Chapter 5

Second-Order Analysis of Random Vectors

5.1 Introduction

In Chapter 4, we developed a characterization of the properties of pairs of random variables X, Y defined on the same probability space $(\Omega, \mathcal{E}, \mathbb{P})$ by defining either a joint probability mass function (PMF) or a joint probability density function (PDF), which can be used to compute probabilities of joint events and expectations of functions of the random variables. Using the joint PMF or joint PDF, we computed statistics such as the expected value of a function $g(X, Y)$.

In this chapter, we focus on second order statistics of a pair of random variables X, Y. These statistics generalize the concepts of variance and standard deviation to pairs of random variables, and are easily computed from sample data. We describe how these statistics change for linear or affine transformations of the pair X, Y. We study the special case of jointly Gaussian random variables X, Y, where the joint PDF is entirely described in terms of its second order statistics, and show special properties of jointly Gaussian random variables that make them suitable models for problems in estimation and detection. We conclude the chapter with a generalization of second order statistics to random vectors involving 2 or more random variables.

5.2 Covariance and Correlation

Consider a pair of random variables X, Y defined on a probability space $(\Omega, \mathcal{E}, \mathbb{P})$. If discrete, these random variables are characterized by a joint PMF $P_{X,Y}(x, y)$ and marginal PMFs $P_X(x), P_Y(y)$ derived from the joint PMF by

$$P_X(x) = \sum_{y \in R_Y} P_{X,Y}(x, y); \quad P_Y(y) = \sum_{x \in R_X} P_{X,Y}(x, y).$$

If X, Y are jointly continuous, the random variables are characterized by the joint PDF $f_{X,Y}(x, y)$, and marginal PDFs $f_X(x), f_Y(y)$ computed as:

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\ dy; \quad f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\ dx.$$

Using the marginal PMFs or PDFs, we can compute the means of X and Y, as $\mathbb{E}[X], \mathbb{E}[Y]$. We also compute the variance of each of the random variables X, Y as

$$\sigma_X^2 = \mathbb{E}\big[\big(X - \mathbb{E}[X]\big)^2\big] = \mathbb{E}\big[X^2\big] - \big(\mathbb{E}[X]\big)^2,$$

$$\sigma_Y^2 = \mathbb{E}\big[\big(Y - \mathbb{E}[Y]\big)^2\big] = \mathbb{E}\big[Y^2\big] - \big(\mathbb{E}[Y]\big)^2.$$

These variances measure how much each of the random variables deviates from their average values. However, as statistics, they provide no information as to how the deviations of the random variables depend on each other.

To capture that information, we define several joint statistics for the random variables X, Y. First, we define the **cross-correlation** between X and Y as $\mathbb{E}[XY]$. An important property of the cross-correlation is

$$\Big(\mathbb{E}[XY]\Big)^2 \le \mathbb{E}[X^2]\mathbb{E}[Y^2].$$

This follows from well-known Cauchy-Schwarz inequality, which states that, for functions $f(x), g(x)$ with finite square integrals,

$$\Big|\int_{-\infty}^{\infty} f(x)g(x)\,dx\Big| \le \Big(\int_{-\infty}^{\infty} f(x)^2\,dx\Big)^{1/2}\Big(\int_{-\infty}^{\infty} g(x)^2\,dx\Big)^{1/2}.$$

Similarly, for square summable sequences x_n, y_n,

$$\Big|\sum_{n=1}^{\infty} x_n y_n\Big| \le \Big(\sum_{n=1}^{\infty} x_n^2\Big)^{1/2}\Big(\sum_{n=1}^{\infty} y_n^2\Big)^{1/2}.$$

For continuous random variables, this implies

$$\begin{aligned}\Big|\mathbb{E}[XY]\Big| &= \Big|\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy f_{X,Y}(x,y)\,dx\,dy\Big| = \Big|\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \big(x f_{X,Y}(x,y)^{\frac{1}{2}}\big)\big(y f_{X,Y}(x,y)^{\frac{1}{2}}\big)\,dx\,dy\Big| \\ &\le \Big(\int_{-\infty}^{\infty} x^2 f_{X,Y}(x,y)\,dx\,dy\Big)^{1/2}\Big(\int_{-\infty}^{\infty} y^2 f_{X,Y}(x,y)\,dx\,dy\Big)^{1/2} = \big(\mathbb{E}[X^2]\big)^{1/2}\big(\mathbb{E}[Y^2]\big)^{1/2}\end{aligned}$$

The cross-correlation depends on the expected value of the individual random variables. To eliminate the dependence on the mean of the random variables, we define the **covariance** of random variables X and Y as

$$\mathsf{Cov}[X,Y] = \mathbb{E}\big[\big(X-\mathbb{E}[X]\big)\big(Y-\mathbb{E}[Y]\big)\big].$$

Intuitively, this captures how X and Y vary together with respect to their expected values. Unlike variances, the covariance between two random variables can be negative. A negative covariance indicates that, when X is greater than its mean $\mathbb{E}[X]$, Y is likely to be less than its mean $\mathbb{E}[Y]$. The covariance will be an important part of how we can estimate the value of one variable (e.g. Y) based on measurements of the other variable (X).

Since X, Y are real-valued random variables, $\mathsf{Cov}[X,Y] = \mathsf{Cov}[Y,X]$. As is the case for variances, there is a useful formula for computing covariances from cross-correlations:

$$\begin{aligned}\mathsf{Cov}[X,Y] &= \mathbb{E}\big[\big(X-\mathbb{E}[X]\big)\big(Y-\mathbb{E}[Y]\big)\big] = \mathbb{E}\big[XY - \mathbb{E}[X]Y - X\mathbb{E}[Y] + \mathbb{E}[X]\mathbb{E}[Y]\big] \\ &= \mathbb{E}[XY] - \mathbb{E}\big[\mathbb{E}[X]Y\big] - \mathbb{E}\big[X\mathbb{E}[Y]\big] + \mathbb{E}\big[\mathbb{E}[X]\mathbb{E}[Y]\big] \\ &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] - \mathbb{E}[X]\mathbb{E}[Y] + \mathbb{E}[X]\mathbb{E}[Y] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]\end{aligned}$$

Using the Cauchy-Schwarz inequality as before, we get the following:

$$|\mathsf{Cov}[X,Y]| \le \sqrt{\mathsf{Var}[X]\mathsf{Var}[Y]}.$$

Using this inequality, we define the **correlation coefficient** $\rho_{X,Y}$ between two random variables X, Y as

$$\rho_{X,Y} = \frac{\mathsf{Cov}[X,Y]}{\sqrt{\mathsf{Var}[X]\mathsf{Var}[Y]}}$$

The correlation coefficient has magnitude less than or equal to 1, so its range is in $[-1,1]$.

Another way of interpreting the correlation coefficient is that it is the covariance of the normalized random variables $F = \frac{X}{\sqrt{\mathsf{Var}[X]}}$ and $G = \frac{Y}{\sqrt{\mathsf{Var}[Y]}}$. Normalizing each of the random variables by dividing by their standard deviation results in random variables F and G with variance 1. This normalization is used extensively in data science and statistics to reduce the effects of measurement units for feature values.

