Chapter 14

Continuous N-Dimensional Random Variables

14.1 Introduction

This chapter extends the results of Chapters 10–13 for one and two continuous random variables to N continuous random variables. Our discussion will mirror Chapter 9 quite closely, the difference being the consideration of continuous rather than discrete random variables. Therefore, the descriptions will be brief and will serve mainly to extend the usual definitions for one and two jointly distributed continuous random variables to an N-dimensional random vector. One new concept that is introduced is the orthogonality principle approach to prediction of the outcome of a random variable based on the outcomes of several other random variables. This concept will be useful later when we discuss prediction of random processes in Chapter 18.

14.2 Summary

The probability of an event defined on an N-dimensional sample space is given by (14.1). The most important example of an N-dimensional PDF is the multivariate Gaussian PDF, which is given by (14.2). If the components of the multivariate Gaussian random vector are uncorrelated, then they are also independent as shown in Example 14.2. Transformations of random vectors yield the transformed PDF given by (14.5). In particular, linear tranformations of Gaussian random vectors preserve the Gaussian nature but change the mean vector and covariance matrix as discussed in Example 14.3. Expected values are described in Section 14.5 with the mean and variance of a linear combination of random variables given by (14.8) and (14.10), respectively. The sample mean random variable is introduced in Example 14.4. The joint moment is defined by (14.13) and the joint characteristic function by (14.15). Joint moments can be found from the characteristic function using (14.17). The PDF for a sum of independent and identically distributed random variables is conveniently determined using (14.22). The prediction of the outcome of a random variable based on a linear combination of the outcomes of other random variables is given by (14.24). The linear prediction coefficients are found by solving the set of simultaneous linear equations in (14.27). The orthogonality principle is summarized by (14.29) and illustrated in Figure 14.3. Section 14.9 describes the computer generation of a multivariate Gaussian random vector. Finally, section 14.10 applies the results of this chapter to the real-world problem of signal detection with the optimal detector given by (14.33).

14.3 Random Vectors and PDFs

An N-dimensional random vector will be denoted by either $(X_1, X_2, \ldots, X_N)$ or $\mathbf{X} = [X_1\, X_2 \ldots X_N]^T$. It is defined as a mapping from the original sample space of the experiment to a numerical sample space $\mathcal{S}_{X_1,X_2,\ldots,X_N} = R^N$. Hence, $\mathbf{X}$ takes on values in the N-dimensional Euclidean space 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 R^N. The number of possible values is uncountably infinite. As an example, we might observe the temperature on each of N successive days. Then, the elements of the random vector would be $X_1(s)$ = temperature on day 1, $X_2(s)$ = temperature on day 2, ..., $X_N(s)$ = temperature on day N, and each temperature measurement would take on a continuum of values.

To compute probabilities of events defined on $\mathcal{S}_{X_1,X_2,\ldots,X_N}$ we will define the N-dimensional joint PDF (or more succinctly just the PDF) as

$$p_{X_1,X_2,\ldots,X_N}(x_1, x_2, \ldots, x_N)$$

and sometimes use the more compact notation $p_{\mathbf{X}}(\mathbf{x})$. The usual properties of a joint PDF must be valid

$$\begin{aligned} p_{X_1,X_2,\ldots,X_N}(x_1, x_2, \ldots, x_N) &\geq 0 \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} p_{X_1,X_2,\ldots,X_N}(x_1, x_2, \ldots, x_N)dx_1\, dx_2 \ldots dx_N &= 1. \end{aligned}$$

Then the probability of an event A defined on R^N is given by

$$P[A] = \int \int \cdots \int_A p_{X_1,X_2,\ldots,X_N}(x_1, x_2, \ldots, x_N) dx_1\, dx_2 \ldots dx_N. \tag{14.1}$$

The most important example of an N-dimensional joint PDF is the *multivariate Gaussian* PDF. This PDF is the extension of the bivariate Gaussian PDF described at length in Chapter 12 (see (12.35)). It is given in vector/matrix form as

$$p_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{N/2} \det^{1/2}(\mathbf{C})} \exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \mathbf{C}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right] \tag{14.2}$$

where $\boldsymbol{\mu} = [\mu_1\, \mu_2 \ldots \mu_N]^T$ is the $N \times 1$ mean vector so that

$$E_{\mathbf{X}}[\mathbf{X}] = \begin{bmatrix} E_{X_1}[X_1] \\ E_{X_2}[X_2] \\ \vdots \\ E_{X_N}[X_N] \end{bmatrix} = \boldsymbol{\mu}$$

and $\mathbf{C}$ is the $N \times N$ covariance matrix defined as

$$\mathbf{C} = \begin{bmatrix} \operatorname{var}(X_1) & \operatorname{cov}(X_1, X_2) & \ldots & \operatorname{cov}(X_1, X_N) \\ \operatorname{cov}(X_2, X_1) & \operatorname{var}(X_2) & \ldots & \operatorname{cov}(X_2, X_N) \\ \vdots & \vdots & \ddots & \vdots \\ \operatorname{cov}(X_N, X_1) & \operatorname{cov}(X_N, X_2) & \ldots & \operatorname{var}(X_N) \end{bmatrix}.$$

Note that $\mathbf{C}$ is assumed to be positive definite and so it is invertible and has $\det(\mathbf{C}) > 0$ (see Appendix C). If the random variables have the multivariate Gaussian PDF, they are said to be *jointly Gaussian distributed.* Note that the covariance matrix can also be written as (see (Problem 9.21))

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

To denote a multivariate Gaussian PDF we will use the notation $\mathcal{N}(\boldsymbol{\mu}, \mathbf{C})$. Clearly, for $N = 2$ we have the bivariate Gaussian PDF. Evaluation of the probability of an event using (14.1) is in general quite difficult. Progress can, however, be made when A is a simple geometric region in R^N and $\mathbf{C}$ is a diagonal matrix. An example follows.

Example 14.1 – Probability of a point lying within a sphere

Assume $N = 3$ and let $\mathbf{X} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$. We will determine the probability that an outcome falls within a sphere of radius R. The event is then given by $A = \{(x_1, x_2, x_3) : x_1^2 + x_2^2 + x_3^2 \leq R^2\}$. This event might represent the probability that a particle with mass m and random velocity components V_x, V_y, V_z has a kinetic energy $\mathcal{E} = (1/2)m(V_x^2 + V_y^2 + V_z^2)$ less than a given amount. This modeling is used in the kinetic theory of gases [Resnick and Halliday 1966] and is known as the Maxwellian distribution. From (14.2) we have with $\boldsymbol{\mu} = \mathbf{0}$, $\mathbf{C} = \sigma^2\mathbf{I}$, and $N = 3$

$$\begin{aligned} P[A] &= \int\int\int_A \frac{1}{(2\pi)^{3/2}\det^{1/2}(\sigma^2\mathbf{I})} \exp\left[-\frac{1}{2}\mathbf{x}^T(\sigma^2\mathbf{I})^{-1}\mathbf{x}\right] dx_1\,dx_2\,dx_3 \\ &= \int\int\int_A \frac{1}{(2\pi\sigma^2)^{3/2}} \exp\left[-\frac{1}{2\sigma^2}(x_1^2 + x_2^2 + x_3^2)\right] dx_1\,dx_2\,dx_3 \end{aligned}$$

since $\det(\sigma^2\mathbf{I}) = (\sigma^2)^3$ and $(\sigma^2\mathbf{I})^{-1} = (1/\sigma^2)\mathbf{I}$. Next we notice that the region of integration is the inside of a sphere. As a result of this and the observation that the integrand only depends on the squared-distance of the point from the origin, a reasonable approach is to convert the Cartesian coordinates to spherical coordinates. Doing so produces the inverse transformation

$$\begin{aligned} x_1 &= r\cos\theta\sin\phi \\ x_2 &= r\sin\theta\sin\phi \\ x_3 &= r\cos\phi \end{aligned}$$

where $r \geq 0$, $0 \leq \theta < 2\pi$, $0 \leq \phi \leq \pi$. We must be sure to include in the integral over r, θ, ϕ the absolute value of the Jacobian determinant of the inverse transformation which is $r^2 \sin\phi$ (see Problem 14.5). Thus,

$$\begin{aligned} P[A] &= \int_0^R \int_0^\pi \int_0^{2\pi} \frac{1}{(2\pi\sigma^2)^{3/2}} \exp\left(-\frac{1}{2\sigma^2}r^2\right) r^2 \sin\phi\, d\theta\, d\phi\, dr \\ &= \int_0^R \int_0^\pi \frac{1}{(2\pi\sigma^2)^{3/2}} r^2 \exp\left(-\frac{1}{2\sigma^2}r^2\right) 2\pi \sin\phi\, d\phi\, dr \\ &= \int_0^R \frac{1}{(2\pi\sigma^2)^{3/2}} r^2 \exp\left(-\frac{1}{2\sigma^2}r^2\right) 2\pi \underbrace{(-\cos\phi)|_0^\pi}_{2}\, dr \\ &= \int_0^R \frac{4\pi}{(2\pi\sigma^2)^{3/2}} r^2 \exp\left(-\frac{1}{2\sigma^2}r^2\right) dr \\ &= \sqrt{\frac{2}{\pi\sigma^2}} \int_0^R \frac{r^2}{\sigma^2} \exp\left(-\frac{1}{2\sigma^2}r^2\right) dr. \end{aligned}$$

