2 days of the week?

9.2 (f,c) Compute the values of a multinomial PMF if $N = 3$, $M = 4$, $p_1 = 0.2$, and $p_2 = 0.4$ for all possible k_1, k_2, k_3. Do the sum of the values equal one? Hint: You will need a computer to do this.

9.3 (t) Prove the multinomial formula given by (9.5) for $N = 3$ by the following method. Use the binomial formula to yield

$$(a_1 + b)^M = \sum_{k_1=0}^{M} \frac{M!}{k_1!(M-k_1)!} a_1^{k_1} b^{M-k_1}.$$

Then let $b = a_2 + a_3$ so that upon using the binomial formula again we have

$$b^{M-k_1} = (a_2 + a_3)^{M-k_1} = \sum_{k_2=0}^{M-k_1} \frac{(M-k_1)!}{k_2!(M-k_1-k_2)!} a_2^{k_2} a_3^{M-k_1-k_2}.$$

Finally, rearrange the sums and note that $k_3 = M - k_1 - k_2$ so that there is actually only a double sum in (9.5) for $N = 3$ due to this constraint.

9.4 (☺) (f) Is the following function a valid PMF?

$$p_{X_1,X_2,X_3}[k_1,k_2,k_3] = \frac{1}{8}\left(\frac{1}{2}\right)^{k_1}\left(\frac{1}{4}\right)^{k_2} \qquad \begin{array}{l} k_1 = 0,1,\ldots \\ k_2 = 0,1,\ldots \\ k_3 = -1,0,1. \end{array}$$

9.5 (w) For the joint PMF

$$p_{X_1,X_2,X_3}[k_1,k_2,k_3] = (1-a)(1-b)(1-c)a^{k_1}b^{k_2}c^{k_3} \qquad \begin{array}{l} k_1 = 0,1,\ldots \\ k_2 = 0,1,\ldots \\ k_3 = 0,1,\ldots \end{array}$$

where $0 < a < 1$, $0 < b < 1$, and $0 < c < 1$, find the marginal PMFs p_{X_1}, p_{X_2} and p_{X_3}.

9.6 (☺) (w) For the joint PMF given below are there any subsets of the random variables that are independent of each other?

$$p_{X_1,X_2,X_3}[k_1,k_2,k_3] = \binom{M}{k_1,k_2} p_1^{k_1} p_2^{k_2} (1-p_3) p_3^{k_3} \qquad \begin{array}{l} k_1 = 0,1,\ldots,M \\ k_2 = M - k_1 \\ k_3 = 0,1,\ldots \end{array}$$

where $0 < p_1 < 1$, $p_2 = 1 - p_1$, and $0 < p_3 < 1$.

9.7 (f) A random vector $\mathbf{X}$ with the joint PMF

$$p_{X_1,X_2,X_3}[k_1,k_2,k_3] = \exp[-(\lambda_1+\lambda_2+\lambda_3)]\frac{\lambda_1^{k_1}\lambda_2^{k_2}\lambda_3^{k_3}}{k_1!k_2!k_3!} \quad \begin{array}{l} k_1 = 0,1,\ldots \\ k_2 = 0,1,\ldots \\ k_3 = 0,1,\ldots \end{array}$$

is transformed according to $\mathbf{Y} = \mathbf{AX}$ where

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}.$$

Find the joint PMF of $\mathbf{Y}$.

9.8 (t) Prove that

$$\int_{-\pi}^{\pi} \exp(j\omega k)\frac{d\omega}{2\pi} = \begin{cases} 0 & k \neq 0 \\ 1 & k = 0. \end{cases}$$

Hint: Expand $\exp(j\omega k)$ into its real and imaginary parts and note that $\int(g(\omega)+jh(\omega))d\omega = \int g(\omega)d\omega + j\int h(\omega)d\omega$.

9.9 (t) Prove that the sum of N independent Poisson random variables with $X_i \sim \text{Pois}(\lambda_i)$ for $i = 1,2,\ldots,N$ is again Poisson distributed but with parameter $\lambda = \sum_{i=1}^{N}\lambda_i$. Hint: See Section 9.4.