Example 5.1
Let X, Y be a pair of random variables, and define $Z = X + Y$. Then,

$$\begin{aligned}
\mathbb{E}[Z] &= \mathbb{E}[X] + \mathbb{E}[Y] \qquad \text{(linearity of expectation)}\\
\mathbb{E}[Z^2] &= \mathbb{E}\big[(X+Y)^2\big] = \mathbb{E}[X^2 + 2XY + Y^2] = \mathbb{E}[X^2] + 2\mathbb{E}[XY] + \mathbb{E}[Y^2]\\
&= \mathbb{E}[X]^2 + \mathsf{Var}[X] + 2\Big(\mathbb{E}[X]\mathbb{E}[Y] + \mathsf{Cov}[X,Y]\Big) + \mathbb{E}[Y]^2 + \mathsf{Var}[Y] \qquad \text{(definitions of variance, covariance)}\\
&= \Big(\mathbb{E}[X]^2 + 2\mathbb{E}[X]\mathbb{E}[Y] + \mathbb{E}[Y]^2\Big) + \mathsf{Var}[X] + 2\mathsf{Cov}[X,Y] + \mathsf{Var}[Y]\\
&= \mathbb{E}[Z]^2 + \mathsf{Var}[X] + 2\mathsf{Cov}[X,Y] + \mathsf{Var}[Y]\\
\mathsf{Var}[Z] &= \mathbb{E}[Z^2] - \mathbb{E}[Z]^2 = \mathsf{Var}[X] + 2\mathsf{Cov}[X,Y] + \mathsf{Var}[Y]
\end{aligned}$$

This provides a quick way of calculating the covariance of a sum of random variables. The result does not depend on the mean of the random variables.

Example 5.2
Can the correlation coefficient have magnitude 1? Let X be a random variable, and let $Y = -3X + 1$. Then,

$$\begin{aligned}
\mathsf{Var}[Y] &= (-3)^2\mathsf{Var}[X] = 9\mathsf{Var}[X]\\
\mathsf{Cov}[X,Y] &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = \mathbb{E}[-3X^2 + X] - \mathbb{E}[X]\mathbb{E}[-3X+1] = -3\big(\mathbb{E}[X^2] - \mathbb{E}[X]^2\big) = -3\mathsf{Var}[X]\\
\rho_{X,Y} &= \frac{\mathsf{Cov}[X,Y]}{\sqrt{\mathsf{Var}[X]\mathsf{Var}[Y]}} = \frac{-3\mathsf{Var}X}{\sqrt{9\mathsf{Var}[X]^2}} = -1
\end{aligned}$$

When the magnitude of the correlation coefficient is either 1 or -1, it usually indicates a linear dependence between the two variables X, Y. Notice that, in this case, the correlation coefficient has a negative sign, suggesting a negative linear dependence.

Note also that the correlation coefficient is a scale-independent measure of how the random variables depend on each other. Thus, the scale factor of -3 between X and Y only affects the correlation coefficient by its sign, not its magnitude.

Example 5.3
Consider a pair of jointly continuous random variables X, Y with $f_{X,Y}(x,y)$ given as

$$f_{X,Y}(x,y) = \begin{cases} xy & 0 \le x \le 1, 0 \le y \le 2,\\ 0 & \text{otherwise.}\end{cases}$$

The marginal distributions are given as

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy = \begin{cases} \int_0^2 xy\,dy = 2x & x \in [0,1]\\ 0 & \text{otherwise.}\end{cases}$$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx = \begin{cases} \int_0^1 xy\,dx = \frac{y}{2} & y \in [0,2]\\ 0 & \text{otherwise.}\end{cases}$$

Note that X, Y are independent, as the range $R_{X,Y} = R_X \times R_Y$ and thus $f_{X,Y}(x,y) = f_X(x)f_Y(y)$.

Using these densities, we compute the first and second order statistics as follows:

$$\begin{aligned}
\mathbb{E}[X] &= \int_{-\infty}^{\infty} x f_X(x)\,dx = \int_0^1 2x^2\,dx = \frac{2}{3}\\
\mathbb{E}[Y] &= \int_{-\infty}^{\infty} y f_Y(y)\,dx = \int_0^2 \frac{y^2}{2}\,dy = \frac{4}{3}\\
\mathsf{Var}[X] &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \int_0^1 2x^3\,dx - \frac{4}{9} = \frac{1}{2} - \frac{4}{9} = \frac{1}{18}\\
\mathsf{Var}[Y] &= \mathbb{E}[Y^2] - \mathbb{E}[Y]^2 = \int_0^2 \frac{y^3}{2}\,dy - \frac{16}{9} = 2 - \frac{16}{9} = \frac{2}{9}\\
\mathbb{E}[XY] &= \mathbb{E}[X]\mathbb{E}[Y] = \frac{8}{9} \quad \text{(because of independence.)}\\
\mathsf{Cov}[X,Y] &= 0, \quad \rho_{X,Y} = 0
\end{aligned}$$

When two random variables are independent, their covariance is 0. The converse is not true, as we will see later.

Two random variables X and Y are **uncorrelated** if $\mathsf{Cov}[X, Y] = 0$ (or $\rho_{X,Y} = 0$).

- If X and Y are uncorrelated, we have that $\mathsf{Var}[X + Y] = \mathsf{Var}[X] + \mathsf{Var}[Y]$ and $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$.
- Independence of X and Y implies that they are uncorrelated. However, uncorrelated X and Y need not be independent.

To clarify, if X, Y are independent, then, for bounded functions f, g, we have

$$\begin{aligned}\mathbb{E}[f(X)g(Y)] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x)g(y)f_{X,Y}(x,y)\,dx\,dy \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x)g(y)f_X(x)f_Y(y)\,dx\,dy \text{ (independence of PDF)} \\ &= \Big(\int_{-\infty}^{\infty} f(x)f_X(x)\,dx\Big)\Big(\int_{-\infty}^{\infty} g(y)f_Y(y)\,dy\Big) = \mathbb{E}[f(X)]\mathbb{E}[g(Y)]\end{aligned}$$

The converse of this is also true: if $\mathbb{E}[f(X)g(Y)] = \mathbb{E}[f(X)]\mathbb{E}[g(Y)]$ for any bounded functions f, g, then X and Y are independent. However, X, Y are uncorrelated if and only if $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$. Thus, the condition for uncorrelated random variables involves only linear functions of X, Y, whereas the condition for independence must hold for the broader class of bounded nonlinear functions of X, Y.

Furthermore, if X, Y are independent, then $f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{f_X(x)f_Y(y)}{f_Y(y)} = f_X(x)$. Hence, $\mathbb{E}[X|Y = y] = \mathbb{E}[X]$ for all $y \in R_Y$. Similarly, $\mathbb{E}[Y|X = x] = \mathbb{E}[Y]$ for all $x \in R_X$. Independence is a *strong* property of the underlying densities of the random variables, while uncorrelatedness is only a property of second order statistics.

One of the interesting properties of uncorrelated random variables X, Y is that, if $Z = X + Y$, then $\mathsf{Var}[Z] = \mathsf{Var}[X] + \mathsf{Var}[Y]$. This is because, as derived in Example 5.1,

$$\mathsf{Var}[Z] = \mathsf{Var}[X] + \mathsf{Var}[Y] + 2\mathsf{Cov}[X, Y] = \mathsf{Var}[X] + \mathsf{Var}[Y],$$

since $\mathsf{Cov}[X, Y] = 0$ because X, Y are uncorrelated. This generalizes to arbitrary sums, so that the variance of a sum of uncorrelated random variables is the sum of the variances of the individual random variables.