To evaluate the integral

$$I = \int_0^R \frac{r^2}{\sigma^2} \exp\left(-\frac{1}{2\sigma^2}r^2\right) dr$$

we use integration by parts (see Problem 11.7) with $U = r$ and hence $dU = dr$ and

$dV = (r/\sigma^2)\exp[-r^2/(2\sigma^2)]dr$ so that $V = -\exp[-r^2/(2\sigma^2)]$. Then

$$\begin{aligned}
I &= -r\exp\left[-\frac{1}{2}r^2/\sigma^2\right]\Big|_0^R + \int_0^R \exp\left[-\frac{1}{2}r^2/\sigma^2\right]dr \\
&= -R\exp\left[-\frac{1}{2}R^2/\sigma^2\right] + \sqrt{2\pi\sigma^2}\int_0^R \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{1}{2}r^2/\sigma^2\right]dr \\
&= -R\exp\left[-\frac{1}{2}R^2/\sigma^2\right] + \sqrt{2\pi\sigma^2}\,[Q(0) - Q(R/\sigma)].
\end{aligned}$$

Finally, we have that

$$\begin{aligned}
P[A] &= \sqrt{\frac{2}{\pi\sigma^2}}\left[-R\exp\left[-\frac{1}{2}R^2/\sigma^2\right] + \sqrt{2\pi\sigma^2}\,[Q(0) - Q(R/\sigma)]\right] \\
&= 1 - 2Q(R/\sigma) - \sqrt{\frac{2}{\pi\sigma^2}}R\exp\left(-\frac{1}{2}R^2/\sigma^2\right).
\end{aligned}$$

◇

The marginal PDFs are found by integrating out the other variables. For example, if $p_{X_1}(x_1)$ is desired, then

$$p_{X_1}(x_1) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} p_{X_1,X_2,\ldots,X_N}(x_1,x_2,\ldots,x_N)dx_2\,dx_3\ldots dx_N.$$

As an example, for the multivariate Gaussian PDF it can be shown that $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$, where $\sigma_i^2 = \mathrm{var}(X_i)$ (see Problem 14.16). Also, the lower dimensional joint PDFs are similarly found. To determine $p_{X_1,X_N}(x_1,x_N)$ for example, we use

$$p_{X_1,X_N}(x_1,x_N) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} p_{X_1,X_2,\ldots,X_N}(x_1,x_2,\ldots,x_N)dx_2\,dx_3\ldots dx_{N-1}.$$

The random variables are defined to be independent if the joint PDF factors into the product of the marginal PDFs as

$$p_{X_1,X_2,\ldots,X_N}(x_1,x_2,\ldots,x_N) = p_{X_1}(x_1)p_{X_2}(x_2)\ldots p_{X_N}(x_N). \tag{14.3}$$

An example follows.

Example 14.2 – Condition for independence of multivariate Gaussian random variables

If the covariance matrix for a multivariate Gaussian PDF is diagonal, then the random variables are not only uncorrelated but also independent as we now show. Assume that

$$\mathbf{C} = \mathrm{diag}\left(\sigma_1^2, \sigma_2^2, \ldots, \sigma_N^2\right)$$

then it follows that

$$\begin{aligned}\det(\mathbf{C}) &= \prod_{i=1}^{N} \sigma_i^2 \\ \mathbf{C}^{-1} &= \operatorname{diag}\left(\frac{1}{\sigma_1^2}, \frac{1}{\sigma_2^2}, \ldots, \frac{1}{\sigma_N^2}\right).\end{aligned}$$

Using these results in (14.2) produces

$$\begin{aligned}p_{\mathbf{X}}(\mathbf{x}) &= \frac{1}{(2\pi)^{N/2}\left(\prod_{i=1}^{N}\sigma_i^2\right)^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \operatorname{diag}\left(\frac{1}{\sigma_1^2}, \frac{1}{\sigma_2^2}, \ldots, \frac{1}{\sigma_N^2}\right)(\mathbf{x}-\boldsymbol{\mu})\right] \\ &= \frac{1}{\prod_{i=1}^{N}\sqrt{2\pi\sigma_i^2}} \exp\left[-\frac{1}{2}\sum_{i=1}^{N}(x_i-\mu_i)^2/\sigma_i^2\right] \\ &= \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left[-\frac{1}{2\sigma_i^2}(x_i-\mu_i)^2\right] \\ &= \prod_{i=1}^{N} p_{X_i}(x_i)\end{aligned}$$

where $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$. *Hence, if a random vector has a multivariate Gaussian PDF and the covariance matrix is diagonal, which means that the random variables are uncorrelated, then the random variables are also independent.*

◇

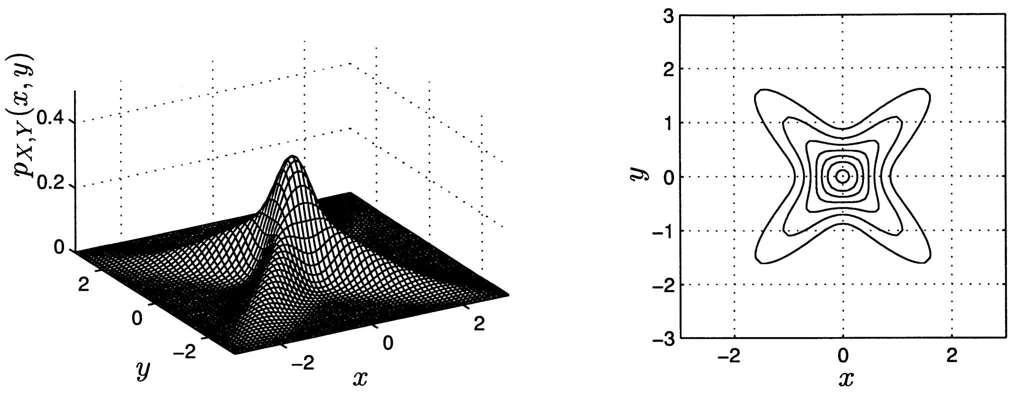
⚠ **Uncorrelated implies independence only for multivariate Gaussian PDF even if marginal PDFs are Gaussian!**

Consider the counterexample of a PDF for the random vector (X, Y) given by

$$\begin{aligned}p_{X,Y}(x,y) &= \frac{1}{2}\frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2(1-\rho^2)}(x^2-2\rho xy+y^2)\right] \\ &\quad + \frac{1}{2}\frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2(1-\rho^2)}(x^2+2\rho xy+y^2)\right] \qquad (14.4)\end{aligned}$$

for $0 < \rho < 1$. This PDF is shown in Figure 14.1 for $\rho = 0.9$. Clearly, the random variables are not independent. Yet, it can be shown that $X \sim \mathcal{N}(0, 1)$, $Y \sim \mathcal{N}(0, 1)$, and X and Y are uncorrelated (see Problem 14.7). The difference here is that the joint PDF is not a bivariate Gaussian PDF.

⚠



(a) Joint PDF

(b) Constant PDF contours

Figure 14.1: Uncorrelated but not independent random variables with Gaussian marginal PDFs.

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

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

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

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

14.4 Transformations

We consider the transformation from $\mathbf{X}$ to $\mathbf{Y}$ where

$$\begin{aligned} Y_1 &= g_1(X_1, X_2, \ldots, X_N) \\ Y_2 &= g_2(X_1, X_2, \ldots, X_N) \\ \vdots &\quad \vdots \\ Y_N &= g_N(X_1, X_2, \ldots, X_N) \end{aligned}$$

and the transformation is one-to-one. Hence $\mathbf{Y}$ is a continuous random vector having a joint PDF (due to the one-to-one property). If we wish to find the PDF of a subset of the Y_i's, then we need only first find the PDF of $\mathbf{Y}$ and then integrate out the undesired variables. The extension of (12.22) for obtaining the joint PDF of two transformed random variables is

$$p_{Y_1,Y_2,\ldots,Y_N}(y_1, y_2, \ldots, y_N)$$

$$= p_{X_1,X_2,\ldots,X_N}(\underbrace{g_1^{-1}(\mathbf{y})}_{x_1}, \underbrace{g_2^{-1}(\mathbf{y})}_{x_2}, \ldots, \underbrace{g_N^{-1}(\mathbf{y})}_{x_N}) \left|\det\left(\frac{\partial(x_1, x_2, \ldots, x_N)}{\partial(y_1, y_2, \ldots, y_N)}\right)\right| \qquad (14.5)$$

where

$$\frac{\partial(x_1, x_2, \ldots, x_N)}{\partial(y_1, y_2, \ldots, y_N)} = \begin{bmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} & \cdots & \frac{\partial x_1}{\partial y_N} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} & \cdots & \frac{\partial x_2}{\partial y_N} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial x_N}{\partial y_1} & \frac{\partial x_N}{\partial y_2} & \cdots & \frac{\partial x_N}{\partial y_N} \end{bmatrix}$$

is the inverse Jacobian matrix. An example follows.