9.10 (☺) (w) The components of a random vector $\mathbf{X} = [X_1\, X_2 \ldots X_N]^T$ all have the same mean $E_X[X]$ and the same variance $\text{var}(X)$. The "sample mean" random variable

$$\bar{X} = \frac{1}{N}\sum_{i=1}^{N} X_i$$

is formed. If the X_i's are independent, find the mean and variance of $\bar{X}$. What happens to the variance as $N \to \infty$? Does this tell you anything about the PMF of $\bar{X}$ as $N \to \infty$?

9.11 (w) Repeat Problem 9.10 if we know that each $X_i \sim \text{Ber}(p)$. How can this result be used to motivate the relative frequency interpretation of probability?

9.12 (f) If the covariance matrix of a 3×1 random vector $\mathbf{X}$ is

$$\mathbf{C}_X = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 2 & 2 \\ 1 & 2 & 4 \end{bmatrix}$$

find the correlation coefficients ρ_{X_1,X_2}, ρ_{X_1,X_3}, and ρ_{X_2,X_3}.

9.13 (☺) (w) A 2×1 random vector is given by

$$\mathbf{X} = \begin{bmatrix} U \\ 2U \end{bmatrix}$$

where $\text{var}(U) = 1$. Find the covariance matrix for $\mathbf{X}$. Next find the correlation coefficient ρ_{X_1,X_2}. Finally, compute the determinant of the covariance matrix. Is the covariance matrix positive definite? Hint: A positive definite matrix must have a positive determinant.

9.14 (t) Prove (9.26) by noting that

$$\mathbf{a}^T \mathbf{C}_X \mathbf{a} = \sum_{i=1}^{N} \sum_{j=1}^{N} a_i a_j \text{cov}(X_i, X_j).$$

9.15 (f) For the covariance matrix given in Problem 9.12, find $\text{var}(X_1 + X_2 + X_3)$.

9.16 (t) Is it ever possible that $\text{var}(X_1 + X_2) = \text{var}(X_1)$ without X_2 being a constant?

9.17 (☺) (w) Which of the following matrices are not valid covariance matrices and why?

a. $\begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$ **b.** $\begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}$ **c.** $\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ **d.** $\begin{bmatrix} 2 & 1 \\ 0 & 1 \end{bmatrix}$

9.18 (f) A positive semidefinite matrix $\mathbf{A}$ must have $\det(\mathbf{A}) \geq 0$. Since a covariance matrix must be positive semidefinite, use this property to prove that the correlation coefficient satisfies $|\rho_{X_1,X_2}| \leq 1$. Hint: Consider a 2×2 covariance matrix.

9.19 (f) If a random vector $\mathbf{X}$ is transformed according to

$$\begin{aligned} Y_1 &= X_1 \\ Y_2 &= X_1 + X_2 \end{aligned}$$

and the mean of $\mathbf{X}$ is

$$E_{\mathbf{X}}[\mathbf{X}] = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$$

find the mean of $\mathbf{Y} = [Y_1\, Y_2]^T$.

9.20 (☺) (f) If the random vector $\mathbf{X}$ given in Problem 9.19 has a covariance matrix

$$\mathbf{C}_X = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

find the covariance matrix for $\mathbf{Y} = [Y_1\, Y_2]^T$.

9.21 (t) For $N = 2$ show that the covariance matrix may be defined as

$$\mathbf{C}_X = E_{\mathbf{X}}\left[(\mathbf{X} - E_{\mathbf{X}}[\mathbf{X}])(\mathbf{X} - E_{\mathbf{X}}[\mathbf{X}])^T\right].$$

Hint: Recall that the expected value of a matrix is the matrix of the expected values of its elements.

9.22 (t) In this problem you are asked to prove that if $\mathbf{Y} = \mathbf{AX}$, where both $\mathbf{X}$ and $\mathbf{Y}$ are $N \times 1$ random vectors and $\mathbf{A}$ is an $N \times N$ matrix, then $E_{\mathbf{Y}}[\mathbf{Y}] = \mathbf{A}E_{\mathbf{X}}[\mathbf{X}]$. If we let $[\mathbf{A}]_{ij}$ be the (i, j) element of $\mathbf{A}$, then you will need to prove that

$$[E_{\mathbf{Y}}[\mathbf{Y}]]_i = \sum_{j=1}^{N} [\mathbf{A}]_{ij}[E_{\mathbf{X}}[\mathbf{X}]]_j.$$

This is because if $\mathbf{b} = \mathbf{Ax}$, then $b_i = \sum_{j=1}^{N} a_{ij}x_j$, for $i = 1, 2, \ldots, N$ where b_i is the ith element of $\mathbf{b}$ and a_{ij} is the (i, j) element of $\mathbf{A}$.