Two random variables X and Y are **orthogonal** if and only if $\mathbb{E}[XY] = 0$. If X and Y are orthogonal, $\mathbb{E}[(X+Y)^2] = \mathbb{E}[X^2]+\mathbb{E}[Y^2]$. Note that orthogonal and uncorrelated random variables are different concepts. If two random variables are both orthogonal and uncorrelated, then the mean of at least one must be zero. For zero mean random variables, orthogonality and uncorrelatedness are equivalent. For instance, the random variables X, Y in Example 5.3 are independent, and thus uncorrelated. However, they are not orthogonal, because neither X nor Y has zero mean.

Example 5.4

Consider a pair of discrete random variables X, Y with joint PMF given by the table on the right. Are X, Y independent? Are X, Y uncorrelated? What is the covariance of X, Y?

$P_{XY}(x,y)$		y		
		0	1	2
x	0	0.01	0.00	0.00
	1	0.09	0.09	0.00
	2	0.00	0.00	0.81

With respect to independence, the answer is clearly not. Note that $P_{X,Y}(0, 1) = 0$, but $P_X(0) = 0.01$ and $P_Y(1) = 0.09$.

Are X, Y uncorrelated? We compute $\mathbb{E}[X] = 0.18 + 2 \cdot 0.81 = 1.80$, and $\mathbb{E}[Y] = 0.09 + 2 \cdot 0.81 = 1.71$. We then compute

$$\mathbb{E}[XY] = 0.09 \cdot 1 \cdot 1 + 0.81 \cdot 2 \cdot 2 = 3.33 \neq (1.71) \cdot (1.89)$$

Hence, they are not uncorrelated.

The covariance $\mathsf{Cov}[X, Y] = 3.33 - (1.71) \cdot (1.89) \approx 0.252$.

Example 5.5

Consider a pair of discrete random variables X, Y with joint PMF given by the table on the right. What are the means and variances of X, Y? Are X, Y independent?

$P_{XY}(x,y)$		$y=0$	$y=1$	$y=2$	$y=3$
x	0	0.06	0.18	0.24	0.12
	1	0.04	0.12	0.16	0.08

We compute the marginal PMFs by doing column and row sums to get

$$P_X(0) = 0.6, P_X(1) = 0.4.$$

$$P_Y(0) = 0.1, P_Y(1) = 0.3, P_Y(2) = 0.4, P_Y(3) = 0.2.$$

With this, we compute $\mathbb{E}[X] = 0.6 \cdot 0 + 0.4 \cdot 1 = 0.4$; similarly, $\mathbb{E}[X^2] = 0.6 \cdot 0^2 + 0.4 \cdot 1^2 = 0.4$. Thus, Var$[X] = 0.4 - (0.4)^2 = 0.24$.

For Y, $\mathbb{E}[Y] = 0 \cdot 0.1 + 1 \cdot 0.3 + 2 \cdot 0.4 + 3 \cdot 0.2 = 1.7$. Similarly, $\mathbb{E}[Y^2] = 0^2 \cdot 0.1 + 1^2 \cdot 0.3 + 2^2 \cdot 0.4 + 3^2 \cdot 0.2 = 4.1$l. Hence, Var$[Y] = 4.10 - (1.7)^2 = 4.10 - 2.89 = 1.21$.

With respect to independence, note that there are no zeros in the table, so $R_{X,Y} = R_X \times R_Y$. We now have to check that $P_{X,Y}(x,y) = P_X(x)P_Y(y)$ for all $(x,y) \in R_{X,Y}$. We quickly verify that this is indeed the case, so X, Y are independent. Therefore, Cov$[X,Y] = 0$.

Example 5.6

Consider a pair of continuous random variables X, Y, uniformly distributed on the unit disk with radius 1, centered at (0,0). Thus, the joint PDF of X, Y is given by

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{\pi} & 0 \le x^2 + y^2 \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

The joint PDF of X, Y is illustrated in Figure 5.1. We saw this example in the previous chapter, as Example 4.11. Are X, Y independent? Are X, Y uncorrelated? What are the means, variances and covariances of X, Y?

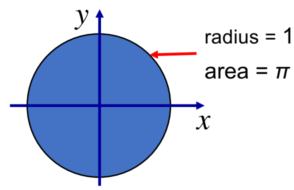

Figure 5.1: Example 5.6.

With respect to independence, consider the point $(x,y) = (0.9, 0.9)$. This point is outside the unit circle, so $f_{X,Y}(0.9, 0.9) = 0$. However, it is clear that a vertical line through that point intersects the unit circle, and so does a horizontal line. This means that $f_X(0.9) > 0, f_Y(0.9) > 0$, and therefore, $f_{X,Y}(0.9, 0.9) = 0 \neq f_X(0.9)f_Y(0.9)$. Hence, X, Y are not independent.

By symmetry, we note that $\mathbb{E}[X] = \mathbb{E}[Y] = 0$. We can also verify these using the results of Example 4.11, where we showed that $f_X(x) = \frac{2\sqrt{1-x^2}}{\pi}$, $f_Y(y) = \frac{2\sqrt{1-y^2}}{\pi}$. Both of these functions are even functions, so $\mathbb{E}[X] = \mathbb{E}[Y] = 0$. By symmetry, we can also show that $E[XY] = 0$. We will show that directly by computation:

$$\mathbb{E}[XY] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f_{X,Y}(x,y)\, dx\, dy = \int_{-1}^{1} \Big(\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} y\, dy \Big) \frac{x}{\pi}\, dx$$

The inner integral evaluates as

$$\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} y\, dy = \frac{y^2}{2} \Big|_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} = 0.$$

Thus, $\mathbb{E}[XY] = 0$, and Cov$[X,Y] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = 0$, so X and Y are uncorrelated. In this case, X, Y are also orthogonal.

It is also clear that Var$[X] =$ Var$[Y]$ by symmetry. To compute Var$[X]$, since X has zero mean, we get

$$\text{Var}[X] = \mathbb{E}[X^2] = \int_{-1}^{1} x^2 \frac{2\sqrt{1-x^2}}{\pi}\, dx = \frac{1}{4} = \text{Var}[Y],$$

where the integral can be evaluated using a trigonometric substitution $x = \sin(\theta)$.

5.3 Algebra of Covariances

Assume we have two random variables X, Y, for which we know their means $\mathbb{E}[X], \mathbb{E}[Y]$, their variances $\mathsf{Var}[X], \mathsf{Var}[Y]$, and their covariance $\mathsf{Cov}[X, Y]$. Define new random variables, linearly related to these, as

$$U = aX + bY + e; \qquad V = cX + dY + f$$

We want to compute the means and variances of U, V and their covariance. To answer this, we exploit the properties of the linearity of the expectation operation, as

$$\mathbb{E}[U] = \mathbb{E}[aX + bY + e] = a\mathbb{E}[X] + b\mathbb{E}[Y] + e\mathbb{E}[1] = a\mathbb{E}[X] + b\mathbb{E}[Y] + e.$$

$$\mathbb{E}[V] = \mathbb{E}[cX + dY + f] = c\mathbb{E}[X] + d\mathbb{E}[Y] + f.$$