Example 14.3 – Linear transformation of multivariate Gaussian random vector

If $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{C})$ and $\mathbf{Y} = \mathbf{GX}$, where $\mathbf{G}$ is an invertible $N \times N$ matrix, then we have from $\mathbf{y} = \mathbf{Gx}$ that

$$\begin{aligned} \mathbf{x} &= \mathbf{G}^{-1}\mathbf{y} \\ \frac{\partial \mathbf{x}}{\partial \mathbf{y}} &= \mathbf{G}^{-1}. \end{aligned}$$

Hence, using (14.5) and (14.2)

$$\begin{aligned} p_{\mathbf{Y}}(\mathbf{y}) &= p_{\mathbf{X}}(\mathbf{G}^{-1}\mathbf{y}) \left|\det(\mathbf{G}^{-1})\right| \\ &= \frac{1}{(2\pi)^{N/2}\det^{1/2}(\mathbf{C})} \exp\left[-\frac{1}{2}(\mathbf{G}^{-1}\mathbf{y} - \boldsymbol{\mu})^T\mathbf{C}^{-1}(\mathbf{G}^{-1}\mathbf{y} - \boldsymbol{\mu})\right] \frac{1}{|\det(\mathbf{G})|} \\ &= \frac{1}{(2\pi)^{N/2}\det^{1/2}(\mathbf{GCG}^T)} \exp\left[-\frac{1}{2}(\mathbf{y} - \mathbf{G}\boldsymbol{\mu})^T(\mathbf{GCG}^T)^{-1}(\mathbf{y} - \mathbf{G}\boldsymbol{\mu})\right] \end{aligned}$$

(see Section 12.7 for details of matrix manipulations) so that $\mathbf{Y} \sim \mathcal{N}(\mathbf{G}\boldsymbol{\mu}, \mathbf{GCG}^T)$. This result is the extension of Theorem 12.7 from 2 to N jointly Gaussian random variables. See also Problems 14.8 and 14.15 for the case where $\mathbf{G}$ is $M \times N$ with $M < N$. It is shown there that the same result holds.

◊

14.5 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 says that we define

$$E_{\mathbf{X}}[\mathbf{X}] = E_{X_1,X_2,\ldots,X_N}\left[\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_N \end{bmatrix}\right] = \begin{bmatrix} E_{X_1}[X_1] \\ E_{X_2}[X_2] \\ \vdots \\ E_{X_N}[X_N] \end{bmatrix}. \tag{14.6}$$

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 (12.28)) this is determined using

$$E_{X_1,X_2,\ldots,X_N}[g(X_1, X_2, \ldots, X_N)]$$
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} g(x_1, x_2, \ldots, x_N) p_{X_1,X_2,\ldots,X_N}(x_1, x_2, \ldots, x_N) dx_1\, dx_2 \ldots dx_N. \tag{14.7}$$

Some specific results of interest are the linearity of the expectation operator or

$$E_{X_1,X_2,\ldots,X_N}\left[\sum_{i=1}^{N} a_i X_i\right] = \sum_{i=1}^{N} a_i E_{X_i}[X_i] \tag{14.8}$$

and in particular if $a_i = 1$ for all i, then we have

$$E_{X_1,X_2,\ldots,X_N}\left[\sum_{i=1}^{N} X_i\right] = \sum_{i=1}^{N} E_{X_i}[X_i]. \tag{14.9}$$

The variance of a linear combination of random variables is given by

$$\operatorname{var}\left(\sum_{i=1}^{N} a_i X_i\right) = \sum_{i=1}^{N}\sum_{j=1}^{N} a_i a_i \operatorname{cov}(X_i, X_j) = \mathbf{a}^T \mathbf{C}_X \mathbf{a} \tag{14.10}$$

where $\mathbf{C}_X$ is the covariance matrix of $\mathbf{X}$ and $\mathbf{a} = [a_1\, a_2 \ldots a_N]^T$. The derivation of (14.10) is identical to that given in the proof of Property 9.2 for discrete random variables. If the random variables are *uncorrelated* so that the covariance matrix is diagonal or

$$\mathbf{C}_X = \operatorname{diag}(\operatorname{var}(X_1), \operatorname{var}(X_2) \ldots, \operatorname{var}(X_N))$$

then (see Problem 14.10)

$$\operatorname{var}\left(\sum_{i=1}^{N} a_i X_i\right) = \sum_{i=1}^{N} a_i^2 \operatorname{var}(X_i). \tag{14.11}$$

If furthermore, $a_i = 1$ for all i, then

$$\operatorname{var}\left(\sum_{i=1}^{N} X_i\right) = \sum_{i=1}^{N} \operatorname{var}(X_i). \tag{14.12}$$

An example follows.

Example 14.4 – Sample mean of independent and identically distributed random variables

Assume that $X_1, X_2, \ldots, X_N$ are independent random variables and each random variable has the same marginal PDF. When random variables have the same marginal PDF, they are said to be *identically distributed*. Hence, we are assuming that the random variables are *independent and identically distributed (IID)*. As a consequence of being identically distributed, $E_{X_i}[X_i] = \mu$ and $\operatorname{var}(X_i) = \sigma^2$ for all i. It is of interest to examine the mean and variance of the *random variable* that we obtain by averaging the X_i's together. This averaged random variable is

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

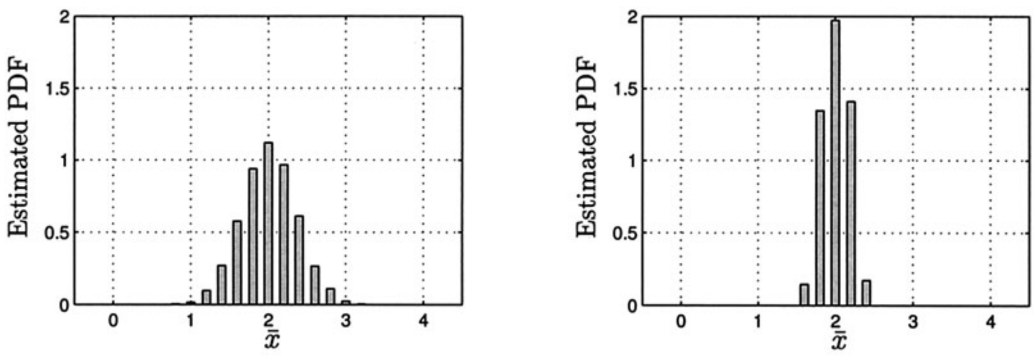
and is called the *sample mean random variable*. We have previously encountered the sample mean when referring to an average of a set of *outcomes* of a repeated experiment, which produced a number. Now, however, $\bar{X}$ is a function of the random variables $X_1, X_2, \ldots, X_N$ and so is a random variable itself. As such we may consider its probabilistic properties such as its mean and variance. The mean is from (14.8) with $a_i = 1/N$

$$E_{X_1,X_2,\ldots,X_N}[\bar{X}] = \frac{1}{N}\sum_{i=1}^{N} E_{X_i}[X_i] = \mu$$

and the variance is from (14.11) with $a_i = 1/N$ (since X_i's are independent and hence uncorrelated)

$$\begin{aligned}
\operatorname{var}(\bar{X}) &= \sum_{i=1}^{N} \frac{1}{N^2}\operatorname{var}(X_i) \\
&= \frac{1}{N^2}\sum_{i=1}^{N} \sigma^2 \\
&= \frac{\sigma^2}{N}.
\end{aligned}$$

Note that on the average the sample mean random variable will yield the value μ, which is the expected value of each X_i. Also as $N \to \infty$, $\operatorname{var}(\bar{X}) \to 0$, so that the PDF of $\bar{X}$ will become more and more concentrated about μ. In effect, as $N \to \infty$, we have that $\bar{X} \to \mu$. This says that the sample mean random variable will converge to the true expected value of X_i. An example is shown in Figure 14.2 in which the marginal PDF of each X_i is $\mathcal{N}(2,1)$. In the next chapter we will prove that $\bar{X}$ does indeed converge to $E_{X_i}[X_i] = \mu$.



(a) $N = 10$

(b) $N = 100$

Figure 14.2: Estimated PDF for sample mean random variable, $\bar{X}$.