9.23 (t) In this problem we prove that

$$E_{\mathbf{X}}[\mathbf{AG}(\mathbf{X})\mathbf{A}^T] = \mathbf{A}E_{\mathbf{X}}[\mathbf{G}(\mathbf{X})]\mathbf{A}^T$$

where $\mathbf{A}$ is an $N \times N$ matrix and $\mathbf{G}(\mathbf{X})$ is an $N \times N$ matrix whose elements are all functions of $\mathbf{X}$. To do so we note that if $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}$ are all $N \times N$ matrices then $\mathbf{D} = \mathbf{ABC}$ is an $N \times N$ matrix with (i, l) element

$$\begin{aligned}
[\mathbf{D}]_{il} &= \sum_{k=1}^{N} [\mathbf{AB}]_{ik}[\mathbf{C}]_{kl} \\
&= \sum_{k=1}^{N} \left(\sum_{j=1}^{N} [\mathbf{A}]_{ij}[\mathbf{B}]_{jk} \right) [\mathbf{C}]_{kl} \\
&= \sum_{k=1}^{N} \sum_{j=1}^{N} [\mathbf{A}]_{ij}[\mathbf{B}]_{jk}[\mathbf{C}]_{kl}.
\end{aligned}$$

Using this result and replacing $\mathbf{A}$ by itself, $\mathbf{B}$ by $\mathbf{G}(\mathbf{X})$, and $\mathbf{C}$ by $\mathbf{A}^T$ will allow the desired result to be proven.

9.24 (f) Prove (9.29) and (9.30) for the case of $N = 2$ by letting

$$\begin{aligned}
\mathbf{A} &= \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \\
\mathbf{b}_1 &= \begin{bmatrix} b_1^{(1)} \\ b_2^{(1)} \end{bmatrix} \qquad \mathbf{b}_2 = \begin{bmatrix} b_1^{(2)} \\ b_2^{(2)} \end{bmatrix} \\
\mathbf{d}_1 &= \begin{bmatrix} d_1^{(1)} \\ d_2^{(1)} \end{bmatrix} \qquad \mathbf{d}_2 = \begin{bmatrix} d_1^{(2)} \\ d_2^{(2)} \end{bmatrix}
\end{aligned}$$

and multiplying out all the matrices and vectors. Then, verify that the relationships are true by showing that the elements of the resultant $N \times N$ matrices are identical.

9.25 (c) Using MATLAB, find the eigenvectors and corresponding eigenvalues for the covariance matrix

$$\mathbf{C}_X = \begin{bmatrix} 26 & 6 \\ 6 & 26 \end{bmatrix}$$

To do so use the statement `[V Lambda]=eig(CX)`.

9.26 (⋱) (f,c) Find a linear transformation to decorrelate the random vector $\mathbf{X} = [X_1\, X_2]^T$ that has the covariance matrix

$$\mathbf{C}_X = \begin{bmatrix} 10 & 6 \\ 6 & 20 \end{bmatrix}.$$

What are the variances of the decorrelated random variables?

9.27 (t) Prove that an orthogonal matrix, i.e., one that has the property $\mathbf{U}^T = \mathbf{U}^{-1}$, *rotates* a vector $\mathbf{x}$ to a new vector $\mathbf{y}$. Do this by letting $\mathbf{y} = \mathbf{U}\mathbf{x}$ and showing that the length of $\mathbf{y}$ is the same as the length of $\mathbf{x}$. The length of a vector is defined to be $||\mathbf{x}|| = \sqrt{\mathbf{x}^T\mathbf{x}} = \sqrt{x_1^2 + x_2^2 + \cdots + x_N^2}$.

9.28 (t) Prove that if the random variables $X_1, X_2, \ldots, X_N$ are independent, then the joint characteristic function factors as

$$\phi_{X_1,X_2,\ldots,X_N}(\omega_1, \omega_2, \ldots \omega_N) = \phi_{X_1}(\omega_1)\phi_{X_2}(\omega_2)\ldots\phi_{X_N}(\omega_N).$$

Alternatively, if the joint characteristic function factors, what does this say about the random variables and why?