What about the variance of U? Since variance is a quadratic statistic, we have to expand a quadratic to compute this. Suppose we compute this as $\mathsf{Var}[U] = \mathbb{E}[U^2] - (\mathbb{E}[U])^2$. Then,

$$\begin{aligned}
\mathbb{E}[U^2] &= \mathbb{E}[(aX + bY + e)^2] = \mathbb{E}[a^2X^2 + 2abXY + b^2Y^2 + 2aeX + 2beY + e^2] \\
&= a^2\ \mathsf{Var}[X] + a^2(\mathbb{E}[X])^2 + 2ab\ \mathsf{Cov}[X, Y] + 2ab\ \mathbb{E}[X]\mathbb{E}[Y] + b^2\ \mathsf{Var}[Y] + b^2(\mathbb{E}[Y])^2 \\
&\quad + 2ae\ \mathbb{E}[X] + 2be\ \mathbb{E}[Y] + e^2 \\
&= \Big(a^2\ \mathsf{Var}[X] + 2ab\ \mathsf{Cov}[X, Y] + b^2\ \mathsf{Var}[Y]\Big) + \Big(a^2(\mathbb{E}[X])^2 + 2ab\ \mathbb{E}[X]\mathbb{E}[Y] + b^2(\mathbb{E}[Y])^2 \\
&\quad + 2ae\ \mathbb{E}[X] + 2be\ \mathbb{E}[Y] + e^2\Big) \\
&= \Big(a^2\ \mathsf{Var}[X] + 2ab\ \mathsf{Cov}[X, Y] + b^2\ \mathsf{Var}[Y]\Big) + (\mathbb{E}[U])^2 \\
\mathsf{Var}[U] &= E[U^2) - (\mathbb{E}[U])^2 = a^2\ \mathsf{Var}[X] + 2ab\ \mathsf{Cov}[X, Y] + b^2\ \mathsf{Var}[Y]
\end{aligned}$$

However, we know that variances do not depend on the mean of the variables. That is, $\mathsf{Var}[U] = \mathsf{Var}\Big[U - \mathbb{E}[U]\Big]$. Indeed, we should have been able to compute the variance of U by assuming all the variables had zero mean. This leads to a much simple computation, as

$$\begin{aligned}
\mathsf{Var}[U] &= \mathsf{Var}[U - a\mathbb{E}[X] - b\mathbb{E}[Y] - e] = \mathbb{E}\Big[(a\tilde{X} + b\tilde{Y})^2\Big] \\
&= \mathbb{E}[a^2(\tilde{X})^2] + 2\mathbb{E}[ab\tilde{X}\tilde{Y}] + \mathbb{E}[b^2(\tilde{Y})^2] \\
&= a^2\ \mathsf{Var}[X] + 2ab\ \mathsf{Cov}[X, Y] + b^2\ \mathsf{Var}[Y].
\end{aligned}$$

where $\tilde{X} = X - \mathbb{E}[X], \tilde{Y} = Y - \mathbb{E}[Y]$, and thus $\mathsf{Var}[X] = \mathbb{E}[(\tilde{X})^2], \mathsf{Var}[Y] = \mathbb{E}[(\tilde{Y})^2]$, and $\mathsf{Cov}[X, Y] = \mathbb{E}[\tilde{X}\tilde{Y}]$. By considering only the zero-mean random variables, we are able to get to a simpler formula for variances without having to consider the extra terms associated with the means. This avoids unnecessary algebraic errors that arise when including all the terms involving the means of the random variables.

Similarly, we compute the variance of V as

$$\mathsf{Var}[V] = \mathsf{Var}\Big[V - \mathbb{E}[V]\Big] = \mathbb{E}\Big[(c\tilde{X} + d\tilde{Y})^2\Big] = c^2\ \mathsf{Var}[X] + 2cd\ \mathsf{Cov}[X, Y] + d^2\ \mathsf{Var}[Y].$$

Furthermore, the covariance of U, V is given by

$$\begin{aligned}
\mathsf{Cov}[U, V] &= \mathsf{Cov}\Big[U - \mathbb{E}[U], V - \mathbb{E}[V]\Big] = \mathbb{E}\Big[(a\tilde{X} + b\tilde{Y})(c\tilde{X} + d\tilde{Y})\Big] \\
&= ac\ \mathsf{Var}[X] + (ad + bc)\ \mathsf{Cov}[X, Y] + bd\ \mathsf{Var}[Y].
\end{aligned}$$

Example 5.7
Consider X, Y as defined in Example 5.6. We know that $\mathbb{E}[X] = \mathbb{E}[Y] = 0, \mathsf{Var}[X] = \mathsf{Var}[Y] = \frac{1}{4}, \mathsf{Cov}[X, Y] = 0$. Thus, X, Y are uncorrelated and orthogonal.

Let $U = 3X + 2Y + 1, V = 2X - 3Y - 1$. Compute the means, variances and covariance of U, V.

The means are easy: Using linearity of expectation, we get

$$\mathbb{E}[U] = 3\mathbb{E}[X] + 2\mathbb{E}[Y] + 1 = 1; \qquad \mathbb{E}[V] = 2\mathbb{E}[X] - 3\mathbb{E}[Y] - 1 = -1.$$

For variances, using the approach that we deal only with zero-mean variables, we get

$$\mathsf{Var}[U] = 9\mathsf{Var}[X] + 12\mathsf{Cov}[X,Y] + 4\mathsf{Var}[Y] = 13\mathsf{Var}[X] = \frac{13}{4}.$$

$$\mathsf{Var}[V] = 4\mathsf{Var}[X] - 12\mathsf{Cov}[X,Y] + 9\mathsf{Var}[Y] = 13\mathsf{Var}[X] = \frac{13}{4}.$$

$$\mathsf{Cov}[U,V] = 6\mathsf{Var}[X] - 9\mathsf{Cov}[X,Y] + 4\mathsf{Cov}[X,Y] - 6\mathsf{Var}[Y] = 0.$$

Our transformations resulted in U, V that are also uncorrelated, but no longer orthogonal, because neither has zero-mean. Why? If we write the transformation as a matrix:

$$\begin{bmatrix} U \\ V \end{bmatrix} = \begin{bmatrix} 3 & 2 \\ 2 & -3 \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix} + \begin{bmatrix} 1 \\ -1 \end{bmatrix},$$

you will notice that the first and second rows of the transformation matrix for X, Y are perpendicular vectors. We will explore this further when we discuss random vectors.

5.4 Jointly Gaussian Random Variables:

There is a class of jointly continuous random variables whose joint PDF is entirely specified by its second order statistics. Recall that Gaussian random variables had PDFs specified entirely in terms of their means and variances. In this section, we define the concept of pairs of jointly Gaussian random variables, where the joint PDFs are specified entirely by first- and second-order statistics, and explore their properties.

We begin by constructing a pair of independent, standard Gaussian random variables. Let U, V be standard Gaussian random variables defined on the same probability space. That is, $U \sim N(0,1), V \sim N(0,1)$ both have zero mean and unit variance. To merge them into joint random variables, we assume that U, V are independent, resulting in a pair of **independent unit Gaussian random variables**. In this case, the joint PDF is

$$f_{U,V}(u,v) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{v^2}{2}} = \frac{1}{2\pi} e^{-\frac{u^2+v^2}{2}}.$$

The joint probability density of a pair of unit Gaussian random variables is shown in Figure 5.2. The density is centered at (0,0), and has a circular symmetry, decaying to 0 as $u^2 + v^2$ approaches infinity. Consider now a pair of random variables X, Y defined in terms of U, V as

$$X = \sigma_X U + \mu_X; \qquad Y = \sigma_Y U + \mu_Y$$

where $\sigma_X, \sigma_Y > 0$, and μ_X, μ_Y are constants. Since X depends only on U and Y depends only on V, X and Y are also independent random variables.