◇

14.6 Joint Moments and the Characteristic Function

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

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

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} x_1^{l_1} x_2^{l_2} \ldots x_N^{l_N} p_{X_1,X_2,\ldots,X_N}(x_1, x_2, \ldots, x_N) dx_1\, dx_2 \ldots dx_N. \tag{14.13}$$

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

$$E_{X_1,X_2,\ldots,X_N}[X_1^{l_1} X_2^{l_2} \ldots X_N^{l_N}] = E_{X_1}[X_1^{l_1}] E_{X_2}[X_2^{l_2}] \ldots E_{X_N}[X_N^{l_N}]. \tag{14.14}$$

The joint characteristic function is defined as

$$\phi_{X_1,X_2,\ldots,X_N}(\omega_1, \omega_2, \ldots, \omega_N) = E_{X_1,X_2,\ldots,X_N}\left[\exp[j(\omega_1 X_1 + \omega_2 X_2 + \cdots + \omega_N X_N)]\right] \tag{14.15}$$

and is evaluated as

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

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\exp[j(\omega_1x_1+\omega_2x_2+\cdots+\omega_Nx_N)] \\ p_{X_1,X_2,\ldots,X_N}(x_1,x_2,\ldots,x_N)dx_1\,dx_2\ldots dx_N.$$

In particular, for independent random variables, we have (see Problem 14.13)

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

Also, the joint PDF can be found from the joint characteristic function using the inverse Fourier transform as

$$p_{X_1,X_2,\ldots,X_N}(x_1,x_2,\ldots,x_N)$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\phi_{X_1,X_2,\ldots,X_N}(\omega_1,\omega_2,\ldots,\omega_N) \\ \cdot\exp[-j(\omega_1x_1+\omega_2x_2+\cdots+\omega_Nx_N)]\frac{d\omega_1}{2\pi}\frac{d\omega_2}{2\pi}\ldots\frac{d\omega_N}{2\pi}. \quad (14.16)$$

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 can let $\omega_3=\omega_4=\cdots=\omega_N=0$ in the joint characteristic function to yield (see Problem 14.14)

$$p_{X_1,X_2}(x_1,x_2)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\underbrace{\phi_{X_1,X_2,\ldots,X_N}(\omega_1,\omega_2,0,\ldots,0)}_{\phi_{X_1,X_2}(\omega_1,\omega_2)}\exp[-j(\omega_1x_1+\omega_2x_2)]\frac{d\omega_1}{2\pi}\frac{d\omega_2}{2\pi}.$$

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

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

$$= \left.\frac{1}{j^{l_1+l_2+\cdots+l_N}}\frac{\partial^{l_1+l_2+\cdots+l_N}}{\partial\omega_1^{l_1}\partial\omega_2^{l_2}\ldots\partial\omega_N^{l_N}}\phi_{X_1,X_2,\ldots,X_N}(\omega_1,\omega_2,\ldots,\omega_N)\right|_{\omega_1=\omega_2=\cdots=\omega_N=0}. \quad (14.17)$$

An example follows.

Example 14.5 – Second-order joint moments for multivariate Gaussian PDF

In this example we derive the second-order moments $E_{X_iX_j}[X_iX_j]$ if $\mathbf{X}\sim\mathcal{N}(\mathbf{0},\mathbf{C})$. The characteristic function can be shown to be [Muirhead 1982]

$$\phi_{\mathbf{X}}(\boldsymbol{\omega})=\exp\left(-\frac{1}{2}\boldsymbol{\omega}^T\mathbf{C}\boldsymbol{\omega}\right)$$

where $\boldsymbol{\omega} = [\omega_1\,\omega_2 \ldots \omega_N]^T$. We first let

$$Q(\boldsymbol{\omega}) = \boldsymbol{\omega}^T \mathbf{C}\boldsymbol{\omega} = \sum_{m=1}^{N}\sum_{n=1}^{N} \omega_m\omega_n[\mathbf{C}]_{mn} \tag{14.18}$$

and note that it is a quadratic form (see Appendix C). Also, we let $[\mathbf{C}]_{mn} = c_{mn}$ to simplify the notation. Then from (14.17) with $l_i = l_j = 1$ and the other l's equal to zero, we have

$$E_{X_i,X_j}[X_iX_j] = \frac{1}{j^2}\frac{\partial^2}{\partial\omega_i\partial\omega_j}\exp\left(-\frac{1}{2}Q(\boldsymbol{\omega})\right)\bigg|_{\boldsymbol{\omega}=\mathbf{0}}.$$

Carrying out the partial differentiation produces

$$\begin{aligned}
\frac{\partial\exp[-(1/2)Q(\boldsymbol{\omega})]}{\partial\omega_i} &= -\frac{1}{2}\frac{\partial Q(\boldsymbol{\omega})}{\partial\omega_i}\exp\left(-\frac{1}{2}Q(\boldsymbol{\omega})\right)\\
\frac{\partial^2\exp[-(1/2)Q(\boldsymbol{\omega})]}{\partial\omega_i\partial\omega_j} &= \frac{1}{4}\frac{\partial Q(\boldsymbol{\omega})}{\partial\omega_i}\frac{\partial Q(\boldsymbol{\omega})}{\partial\omega_j}\exp\left(-\frac{1}{2}Q(\boldsymbol{\omega})\right)\\
&\quad -\frac{1}{2}\frac{\partial^2 Q(\boldsymbol{\omega})}{\partial\omega_i\partial\omega_j}\exp\left(-\frac{1}{2}Q(\boldsymbol{\omega})\right).
\end{aligned} \tag{14.19}$$

But

$$\begin{aligned}
\frac{\partial Q(\boldsymbol{\omega})}{\partial\omega_i}\bigg|_{\boldsymbol{\omega}=\mathbf{0}} &= \sum_{m=1}^{N}\sum_{n=1}^{N}\frac{\partial\omega_m\omega_n}{\partial\omega_i}c_{mn}\bigg|_{\boldsymbol{\omega}=\mathbf{0}} \qquad \text{(from (14.18))}\\
&= \sum_{m=1}^{N}\sum_{n=1}^{N}\left[\omega_m\frac{\partial\omega_n}{\partial\omega_i}c_{mn} + \omega_n\frac{\partial\omega_m}{\partial\omega_i}c_{mn}\right]\bigg|_{\boldsymbol{\omega}=\mathbf{0}}\\
&= 0
\end{aligned} \tag{14.20}$$

and also

$$\frac{\partial^2 Q(\boldsymbol{\omega})}{\partial\omega_i\partial\omega_j}\bigg|_{\boldsymbol{\omega}=\mathbf{0}} = \sum_{m=1}^{N}\sum_{n=1}^{N}\frac{\partial^2\omega_m\omega_n}{\partial\omega_i\partial\omega_j}c_{mn}\bigg|_{\boldsymbol{\omega}=\mathbf{0}}. \tag{14.21}$$

But

$$\begin{aligned}
\frac{\partial^2\omega_m\omega_n}{\partial\omega_i} &= \omega_m\frac{\partial\omega_n}{\partial\omega_i} + \omega_n\frac{\partial\omega_m}{\partial\omega_i}\\
&= \omega_m\delta_{ni} + \omega_n\delta_{mi}
\end{aligned}$$

where δ_{ij} is the Kronecker delta, which is defined to be 1 if $i = j$ and 0 otherwise. Hence

$$\frac{\partial^2\omega_m\omega_n}{\partial\omega_i\partial\omega_j} = \delta_{mj}\delta_{ni} + \delta_{nj}\delta_{mi}$$

and $\delta_{mj}\delta_{ni}$ equals 1 if $(m,n)=(j,i)$ and equals 0 otherwise, and $\delta_{nj}\delta_{mi}$ equals 1 if $(m,n)=(i,j)$ and equals 0 otherwise. Thus,

$$\begin{aligned}\left.\frac{\partial^2 Q(\boldsymbol{\omega})}{\partial\omega_i\partial\omega_j}\right|_{\boldsymbol{\omega}=\mathbf{0}} &= c_{ji}+c_{ij} \quad \text{(from (14.21))}\\ &= 2c_{ij} \qquad \text{(recall that } \mathbf{C}^T=\mathbf{C}).\end{aligned}$$

Finally, we have the expected result from (14.19) and (14.20) that

$$\begin{aligned}E_{X_i,X_j}[X_iX_j] &= \frac{1}{j^2}\left[-\frac{1}{2}\frac{\partial^2 Q(\boldsymbol{\omega})}{\partial\omega_i\partial\omega_j}\exp\left(-\frac{1}{2}Q(\boldsymbol{\omega})\right)\right]\Bigg|_{\boldsymbol{\omega}=\mathbf{0}}\\ &= \frac{1}{j^2}\left(-\frac{1}{2}\right)(2c_{ij})=c_{ij}=[\mathbf{C}]_{ij}.\end{aligned}$$