9.29 (f) For the random walk described in Example 9.5 find the mean and the variance of X_n as a function of n if $p = 3/4$. What do they indicate about the probable outcomes of $X_1, X_2, \ldots, X_N$?

9.30 (c) For the random walk of Problem 9.29 simulate several realizations of the random vector $\mathbf{X} = [X_1\, X_2 \ldots X_N]^T$ and plot these as x_n versus n for $n = 1, 2, \ldots, N = 50$. Does the appearance of the outcomes corroborate your results in Problem 9.29? Also, compare your results to those shown in Figure 9.3b.

9.31 (t) Prove the relationship given by (9.45) as follows. Consider the (i, j) element of $\mathbf{C}_X$, which is $\text{cov}(X_i, X_j) = E_{X_i,X_j}[X_iX_j] - E_{X_i}[X_i]E_{X_j}[X_j]$. Then, show that the latter is just the (i, j) element of the right-hand side of (9.45). Recall the definition of the expected value of a matrix/vector as the matrix/vector of expected values.

9.32 (c) A random vector is defined as $\mathbf{X} = [X_1 X_2 \ldots X_N]^T$, where each component is $X_i \sim \text{Ber}(1/2)$ and all the random variables are independent. Since the random variables are independent, the covariance matrix should be diagonal. Using MATLAB, generate realizations of $\mathbf{X}$ for $N = 10$ by using `x=floor(rand(10,1)+0.5)` to generate a single vector realization. Next generate multiple random vector realizations and use them to estimate the covariance matrix. Presumably the random numbers that MATLAB produces are "pseudo-independent" and hence "pseudo-uncorrelated". Does this appear to be the case? Hint: Use the MATLAB command `mesh(CXest)` to plot the estimated covariance matrix `CXest`.

9.33 (w) Prove that if X_1, X_2, X_3 are zero mean random variables, then $E[(X_3 - (X_1 + X_2))^2] = 0$ for the covariance matrix given by (9.48).

9.34 (t) In this problem we explain how to generate a computer realization of a random vector with a given covariance matrix. This procedure was used to produce the realizations shown in Figure 9.4a. For simplicity the desired $N \times 1$ random vector $\mathbf{X}$ is assumed to have a zero mean vector. The procedure is to first generate an $N \times 1$ random vector $\mathbf{U}$ whose elements are zero mean, uncorrelated random variables with unit variances so that its covariance matrix is $\mathbf{I}$. Then transform $\mathbf{U}$ according to $\mathbf{X} = \mathbf{BU}$, where $\mathbf{B}$ is an appropriate $N \times N$ matrix. The matrix $\mathbf{B}$ is obtained from the $N \times N$ matrix $\sqrt{\mathbf{\Lambda}}$ whose elements are obtained from the eigenvalue matrix $\mathbf{\Lambda}$ of $\mathbf{C}_X$ by taking the square root of the elements of $\mathbf{\Lambda}$, and $\mathbf{V}$, where $\mathbf{V}$ is the eigenvector matrix of $\mathbf{C}_X$, to form $\mathbf{B} = \mathbf{V}\sqrt{\mathbf{\Lambda}}$. Prove that the covariance matrix of $\mathbf{BU}$ will be $\mathbf{C}_X$.

9.35 (☺) (f) Using the results of Problem 9.34 find a matrix transformation $\mathbf{B}$ of $\mathbf{U} = [U_1 U_2]^T$, where $\mathbf{C}_U = \mathbf{I}$, so that $\mathbf{X} = \mathbf{BU}$ has the covariance matrix

$$\mathbf{C}_X = \begin{bmatrix} 4 & 1 \\ 1 & 4 \end{bmatrix}.$$

9.36 (☺) (c) Generate 30 realizations of a 2×1 random vector $\mathbf{X}$ that has a zero mean vector and the covariance matrix given in Problem 9.35. To do so use the results from Problem 9.35. For the random vector $\mathbf{U}$ assume that U_1 and U_2 are uncorrelated and have the same PMF

$$p_U[k] = \begin{cases} \frac{1}{2} & k = -1 \\ \frac{1}{2} & k = 1. \end{cases}$$

Note that the mean of $\mathbf{U}$ is zero and the covariance matrix of $\mathbf{U}$ is $\mathbf{I}$. Next estimate the covariance matrix $\mathbf{C}_X$ using your realizations and compare it to the true covariance matrix.