Note that $\mathbb{E}[X] = \sigma_X \mathbb{E}[U] + \mu_X = \mu_X$, $\mathsf{Var}[X] = \sigma_X^2 \mathsf{Var}[U] = \sigma_X^2$. Similarly, $\mathbb{E}[Y] = \mu_Y, \mathsf{Var}[Y] = \sigma_Y^2$. Since X is a linear transformation of the variable U, we can obtain the density of X using the methods of Chapter 3 as

$$f_X(x) = \frac{1}{|\sigma_X|} f_U(\frac{X - \mu_X}{\sigma_X}) = \frac{1}{\sqrt{2\pi\sigma_X^2}} e^{-\frac{(X-\mu_X)^2}{2\sigma_X^2}}.$$

which also follows because a linear transformation of a Gaussian random variable results in another Gaussian random variable. Similarly,

$$f_Y(y) = \frac{1}{\sqrt{2\pi\sigma_Y^2}} e^{-\frac{(Y-\mu_Y)^2}{2\sigma_Y^2}}.$$

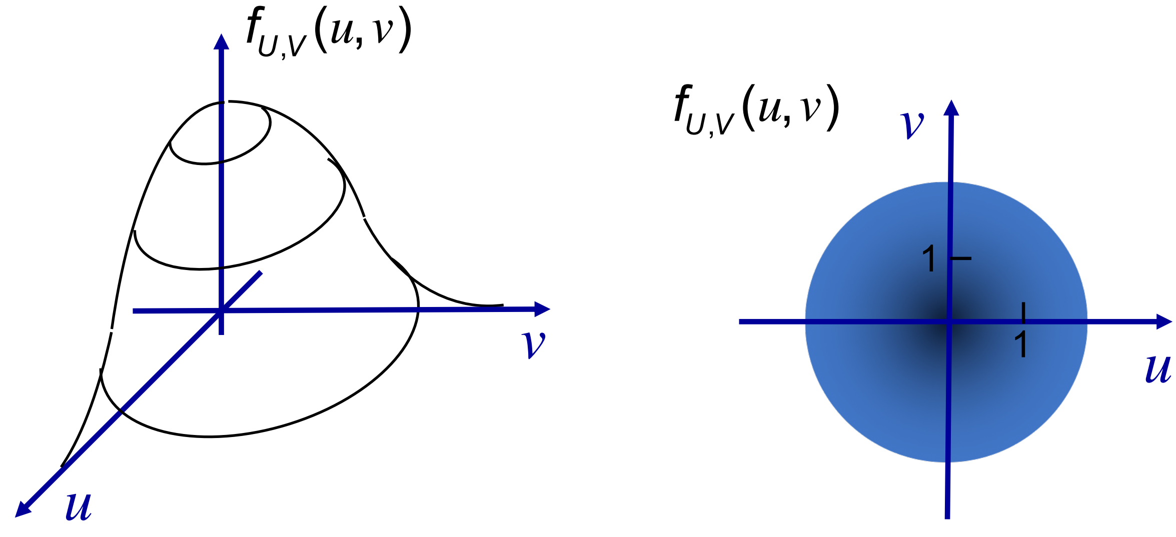

Figure 5.2: Illustration of the density of a pair of independent unit Gaussian random variables.

and, because X, Y are independent, their joint PDF is given by

$$f_{X,Y}(x,y) = \frac{1}{\sqrt{2\pi\sigma_X^2}} e^{-\frac{(X-\mu_X)^2}{2\sigma_X^2}} \frac{1}{\sqrt{2\pi\sigma_Y^2}} e^{-\frac{(Y-\mu_Y)^2}{2\sigma_Y^2}} = \frac{1}{2\pi\sigma_X\sigma_Y} e^{-\frac{(X-\mu_X)^2}{2\sigma_X^2}+\frac{(Y-\mu_Y)^2}{2\sigma_Y^2}}$$

An illustration of the joint PDF of X, Y is shown in the figure on the right. Note that the level sets of the probability density function (curves where $f_{X,Y}(x,y) = K$ for some constant K) are now ellipses, and the center of the PDF has shifted to the mean (μ_X, μ_Y). The individual standard deviations are measures of the relative elongation of the ellipses along each axis. The major axes of the ellipses are aligned with the x and y axes, because X and Y are still independent.

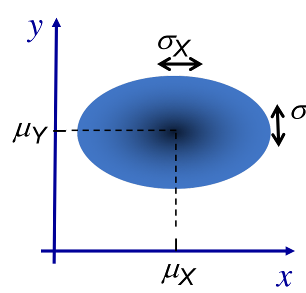

Figure 5.3: Gaussian PDF with unequal variances.

Consider now $Z = X + Y$ to be the sum of two independent jointly Gaussian random variables. We want to show that this is also a Gaussian random variable. If we know this, then the PDF of Z can be computed trivially by knowing $\mathbb{E}[Z] = \mu_X + \mu_Y$, and $\mathsf{Var}[Z] = \sigma_X^2 + \sigma_Y^2$, since the variances add when X, Y are uncorrelated and hence independent. We show this for the case where the means $\mu_X = \mu_Y = 0$, as we can always add a constant to shift the means. We refer to Section 4.7.1 for determining the density of a sum of two jointly continuous random variables, as

$$
\begin{aligned}
f_Z(z) &= \int_{-\infty}^{\infty} f_{X,Y}(x, z-x)\, dx = C \int_{-\infty}^{\infty} e^{-\frac{x^2}{2\sigma_X^2} - \frac{(z-x)^2}{2\sigma_Y^2}}\, dx \\
&= Ce^{-\frac{z^2}{2\sigma_Y^2}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}(\frac{1}{\sigma_X^2}+\frac{1}{\sigma_Y^2}) + \frac{xz}{\sigma_Y^2}}\, dx \\
&= Ce^{-\frac{z^2}{2\sigma_Y^2}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}(\frac{1}{\sigma_X^2}+\frac{1}{\sigma_Y^2}) + \frac{xz}{\sigma_Y^2} - \frac{z^2}{2}\frac{\sigma_X^2}{\sigma_Y^2(\sigma_X^2+\sigma_Y^2)} + \frac{z^2}{2}\frac{\sigma_X^2}{\sigma_Y^2(\sigma_X^2+\sigma_Y^2)}}\, dx \quad \text{(add and subtract same term)} \\
&= Ce^{-\frac{z^2}{2\sigma_Y^2} + \frac{z^2}{2}\frac{\sigma_X^2}{\sigma_Y^2(\sigma_X^2+\sigma_Y^2)}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}(\frac{1}{\sigma_X^2}+\frac{1}{\sigma_Y}^2) + \frac{xz}{\sigma_Y^2} - \frac{z^2}{2}\frac{\sigma_X^2}{\sigma_Y^2(\sigma_X^2+\sigma_Y^2)}}\, dx \\
&= Ce^{-\frac{z^2}{2(\sigma_Y^2+\sigma_X^2)}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\left(\frac{1}{\sigma_X^2}+\frac{1}{\sigma_Y^2}\right)\left(x^2 - \frac{\sigma_X^2\sigma_Y^2}{\sigma_Y^2+\sigma_X^2}\frac{2xz}{\sigma_Y^2} + z^2\frac{\sigma_X^2}{\sigma_Y^2(\sigma_X^2+\sigma_Y^2)}\frac{\sigma_X^2\sigma_Y^2}{\sigma_X^2+\sigma_Y^2}\right)}\, dx \\
&= Ce^{-\frac{z^2}{2(\sigma_Y^2+\sigma_X^2)}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\left(\frac{1}{\sigma_X^2}+\frac{1}{\sigma_Y^2}\right)\left(x - \frac{\sigma_X^2}{\sigma_X^2+\sigma_Y^2}\right)^2}\, dx \\
&= C_1 e^{-\frac{z^2}{2(\sigma_Y^2+\sigma_X^2)}}
\end{aligned}
$$

where the constant is chosen C_1 to satisfy the normalization property $\int_{-\infty}^{\infty} f_Z(z)\, dz = 1$. The result shows that Z is a Gaussian random variable with zero mean, and variance $\sigma_X^2 + \sigma_Y^2$.