◇

Lastly, we extend the characteristic function approach to determining the PDF for a sum of IID random variables. Letting $Y=\sum_{i=1}^N X_i$, the characteristic function of Y is defined by

$$\phi_Y(\omega)=E_Y[\exp(j\omega Y)]$$

and is evaluated using (14.7) with $g(X_1,X_2,\ldots,X_N)=\exp[j\omega\sum_{i=1}^N X_i]$ (the real and imaginary parts are evaluated as separate integrals) as

$$\begin{aligned}\phi_Y(\omega) &= E_{X_1,X_2,\ldots,X_N}\left[\exp\left(j\omega\sum_{i=1}^N X_i\right)\right]\\ &= E_{X_1,X_2,\ldots,X_N}\left[\prod_{i=1}^N \exp(j\omega X_i)\right].\end{aligned}$$

Now using the fact that the X_i's are IID, we have that

$$\begin{aligned}\phi_Y(\omega) &= \prod_{i=1}^N E_{X_i}[\exp(j\omega X_i)] \qquad \text{(independence)}\\ &= \prod_{i=1}^N \phi_{X_i}(\omega)\\ &= [\phi_X(\omega)]^N \qquad \text{(identically distributed)}\end{aligned}$$

where $\phi_X(\omega)$ is the common characteristic function of the random variables. To finally obtain the PDF of the sum random variable we use an inverse Fourier transform to yield

$$p_Y(y)=\int_{-\infty}^{\infty}[\phi_X(\omega)]^N\exp(-j\omega y)\frac{d\omega}{2\pi}. \tag{14.22}$$

This formula will form the basis for the exploration of the PDF of a sum of IID random variables in Chapter 15. See Problems 14.17 and 14.18 for some examples of its use.

14.7 Conditional PDFs

The discussion of Section 9.7 of the definitions and properties of the conditional PMF also hold for the conditional PDF. To accommodate continuous random variables we need only replace the PMF notation of the "bracket" with that of the PDF notation of the "parenthesis." Hence, we do not pursue this topic further.

14.8 Prediction of a Random Variable Outcome

We have seen in Section 7.9 that the optimal linear prediction of the outcome of Y when $X = x$ is observed to occur is

$$\hat{Y} = E_Y[Y] + \frac{\text{cov}(X,Y)}{\text{var}(X)}(x - E_X[X]). \tag{14.23}$$

If (X, Y) has a bivariate Gaussian PDF, then the linear predictor is also the optimal predictor, amongst all linear *and* nonlinear predictors. We now extend these results to the prediction of a random variable after having observed the outcomes of *several* other random variables. In doing so the *orthogonality principle* will be introduced. Our discussions will assume only zero mean random variables, although the results are easily modified to yield the prediction for a nonzero mean random variable. To do so note that (14.23) can also be written as

$$\hat{Y} - E_Y[Y] = \frac{\text{cov}(X,Y)}{\text{var}(X)}(x - E_X[X]).$$

But if X and Y had been zero mean, then we would have obtained

$$\hat{Y} = \frac{\text{cov}(X,Y)}{\text{var}(X)}x.$$

It is clear that the modification from the zero mean case to the nonzero mean case is to replace each x_i by $x_i - E_{X_i}[X_i]$ and also $\hat{Y}$ by $\hat{Y} - E_Y[Y]$.

Now consider the $p + 1$ continuous random variables $\{X_1, X_2, \ldots, X_p, X_{p+1}\}$ and say we wish to predict X_{p+1} based on the knowledge of the outcomes of $X_1, X_2, \ldots, X_p$. Letting $X_1 = x_1, X_2 = x_2 \ldots, X_p = x_p$ be those outcomes, we consider the linear prediction

$$\hat{X}_{p+1} = \sum_{i=1}^{p} a_i x_i \tag{14.24}$$

where the a_i's are the *linear prediction coefficients*, which are to be determined. The optimal coefficients are chosen to minimize the mean square error (MSE)

$$\text{mse} = E_{X_1,X_2,\ldots,X_{p+1}}[(X_{p+1} - \hat{X}_{p+1})^2]$$

or written more explicitly as

$$\text{mse} = E_{X_1,X_2,\ldots,X_{p+1}}\left[\left(X_{p+1} - \sum_{i=1}^{p} a_i X_i\right)^2\right]. \tag{14.25}$$

We have used $\sum_{i=1}^{p} a_i X_i$, which is a random variable, as the predictor in order that the error measure be the *average* over all predictions. If we now differentiate the MSE with respect to a_1 we obtain

$$\frac{\partial E_{X_1,X_2,\ldots,X_{p+1}}[(X_{p+1} - \sum_{i=1}^{p} a_i X_i)^2]}{\partial a_1}$$

$$\begin{aligned} &= E_{X_1,X_2,\ldots,X_{p+1}}\left[\frac{\partial}{\partial a_1}(X_{p+1} - \sum_{i=1}^{p} a_i X_i)^2\right] \quad \text{(interchange integration and differentiation)} \\ &= E_{X_1,X_2,\ldots,X_{p+1}}\left[-2(X_{p+1} - \sum_{i=1}^{p} a_i X_i)X_1\right] = 0. \end{aligned} \tag{14.26}$$

This produces

$$E_{X_1,X_2,\ldots,X_{p+1}}[X_1 X_{p+1}] = E_{X_1,X_2,\ldots,X_{p+1}}\left[\sum_{i=1}^{p} a_i X_1 X_i\right]$$

or

$$E_{X_1,X_{p+1}}[X_1 X_{p+1}] = \sum_{i=1}^{p} a_i E_{X_1,X_i}[X_1 X_i].$$

Letting $c_{ij} = E_{X_i,X_j}[X_i X_j]$ denote the covariance (since the X_i's are zero mean) we have the equation

$$\sum_{i=1}^{p} c_{1i} a_i = c_{1,p+1}.$$

If we differentiate with respect to the other coefficients, similar equations are obtained. In all, there will be p simultaneous linear equations given by

$$\sum_{i=1}^{p} c_{ki} a_i = c_{k,p+1} \qquad k = 1, 2, \ldots, p$$

that need to be solved to yield the a_i's. These equations can be written in vector/matrix form as

$$\underbrace{\begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1p} \\ c_{21} & c_{22} & \cdots & c_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ c_{p1} & c_{p2} & \cdots & c_{pp} \end{bmatrix}}_{\mathbf{C}} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} = \underbrace{\begin{bmatrix} c_{1,p+1} \\ c_{2,p+1} \\ \vdots \\ c_{p,p+1} \end{bmatrix}}_{\mathbf{c}}. \tag{14.27}$$

We note that $\mathbf{C}$ is the covariance matrix of the random vector $[X_1\, X_2 \ldots X_p]^T$ and $\mathbf{c}$ is the vector of covariances between X_{p+1} and each X_i used in the predictor. The linear prediction coefficients are found by solving these linear equations. An example follows.

Example 14.6 – Linear prediction based on two random variable outcomes

Consider the prediction of X_3 based on the outcomes of X_1 and X_2 so that $\hat{X}_3 = a_1x_1 + a_2x_2$, where $p = 2$. If we know the covariance matrix of $\mathbf{X} = [X_1\, X_2\, X_3]^T$ say $\mathbf{C}_X$, then all the c_{ij}'s needed for (14.27) are known. Hence, suppose that

$$\mathbf{C}_X = \begin{bmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{bmatrix} = \begin{bmatrix} 1 & 2/3 & 1/3 \\ 2/3 & 1 & 2/3 \\ 1/3 & 2/3 & 1 \end{bmatrix}.$$

Thus, X_3 is correlated with X_2 with a correlation coefficient of 2/3 and X_3 is correlated with X_1 but with a smaller correlation coefficient of 1/3. Using (14.27) with $p = 2$ we must solve

$$\begin{bmatrix} 1 & 2/3 \\ 2/3 & 1 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 1/3 \\ 2/3 \end{bmatrix}.$$

By inverting the covariance matrix we have the solution

$$\begin{bmatrix} a_{1_{\text{opt}}} \\ a_{2_{\text{opt}}} \end{bmatrix} = \frac{1}{1-(2/3)^2} \begin{bmatrix} 1 & -2/3 \\ -2/3 & 1 \end{bmatrix} \begin{bmatrix} 1/3 \\ 2/3 \end{bmatrix} = \begin{bmatrix} -\frac{1}{5} \\ \frac{4}{5} \end{bmatrix}.$$

Due to the larger correlation of X_3 with X_2, the prediction coefficient a_2 is larger. Note that if the covariance matrix is $\mathbf{C}_X = \sigma^2\mathbf{I}$, then $c_{13} = c_{23} = 0$ and $a_{1_{\text{opt}}} = a_{2_{\text{opt}}} = 0$. This results in $\hat{X}_3 = 0$ or more generally for random variables with nonzero means, $\hat{X}_3 = E_{X_3}[X_3]$, as one might expect. See also Problem 14.24 to see how to determine the minimum value of the MSE.