Using the above argument, we can show that a random variable $X = aU + bV + \mu_X$ will be Gaussian, with mean $\mathbb{E}[X] = \mathbb{E}[aU + bV + \mu_X] = \mu_X$, and variance $\mathsf{Var}[X] = \mathsf{Var}[aU + bV] = a^2\mathsf{Var}[U] + b^2\mathsf{Var}[V] = a^2 + b^2$. Similarly, a random variable $Y = cU + dV + \mu_Y$ will be Gaussian, with mean $\mathbb{E}[Y] = \mu_Y$, and variance $\mathsf{Var}[Y] = c^2 + d^2$.

We formally define jointly Gaussian random variables as follows: A pair of random variables X and Y are **jointly Gaussian random variables** if they are linear functions of independent unit Gaussian random variables U and V:

$$X = aU + bV + \mu_X \qquad Y = cU + dV + \mu_Y \ .$$

We now compute the covariance of X, Y as

$$\mathsf{Cov}[X, Y] = \mathbb{E}[(aU + bV)(cU + dV)]) = ac\mathbb{E}[U^2] + (ad + bc)\mathbb{E}[UV] + bd\mathbb{E}[V^2] = ac + bd,$$

since U, V are zero-mean, independent, unit variance random variables. The resulting correlation coefficient is

$$\rho_{X,Y} = \frac{ac + bd}{\sqrt{(a^2 + b^2)(c^2 + d^2)}}.$$

When the correlation coefficient of X, Y has magnitude less than 1, we can write the joint PDF of X, Y as

$$f_{X,Y}(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1 - \rho_{X,Y}^2}} e^{-\frac{1}{2(1-\rho_{X,Y}^2)}\left(\frac{(x-\mu_X)^2}{\sigma_X^2} - 2\rho_{X,Y}\frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} + \frac{(y-\mu_Y)^2}{\sigma_Y^2}\right)}$$

Thus, the joint PDF is fully specified by the first- and second-order statistics: the means μ_X, μ_Y, the variances σ_X^2, σ_Y^2, and the correlation coefficient $\rho_{X,Y}$.

This is a difficult formula to remember, and it does not generalize to more than two Gaussian random variables. However, we can write this in terms of vectors and matrices as follows:

$$f_{X,Y}(x, y) = \frac{1}{\sqrt{\det\left(2\pi \begin{bmatrix} \sigma_X^2 & \mathsf{Cov}[X,Y] \\ \mathsf{Cov}[X,Y] & \sigma_Y^2 \end{bmatrix}\right)}} e^{-\frac{1}{2}\begin{bmatrix} x - \mu_x & y - \mu_y \end{bmatrix} \begin{bmatrix} \sigma_X^2 & \mathsf{Cov}[X,Y] \\ \mathsf{Cov}[X,Y] & \sigma_Y^2 \end{bmatrix}^{-1} \begin{bmatrix} x - \mu_x \\ y - \mu_y \end{bmatrix}}.$$

This form that uses the inverse of a matrix formed from the individual covariances generalizes well to three or more Gaussian random variables.

An illustration of the joint PDF of X, Y in this general case is shown in the figure on the right. Note that the level sets of the probability density function (curves where $f_{X,Y}(x, y) = K$ for some constant K) are still ellipses, and the center of the PDF has shifted to the mean (μ_X, μ_Y). The individual standard deviations are measures of the relative elongation of the ellipses along each axis. However, note that the major axes of the ellipses are no longer aligned with the x and y axes, because X and Y are now correlated and not independent. That is seen in the joint PDF by the presence of xy terms in the exponent of the density. Note that, if $\rho_{X,Y} = 0$, these terms vanish.

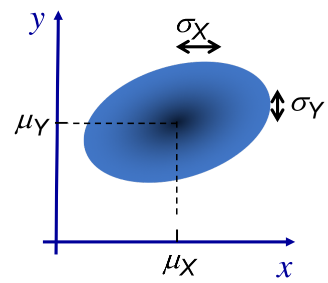

Figure 5.4: Correlated Gaussian PDF.

Jointly Gaussian random variables satisfy the following properties:

- **Any linear function of X and Y plus a constant is Gaussian:** If $Z = \alpha X + \beta Y + \gamma$, then Z is Gaussian with $\mathbb{E}[Z] = \mu_Z, \mathsf{Var}[Z] = \sigma_Z^2$ where

$$\mu_Z = \alpha\mu_X + \beta\mu_Y + \gamma, \qquad \sigma_Z^2 = \alpha^2\sigma_X^2 + \beta^2\sigma_Y^2 + 2\alpha\beta\mathsf{Cov}[X, Y]$$

  The reason for this is that, since X, Y are linear combinations of independent, unit Gaussian random variables U, V plus a constant, we can substitute for X, Y and write Z as a linear combination of U, V plus a constant. We have already shown this is a Gaussian random variable.

- **Marginal PDFs are Gaussian:** X is Gaussian with $\mathbb{E}[X] = \mu_X, \mathsf{Var}[X] = \sigma_X^2$ and Y is Gaussian with $\mathbb{E}[Y] = \mu_Y, \mathsf{Var}[Y] = \sigma_Y^2$.

  The function $Z = 1 \cdot X + 0 \cdot Y$ is a linear combination, and hence it is Gaussian. We know its mean and variance by computation as above.

- **Uncorrelated $\implies$ Independence:** X and Y are uncorrelated ($\mathsf{Cov}[X, Y] = 0$ or $\rho_{X,Y} = 0$) if and only if X and Y are independent.

  This follows by examining the form of the joint density function described above. If $\rho_{X,Y} = 0$, then we can separate $F_{X,Y} = f_X(x)f_Y(y)$. In general, uncorrelated random variables are not independent. However, for jointly Gaussian random variables, uncorrelated Gaussian random variables are independent. This means we can verify independence strictly using second-order statistics.

- $|\rho_{X,Y}| = 1$ if and only if Y is a deterministic linear function of X (and vice versa). In this case, we can write Y as $Y = \rho_{X,Y}\dfrac{\sigma_Y}{\sigma_X}(X - \mu_X) + \mu_Y$.