◊

As another simple example, observe what happens if $p = 1$ so that we wish to predict X_2 based on the outcome of X_1. In this case we have that $\hat{X}_2 = a_1x_1$ and from (14.27), the solution for a_1 is $a_{1_{\text{opt}}} = c_{12}/c_{11} = \text{cov}(X_1, X_2)/\text{var}(X_1)$. Hence, $\hat{X}_2 = [\text{cov}(X_1, X_2)/\text{var}(X_1)]x_1$ and we recover our previous results for the bivariate case (see (14.23) and let $E_X[X] = E_Y[Y] = 0$) by replacing X_1 with X, x_1 with x, and X_2 with Y.

An interesting and quite useful interpretation of the linear prediction procedure can be made by reexamining (14.26). To simplify the discussion let $p = 2$ so that the equations to be solved are

$$\begin{aligned} E_{X_1,X_2,X_3}[(X_3 - a_1X_1 - a_2X_2)X_1] &= 0 \\ E_{X_1,X_2,X_3}[(X_3 - a_1X_1 - a_2X_2)X_2] &= 0. \end{aligned} \tag{14.28}$$

Let the predictor error be denoted by ϵ, which is explicitly $\epsilon = X_3 - a_1X_1 - a_2X_2$. Then (14.28) becomes

$$\begin{aligned} E_{X_1,X_2,X_3}[\epsilon X_1] &= 0 \\ E_{X_1,X_2,X_3}[\epsilon X_2] &= 0 \end{aligned} \tag{14.29}$$

which says that the optimal prediction coefficients a_1, a_2 are found by making the predictor error uncorrelated with the random variables used to predict X_3. Presumably if this were not the case, then some correlation would remain between the error and X_1, X_2, and this correlation could be exploited to reduce the error further (see Problem 14.23).

A geometric interpretation of (14.29) becomes apparent by considering X_1, X_2, and X_3 as vectors in a Euclidean space as depicted in Figure 14.3a. Since $\hat{X}_3 =$

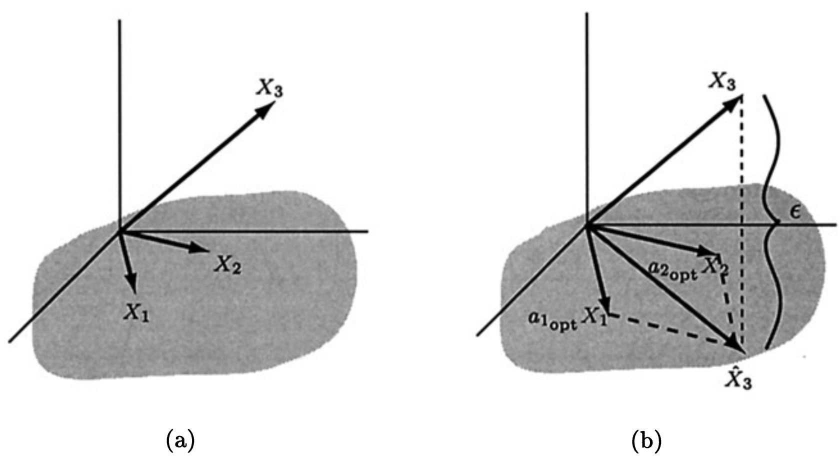


Figure 14.3: Geometrical interpretation of linear prediction.

$a_1X_1 + a_2X_2$, $\hat{X}_3$ can be any vector in the shaded region, which is the X_1-X_2 plane, depending upon the choice of a_1 and a_2. To minimize the error we should choose $\hat{X}_3$ as the *orthogonal projection* onto the plane as shown in Figure 14.3b. But this is equivalent to making the error vector ϵ orthogonal to any vector in the plane. In particular, then we have the requirement that

$$\begin{aligned} \epsilon &\perp X_1 \\ \epsilon &\perp X_2 \end{aligned} \tag{14.30}$$

where $\perp$ denotes orthogonality. To relate these conditions back to those of (14.29) we define two zero mean random variables X and Y to be orthogonal if $E_{X,Y}[XY] = 0$. Hence, we have that (14.30) is equivalent to

$$\begin{aligned} E_{X_1,X_2,X_3}[\epsilon X_1] &= 0 \\ E_{X_1,X_2,X_3}[\epsilon X_2] &= 0 \end{aligned}$$

or just the condition given by (14.29). (Since ϵ depends on (X_1, X_2, X_3), the expectation reflects this dependence.) This is called the *orthogonality principle.* It asserts that *to minimize the MSE the error "vector" should be orthogonal to each of the "data vectors" used to predict the desired "vector".* The "vectors" X and Y are defined to be orthogonal if $E_{X,Y}[XY] = 0$, which is equivalent to being uncorrelated since we have assumed zero mean random variables. See also Problem 14.22 for the one-dimensional case of the orthogonality principle.

14.9 Computer Simulation of Gaussian Random Vectors

The method described in Section 12.11 for generating a bivariate Gaussian random vector is easily extended to the N-dimensional case. To generate a realization of $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{C})$ we proceed as follows:

1. Perform a Cholesky decomposition of $\mathbf{C}$ to yield the $N \times N$ nonsingular matrix $\mathbf{G}$, where $\mathbf{C} = \mathbf{G}\mathbf{G}^T$.

2. Generate a realization $\mathbf{u}$ of an $N \times 1$ random vector $\mathbf{U}$ whose PDF is $\mathcal{N}(\mathbf{0}, \mathbf{I})$.

3. Form the realization of $\mathbf{X}$ as $\mathbf{x} = \mathbf{G}\mathbf{u} + \boldsymbol{\mu}$.

As an example, if $\boldsymbol{\mu} = \mathbf{0}$ and

$$\mathbf{C} = \begin{bmatrix} 1 & 2/3 & 1/3 \\ 2/3 & 1 & 2/3 \\ 1/3 & 2/3 & 1 \end{bmatrix} \tag{14.31}$$

then

$$\mathbf{G} = \begin{bmatrix} 1 & 0 & 0 \\ 0.6667 & 0.7454 & 0 \\ 0.3333 & 0.5963 & 0.7303 \end{bmatrix}.$$

We plot 100 realizations of $\mathbf{X}$ in Figure 14.4. The MATLAB code is given next.

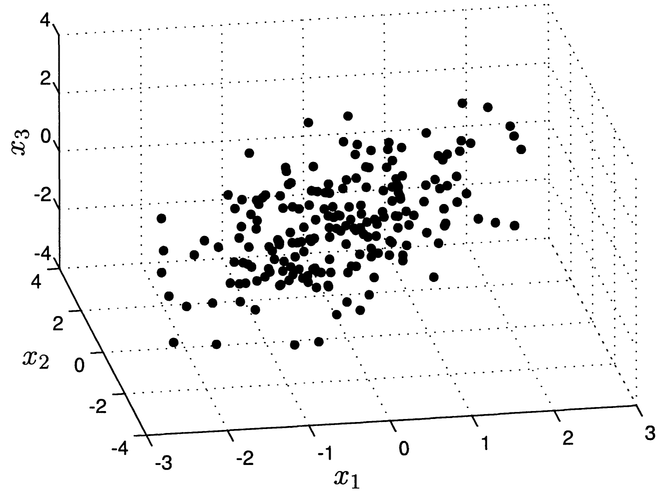


Figure 14.4: Realizations of 3 × 1 multivariate Gaussian random vector.

```
C=[1 2/3 1/3;2/3 1 2/3;1/3 2/3 1];
G=chol(C)'; % perform Cholesky decomposition
            % MATLAB produces C=A'*A so G=A'
M=200;
for m=1:M % generate realizations of x
  u=[randn(1,1) randn(1,1) randn(1,1)]';
  x(:,m)=G*u; % realizations stored as columns of 3 x 200 matrix
end
```

14.10 Real-World Example – Signal Detection

An important problem in sonar and radar is to be able to determine when an object, such as a submarine in sonar or an aircraft in radar, is present. To make this decision a pulse is transmitted into the water (sonar) or air (radar) and one looks to see if a reflected pulse from the object is returned. Typically, a digital computer is used to sample the received waveform in time and store the samples in memory for further processing. We will denote the received samples as $X_1, X_2, \ldots, X_N$. If there is no reflection, indicating no object is present, the received samples are due to noise only. If, however, there is a reflected pulse, also called an *echo*, the received samples will consist of a signal added to the noise. A standard model for the received samples is to assume that $X_i = W_i$, where $W_i \sim \mathcal{N}(0, \sigma^2)$ for noise only present and $X_i = s_i + W_i$ for a signal plus noise present. The noise samples W_i are usually also assumed to be independent and hence they are IID. With this modeling we can formulate the signal detection problem as the problem of deciding between the following two *hypotheses*

$$\begin{aligned} \mathcal{H}_W &: X_i = W_i \qquad i = 1, 2, \ldots, N \\ \mathcal{H}_{s+W} &: X_i = s_i + W_i \qquad i = 1, 2, \ldots, N. \end{aligned}$$