- **Conditional PDF of X given $Y = y$ is Gaussian:** The conditional PDF $f_{X|Y}(x|y)$ of X given $Y = y$ is Gaussian with mean $\mathbb{E}[X|Y = y]$ and variance $\mathsf{Var}[X|Y = y]$ to be computed as:

$$\mathbb{E}[X|Y = y] = \mu_X + \rho_{X,Y}\frac{\sigma_X}{\sigma_Y}(y - \mu_Y) = \mu_X + \frac{\mathsf{Cov}[X, Y]}{\mathsf{Var}[Y]}(y - \mu_Y)$$

$$\mathsf{Var}[X|Y = y] = (1 - \rho_{X,Y}^2)\sigma_X^2 = \mathsf{Var}[X] - \frac{\mathsf{Cov}[X, Y]^2}{\mathsf{Var}[Y]}$$

Let's derive this result. We know the following:

$$
\begin{aligned}
f_{X,Y}(x,y) &= \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho_{X,Y}^2}} e^{-\frac{1}{2(1-\rho_{X,Y}^2)}\left(\frac{(x-\mu_X)x^2}{\sigma_X^2} - 2\rho_{X,Y}\frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} + \frac{(y-\mu_Y)^2}{\sigma_Y^2}\right)} \\
&= \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho_{X,Y}^2}} e^{Q(x-\mu_X,y-\mu_Y)}, \quad \text{where} \\
Q(x,y) &= -\frac{1}{2(1-\rho_{X,Y}^2)}\left((\frac{x}{\sigma_X})^2 \,-\, 2\rho_{X,Y}(\frac{x}{\sigma_X})(\frac{y}{\sigma_X\sigma_Y}) \,+\, (\frac{y}{\sigma_Y})^2\right) \\
&= -\frac{1}{2(1-\rho_{X,Y}^2)}\left((\frac{x}{\sigma_X})^2 \,-\, 2\rho_{X,Y}(\frac{x}{\sigma_X})(\frac{y}{\sigma_X\sigma_Y}) \,+\, (\rho_{X,Y}\frac{y}{\sigma_Y})^2 + (1-\rho_{X,Y}^2)(\frac{y}{\sigma_Y})^2\right) \\
&= -\frac{1}{2(1-\rho_{X,Y}^2)\sigma_X^2}\left(x \,-\, \rho_{X,Y}\frac{\sigma_X}{\sigma_Y}y\right)^2 - \frac{y^2}{2\sigma_Y^2}. \quad \text{Hence,} \\
f_{X,Y}(x,y) &= \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho_{X,Y}^2}} e^{-\frac{1}{2(1-\rho_{X,Y}^2)\sigma_X^2}\left((x-\mu_X)-\rho_{X,Y}\frac{\sigma_X}{\sigma_Y}(y-\mu_Y)\right)^2 - \frac{(y-\mu_Y)^2}{2\sigma_Y^2}} \\
f_Y(y) &= \frac{1}{\sqrt{2\pi}\sigma_Y} e^{-\frac{(y-\mu_Y)^2}{2\sigma_Y^2}} \\
f_{X|Y}(x|y) &= \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho_{X,Y}^2}}(\sqrt{2\pi}\sigma_Y) e^{-\frac{1}{2(1-\rho_{X,Y}^2)\sigma_X^2}\left((x-\mu_X)-\rho_{X,Y}\frac{\sigma_X}{\sigma_Y}(y-\mu_Y)\right)^2} \\
&= \frac{1}{\sigma_X\sqrt{2\pi(1-\rho_{X,Y}^2)}} e^{-\frac{1}{2(1-\rho_{X,Y}^2)\sigma_X^2}\left((x-\mu_X)-\rho_{X,Y}\frac{\sigma_X}{\sigma_Y}(y-\mu_Y)\right)^2}
\end{aligned}
$$

We recognize the above expression as a Gaussian density, with statistics

$$
\begin{aligned}
\mathbb{E}[X|Y=y] &= \mu_X + \rho_{X,Y}\frac{\sigma_X}{\sigma_Y}(y-\mu_Y) = \mu_X + \frac{\mathsf{Cov}[X,Y]}{\mathsf{Var}[Y]}(y-\mu_Y) \\
\mathsf{Var}[X|Y=y] &= (1-\rho_{X,Y}^2)\sigma_X^2 = \mathsf{Var}[X] - \frac{\mathsf{Cov}[X,Y]^2}{\mathsf{Var}[Y]}
\end{aligned}
$$

Notice that the conditional covariance does not depend on the actual observed value $Y=y$; it only depends on the second order statistics of X, Y. Notice also that the conditional covariance $\mathsf{Var}[X|Y=y]$ is no larger than the unconditional covariance $\mathsf{Var}[X]$, as we are subtracting a nonnegative term.

The above formulas for the conditional mean and variance are very important in estimation, as we will illustrate in a subsequent chapter. Specifically, $\mathbb{E}[X|Y=y]$ is an estimate of the random variable X based on measuring that the random variable Y has value y. Define

$$
e = X - \mathbb{E}[X|Y] = X - \mu_X - \frac{\mathsf{Cov}[X,Y]}{\mathsf{Var}[Y]}(Y-\mu_Y).
$$

Then, this is the error in the estimate of X given observation Y. In this case, e is a linear function of X and Y plus a constant.

Note some important properties of the estimation error:

- $\mathbb{E}[e(y)] = 0$. This follows directly by noting that $e = \tilde{X} - \frac{\mathsf{Cov}[X,Y]}{\mathsf{Var}[Y]}\tilde{Y}$, and thus it is a linear combination of zero-mean random variables.

- $\mathbb{E}[e^2] = \mathsf{Var}[X|Y=y]$. Note that $\mathsf{Var}[X|Y=y]$ is a constant that does not depend on y. This follows because

$$\begin{aligned}\mathbb{E}[e^2] &= \mathbb{E}[\tilde{X}^2] - 2\frac{\mathsf{Cov}[X,Y]}{\mathsf{Var}[Y]}\mathbb{E}[\tilde{X}\tilde{Y}] + (\frac{\mathsf{Cov}[X,Y]}{\mathsf{Var}[Y]})^2\mathbb{E}[\tilde{Y}^2] \\ &= \mathsf{Var}[X] - 2\frac{\mathsf{Cov}[X,Y]}{\mathsf{Var}[Y]}\mathsf{Cov}[X,Y] + (\frac{\mathsf{Cov}[X,Y]}{\mathsf{Var}[Y]})^2\mathsf{Var}[Y] \\ &= \mathsf{Var}[X] - \frac{\mathsf{Cov}[X,Y]^2}{\mathsf{Var}[Y]}\end{aligned}$$

  A more subtle proof of the above uses iterated expectations, as

$$\mathbb{E}[e^2] = \mathbb{E}\Big[\mathbb{E}[e^2|Y]\Big] = \mathbb{E}\Big[\mathsf{Var}[X|Y]\Big] = \mathsf{Var}[X|Y],$$

  which follows because

$$\mathsf{Var}[X|Y]$$

  is a constant that does not depend on

$$Y$$

  .

- $\mathsf{Cov}[e,Y] = 0$. This states that the estimation error is uncorrelated with the measurement Y. We compute this directly as

$$\mathsf{Cov}[e,Y] = \mathbb{E}[(\tilde{X} - \frac{\mathsf{Cov}[X,Y]}{\mathsf{Var}[Y]}\tilde{Y})\tilde{Y}] = \mathsf{Cov}[X,Y] - \frac{\mathsf{Cov}[X,Y]}{\mathsf{Var}[Y]}\mathsf{Var}[Y] = 0.$$

- $\mathbb{E}[eY] = 0$, so that the estimation error is orthogonal to the measurement Y. This is because $\mathbb{E}[eY] = \mathsf{Cov}[e,Y] + \mathbb{E}[e]\mathbb{E}[Y] = 0$ because $\mathbb{E}[e] = 0$ and $\mathsf{Cov}[e,Y] = 0$.