It can be shown that a good decision procedure is to choose the hypothesis for which the received data samples have the highest probability of occurring. In other words, if the received data is more probable when $\mathcal{H}_{s+W}$ is true than when $\mathcal{H}_W$ is true, we say that a signal is present. Otherwise, we decide that noise only is present. To implement this approach we let $p_{\mathbf{X}}(\mathbf{x}; \mathcal{H}_W)$ be the PDF when noise only is present and $p_{\mathbf{X}}(\mathbf{x}; \mathcal{H}_{s+W})$ be the PDF when a signal plus noise is present. Then we decide a signal is present if

$$p_{\mathbf{X}}(\mathbf{x}; \mathcal{H}_{s+W}) > p_{\mathbf{X}}(\mathbf{x}; \mathcal{H}_W). \tag{14.32}$$

But from the modeling we have that $\mathbf{X} = \mathbf{W} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$ for no signal present and $\mathbf{X} = \mathbf{s} + \mathbf{W} \sim \mathcal{N}(\mathbf{s}, \sigma^2\mathbf{I})$ when a signal is present. Here we have defined the signal vector as $\mathbf{s} = [s_1\, s_2 \ldots s_N]^T$. Hence, (14.32) becomes from (14.2)

$$\frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left[-\frac{1}{2\sigma^2}(\mathbf{x} - \mathbf{s})^T(\mathbf{x} - \mathbf{s})\right] > \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left[-\frac{1}{2\sigma^2}\mathbf{x}^T\mathbf{x}\right]$$

An equivalent inequality is

$$-(\mathbf{x} - \mathbf{s})^T(\mathbf{x} - \mathbf{s}) > -\mathbf{x}^T\mathbf{x}$$

since the constant $1/(2\pi\sigma^2)^{N/2}$ is positive and the exponential function increases with its argument. Expanding the terms we have

$$-\mathbf{x}^T\mathbf{x} + \mathbf{x}^T\mathbf{s} + \mathbf{s}^T\mathbf{x} - \mathbf{s}^T\mathbf{s} > -\mathbf{x}^T\mathbf{x}$$

and since $\mathbf{s}^T\mathbf{x} = \mathbf{x}^T\mathbf{s}$ we have

$$\mathbf{x}^T\mathbf{s} > \frac{1}{2}\mathbf{s}^T\mathbf{s}$$

or finally we decide a signal is present if

$$\sum_{i=1}^{N} x_i s_i > \frac{1}{2}\sum_{i=1}^{N} s_i^2. \tag{14.33}$$

This detector is called a *replica correlator* [Kay 1998] since it correlates the data $x_1, x_2, \ldots, x_N$ with a replica of the signal $s_1, s_2, \ldots, s_N$. The quantity on the right-hand-side of (14.33) is called the *threshold*. If the value of $\sum_{i=1}^N x_i s_i$ exceeds the threshold, the signal is declared as being present.

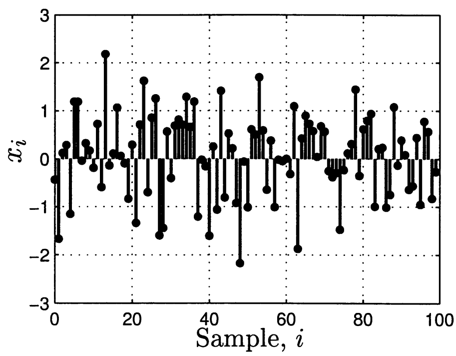
As an example, assume that the signal is a "DC level" pulse or $s_i = A$ for $i = 1, 2, \ldots, N$ and that $A > 0$. Then (14.33) reduces to

$$A\sum_{i=1}^{N} x_i > \frac{1}{2}NA^2$$

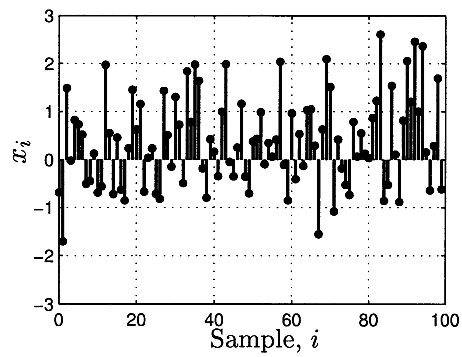
and since $A > 0$, we decide a signal is present if

$$\frac{1}{N}\sum_{i=1}^{N} x_i > \frac{A}{2}.$$

Hence, the sample mean is compared to a threshold of $A/2$. To see how this detector performs we choose $A = 0.5$ and $\sigma^2 = 1$. The received data samples are shown in Figure 14.5a for the case of noise only and in Figure 14.5b for the case of a signal plus noise. A total of 100 received data samples are shown. Note that the noise samples



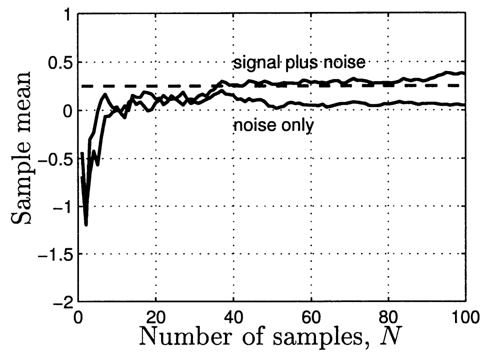
(a) Noise only



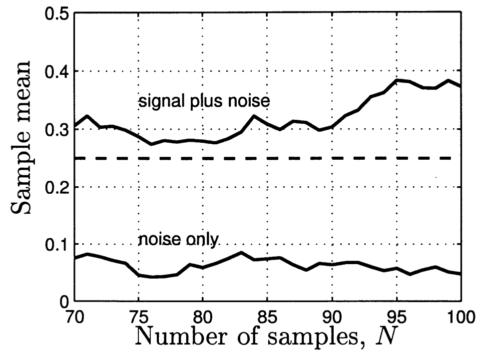
(b) Signal plus noise

Figure 14.5: Received data samples. Signal is $s_i = A = 0.5$ and noise consists of IID standard Gaussian random variables.

generated are different for each figure. The value of the sample mean $(1/N)\sum_{i=1}^{N} x_i$ is shown in Figure 14.6 versus the number of data samples N used in the averaging. For example, if $N = 10$, then the value shown is $(1/10)\sum_{i=1}^{10} x_i$, where x_i is found from the first 10 samples of Figure 14.5. To more easily observe the results they have been plotted as a continuous curve by connecting the points with straight lines. Also, the threshold of $A/2 = 0.25$ is shown as the dashed line. It is seen that as the number of data samples averaged increases, the sample mean converges to the mean of X_i (see also Example 14.4). When noise only is present, this becomes $E_X[X] = 0$ and when a signal is present, it becomes $E_X[X] = A = 0.5$. Thus by comparing the sample mean to the threshold of $A/2 = 0.25$ we should be able to decide if a signal is present or not most of the time (see also Problem 14.26).



(a) Total view



(b) Expanded view for $70 \leq N \leq 100$

Figure 14.6: Value of sample mean versus the number of data samples averaged.

References

Kay, S., *Fundamentals of Statistical Signal Processing: Detection Theory*, Prentice-Hall, Englewood Cliffs, NJ, 1998.

Muirhead, R.J., *Aspects of Multivariate Statistical Theory*, John Wiley & Sons, New York, 1982.

Resnick, R., D. Halliday, *Physics, Part I*, John Wiley & Sons, New York, 1966.

Problems

14.1 (⌣̈) **(w,f)** If $Y = X_1 + X_2 + X_3$, where $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{C})$ and

$$\boldsymbol{\mu} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

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

find the mean and variance of Y.

14.2 (w,c) If $[X_1\, X_2]^T \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$, find $P[X_1^2 + X_2^2 > R^2]$. Next, let $\sigma^2 = 1$ and $R = 1$ and lend credence to your result by performing a computer simulation to estimate the probability.

14.3 (f) Find the PDF of $Y = X_1^2 + X_2^2 + X_3^2$ if $\mathbf{X} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. Hint: Use the results of Example 14.1. Note that you should obtain the PDF for a χ_3^2 random variable.

14.4 (w) An airline has flights that depart according to schedule 95% of the time. This means that they depart late 1/2 hour or more 5% of the time due to mechanical problems, traffic delays, etc. (for less than 1/2 hour the plane is considered to be "on time"). The amount of time that the plane is late is modeled as an $\exp(\lambda)$ random variable. If a person takes a plane that makes two stops at intermediate destinations, what is the probability that he will be more than 1 1/2 hours late? Hint: You will need the PDF for a sum of independent exponential random variables.

14.5 (f) Consider the transformation from spherical to Cartesian coordinates. Show that the Jacobian has a determinant whose absolute value is equal to $r^2 \sin \phi$.

14.6 (⌣̈) (w) A large group of college students have weights that can be modeled as a $\mathcal{N}(150, 30)$ random variable. If 4 students are selected at random, what is the probability that they will all weigh more than 150 lbs?

14.7 (t) Prove that the joint PDF given by (14.4) has $\mathcal{N}(0, 1)$ marginal PDFs and that the random variables are uncorrelated. Hint: Use the known properties of the standard bivariate Gaussian PDF.

14.8 (t) Assume that $\mathbf{X} \sim \mathcal{N}(\mathbf{0}, \mathbf{C})$ for $\mathbf{X}$ an $N \times 1$ random vector and that $\mathbf{Y} = \mathbf{G}\mathbf{X}$, where $\mathbf{G}$ is an $M \times N$ matrix with $M < N$. If the characteristic function of $\mathbf{X}$ is $\phi_{\mathbf{X}}(\boldsymbol{\omega}) = \exp\left(-\frac{1}{2}\boldsymbol{\omega}^T \mathbf{C} \boldsymbol{\omega}\right)$, find the characteristic function of $\mathbf{Y}$. Use the following

$$\phi_{\mathbf{Y}}(\boldsymbol{\omega}) = E_{\mathbf{Y}}[\exp(j\boldsymbol{\omega}^T \mathbf{Y})] = E_{\mathbf{X}}[\exp(j\boldsymbol{\omega}^T \mathbf{G}\mathbf{X})] = E_{\mathbf{X}}[\exp(j(\mathbf{G}^T \boldsymbol{\omega})^T \mathbf{X})].$$

Based on your results conclude that $\mathbf{Y} \sim \mathcal{N}(\mathbf{0}, \mathbf{G}\mathbf{C}\mathbf{G}^T)$.

14.9 (⌣̈) (f) If $Y = X_1 + X_2 + X_3$, where $\mathbf{X} \sim \mathcal{N}(\mathbf{0}, \mathbf{C})$ and $\mathbf{C} = \operatorname{diag}(\sigma_1^2, \sigma_2^2, \sigma_3^2)$, find the PDF of Y. Hint: See Problem 14.8.

14.10 (f) Show that if $\mathbf{C}_X$ is a diagonal matrix, then $\mathbf{a}^T \mathbf{C}_X \mathbf{a} = \sum_{i=1}^N a_i^2 \operatorname{var}(X_i)$.

14.11 (c) Simulate a single realization of a random vector composed of IID random variables with PDF $X_i \sim \mathcal{N}(1, 2)$ for $i = 1, 2, \ldots, N$. Do this by repeating an experiment that successively generates $X \sim \mathcal{N}(1, 2)$. Then, find the outcome of the sample mean random variable and discuss what happens as N becomes large.

14.12 (⌣̈) (w,c) An $N \times 1$ random vector $\mathbf{X}$ has $E_{X_i}[X_i] = \mu$ and $\operatorname{var}(X_i) = i\sigma^2$ for $i = 1, 2, \ldots, N$. The components of $\mathbf{X}$ are independent. Does the sample mean random variable converge to μ as N becomes large? Carry out a computer simulation for this problem and explain your results.

14.13 (t) Prove that if $X_1, X_2, \ldots, X_N$ are independent random variables, then $\phi_{X_1,X_2,\ldots,X_N}(\omega_1, \omega_2, \ldots, \omega_N) = \prod_{i=1}^{N} \phi_{X_i}(\omega_i)$.

14.14 (t) Prove that $\phi_{X_1,X_2,\ldots,X_N}(\omega_1, \omega_2, 0, 0 \ldots, 0) = \phi_{X_1,X_2}(\omega_1, \omega_2)$.

14.15 (t) If $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{C})$ with $\mathbf{X}$ an $N \times 1$ random vector, prove that the characteristic function is

$$\phi_{\mathbf{X}}(\boldsymbol{\omega}) = \exp\left(j\boldsymbol{\omega}^T\boldsymbol{\mu} - \frac{1}{2}\boldsymbol{\omega}^T\mathbf{C}\boldsymbol{\omega}\right).$$

To do so note that the characteristic function of a random vector distributed according to $\mathcal{N}(\mathbf{0}, \mathbf{C})$ is $\exp\left(-\frac{1}{2}\boldsymbol{\omega}^T\mathbf{C}\boldsymbol{\omega}\right)$. With these results show that the PDF of $\mathbf{Y} = \mathbf{GX}$ for $\mathbf{G}$ an $M \times N$ matrix with $M < N$ is $\mathcal{N}(\mathbf{G}\boldsymbol{\mu}, \mathbf{GCG}^T)$.

14.16 (t) Prove that if $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{C})$ for $\mathbf{X}$ an $N \times 1$ random vector, then the marginal PDFs are $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$. Hint: Examine the PDF of $Y = \mathbf{e}_i^T\mathbf{X}$, where $\mathbf{e}_i$ is the $N \times 1$ vector whose elements are all zeros except for the ith element, which is a one. Also, make use of the results of Problem 14.15.

14.17 (f) Prove that if $X_i \sim \mathcal{N}(0, 1)$ for $i = 1, 2 \ldots, N$ and the X_i's are IID, then $\sum_{i=1}^{N} X_i^2 \sim \chi_N^2$. To do so first find the characteristic function of X_i^2. Hint: You will need the result that

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi c}} \exp\left(-\frac{1}{2}\frac{x^2}{c}\right) dx = 1$$

for c a *complex number*. Also, see Table 11.1.

14.18 (t) Prove that if $X_i \sim \exp(\lambda)$ and the X_i's are IID, then $\sum_{i=1}^{N} X_i$ has an Erlang PDF. Hint: See Table 11.1.

14.19 (⌣̈) (w,c) Find the mean and variance of the random variable

$$Y = \sum_{i=1}^{12} (U_i - 1/2)$$

where $U_i \sim \mathcal{U}(0, 1)$ and the U_i's are IID. Estimate the PDF of Y using a computer simulation and compare it to a standard Gaussian PDF. See Section 15.5 for a theoretical justification of your results.

14.20 (w) Three different voltmeters measure the voltage of a 100 volt source. The measurements can be modeled as random variables with

$$\begin{aligned} V_1 &\sim \mathcal{N}(100, 1) \\ V_2 &\sim \mathcal{N}(100, 10) \\ V_3 &\sim \mathcal{N}(100, 5). \end{aligned}$$

Is it better to average the results or just use the most accurate voltmeter?

14.21 (⌣̈) (f) If a 3×1 random vector has mean zero and covariance matrix

$$\mathbf{C}_X = \begin{bmatrix} 3 & 2 & 1 \\ 2 & 3 & 2 \\ 1 & 2 & 3 \end{bmatrix}$$

find the optimal prediction of X_3 given that we have observed $X_1 = 1$ and $X_2 = 2$.

14.22 (t) Consider the prediction of the random variable Y based on observing that $X = x$. Assuming (X, Y) is a zero mean random vector, we propose using the linear prediction $\hat{Y} = ax$. Determine the optimal value of a (being the value that minimizes the MSE) by using the orthogonality principle. Explain your results by drawing a diagram.

14.23 (f) If a 3×1 random vector $\mathbf{X}$ has a zero mean and covariance matrix

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

determine the optimal linear prediction of X_3 based on the observed outcomes of X_1 and X_2. Why is $a_{1_{\text{opt}}} = 0$? Hint: Consider the covariance between $\epsilon = X_3 - \rho X_2$, which is the predictor error for X_3 based on observing only X_2, and X_1.

14.24 (⌣̈) (t,f) Explain why the minimum MSE of the predictor $\hat{X}_3 = a_{1_{\text{opt}}} X_1 + a_{2_{\text{opt}}} X_2$ is

$$\begin{aligned} \text{mse}_{\text{min}} &= E_{X_1,X_2,X_3}\left[(X_3 - a_{1_{\text{opt}}} X_1 - a_{2_{\text{opt}}} X_2)^2\right] \\ &= E_{X_1,X_2,X_3}\left[(X_3 - a_{1_{\text{opt}}} X_1 - a_{2_{\text{opt}}} X_2) X_3\right] \\ &= c_{33} - a_{1_{\text{opt}}} c_{13} - a_{2_{\text{opt}}} c_{23}. \end{aligned}$$

Next use this result to find the minimum MSE for Example 14.6.

14.25 (⌣) **(c)** Use a computer simulation to generate realizations of the random vector $\mathbf{X}$ described in Example 14.6. Then, predict X_3 based on the outcomes of X_1 and X_2 and plot the true realizations and the predictions. Finally, estimate the average predictor error and compare your results to the theoretical minimum MSE obtained in Problem 14.24.

14.26 (w) For the signal detection example described in Section 14.9 prove that the probability of saying a signal is present when indeed there is one goes to 1 as $A \to \infty$.

14.27 (c) Generate on a computer 1000 realizations of the two different random variables $X_W \sim \mathcal{N}(0, 1)$ and $X_{s+W} \sim \mathcal{N}(0.5, 1)$. Next plot the outcomes of the sample mean random variable versus N, the number of successive samples averaged, or $\bar{x}_N = (1/N) \sum_{i=1}^{N} x_i$. What can you say about the sample means as N becomes large? Explain what this has to do with signal detection.