- e, Y are jointly Gaussian, since e is a linear transformation of X, Y, and $\mathsf{Cov}[e,Y] = 0$, then e, Y are independent!

Example 5.8
Let X, Y be zero-mean, unit variance Gaussian random variables with correlation coefficient $\rho_{X,Y} = 0.5$. Compute the covariance of X and Y. Compute the conditional probability density of X given $Y = 2$.

From the correlation coefficient definition,

$$\rho_{X,Y} = \frac{\mathsf{Cov}[X,Y]}{\sqrt{\mathsf{Var}[X]\mathsf{Var}[Y]}} = \mathsf{Cov}[X,Y] = 0.5.$$

For the conditional density, we know it is Gaussian, so we compute the conditional mean and the conditional covariance.

$$\begin{aligned}\mathbb{E}[X|Y=2] &= \mathbb{E}[X] + \frac{\mathsf{Cov}[X,Y]}{\mathsf{Var}[Y]}(2 - \mathbb{E}[Y]) = 0.5 \cdot 2 = 1. \\ \mathsf{Var}[X|Y=2] &= \mathsf{Var}[X] - \frac{cov[X,Y]^2}{\mathsf{Var}[Y]} = 1 - 0.25 = 0.75.\end{aligned}$$

The conditional density is Gaussian with mean 1, variance 0.75.

Example 5.9
Assume that X, Y are *correlated,* jointly Gaussian random variables, such that $\mathbb{E}[X] = \mathbb{E}[Y] = 1, \mathsf{Var}[X] = 1, \mathsf{Var}[Y] = 1$ and $\mathsf{Cov}[X,Y] = 0.5$. Define derived random variables $A = 2X - 3, B = X - 2Y$.

1. Are A, B Gaussian?

   Yes. Linear combinations of joint Gaussians are Gaussian.

2. What are $\mathbb{E}[A], \mathbb{E}[B]$?
   Using the linearity of expectations, $\mathbb{E}[A] = 2\mathbb{E}[X] - 3 = -1$. $\mathbb{E}[B] = \mathbb{E}[X] - 2\mathbb{E}[Y] = -1$.
3. Compute $\mathsf{Var}[A], \mathsf{Var}[B]$.
   Since A is a scaled version of X, translated, we have $\mathsf{Var}[A] = (2)^2\mathsf{Var}[X] = 4$. For B, we use the method for representing the zero-mean random variables $\tilde{B}, \tilde{X}, \tilde{Y}$, so that
   $$\mathsf{Var}[B] = \mathbb{E}[\tilde{B}^2] = \mathbb{E}[(\tilde{X} - 2\tilde{Y})^2] = \mathbb{E}[\tilde{X}^2] - 4\mathbb{E}[\tilde{X}\tilde{Y}] + 4\mathbb{E}[\tilde{Y}^2] = \mathsf{Var}[X] - 4\mathsf{Cov}[X, Y] + 4\mathsf{Var}[Y] = 3.$$
4. Compute $\mathsf{Cov}(A, B)$.
   Proceeding as before with the zero-mean representations,
   $$\mathsf{Cov}[A, B] = 2\mathbb{E}[\tilde{X}^2] - 4\mathbb{E}[\tilde{X}\tilde{Y}] = 2\mathsf{Var}[X] - 4\mathsf{Cov}[X, Y] = 2 - 2 = 0.$$
5. Are X, Y independent? Explain.
   They are clearly not independent, since the covariance is non-zero.
6. Are A, B independent? Explain.
   Yes, they are independent, because they are uncorrelated and Gaussian.
7. Compute $\mathbb{E}[Y|A = a]$.
   We know $\mathbb{E}[Y|A = a] = \mathbb{E}[Y] + \frac{\mathsf{Cov}[A,Y]}{\mathsf{Var}[A]}(a - \mathbb{E}[A])$. We have most of those terms computed, except for $\mathsf{Cov}[A, Y]$, which is $\mathsf{Cov}[A, Y] = \mathbb{E}[\tilde{A}\tilde{Y}] = \mathbb{E}[2\tilde{X}\tilde{Y}] = 1$. Hence, $\mathbb{E}[Y|A = a] = 1 + \frac{1}{4}(a + 1)$.
8. Let $e = Y - \mathbb{E}[Y|A = a]$. Compute $\mathbb{E}[e^2]$.
   Since e is the conditional estimation error, this is asking for the conditional variance $\mathsf{Var}[Y|A = a] = \mathsf{Var}[Y] - \frac{\mathsf{Cov}[Y,A]^2}{\mathsf{Var}[A]} = 1 - \frac{1}{4} = \frac{3}{4}$.
9. Compute the covariance between B and Y.
   By now, we know how to do this with the zero-mean versions:
   $$\mathsf{Cov}[B, Y] = \mathbb{E}[\tilde{B}\tilde{Y}] = \mathbb{E}[(\tilde{X} - 2\tilde{Y})\tilde{Y}] = \mathsf{Cov}[X, Y] - 2\mathsf{Var}[Y] = -\frac{3}{2}.$$

Example 5.10
Suppose we have two jointly continuous random variables X, Y, with marginal probability densities $f_X(x), f_Y(y)$ that are Gaussian. Must the pair X, Y be jointly Gaussian random variables?
Surprisingly, the answer to this is no. Consider the following jointly continuous random variables X, Y with joint PDF given by

$$f_{X,Y}(x, y) = \begin{cases} \frac{1}{\pi}e^{-\frac{x^2+y^2}{2}} & 0 \le xy, \\ 0 & \text{otherwise.} \end{cases}$$

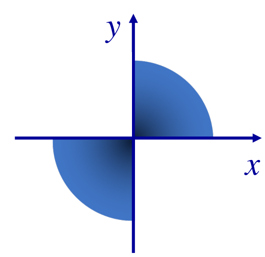

Figure 5.5: Non Gaussian PDF with Gaussian marginals.

This density is illustrated in the figure on the right. As you can see, it is definitely not a Gaussian, since the range of (X, Y) is not all of $\Re^2$. The marginal density of X is:

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dy = \begin{cases} \frac{1}{\pi}e^{-\frac{x^2}{2}} \int_0^{\infty} e^{-\frac{y^2}{2}}\, dy = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}} & x \ge 0, \\ \frac{1}{\pi}e^{-\frac{x^2}{2}} \int_{-\infty}^{0} e^{-\frac{y^2}{2}}\, dy = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}} & x < 0. \end{cases}$$

which is Gaussian. Similarly, the marginal density of Y is Gaussian. This shows that having Gaussian marginal densities does not guarantee that the joint density is Gaussian.

5.5 Random Vectors

So far, we have focused our analysis on pairs of random variables X, Y. Nevertheless, the theory that we introduced for pairs of random variables extends easily to higher dimensional vectors. Given a probability space $(\Omega, \mathcal{E}, \mathbb{P})$, we can define a **random vector** as a function that maps outcomes $\omega \in \Omega$ to vectors $\underline{x}(\omega)$ that take values in $\Re^n$, an n-dimensional Euclidean space. The theory of random vectors parallels the development we have presented for pairs of random variables. We can define the cumulative distribution function $F_{\underline{X}}(\underline{x})$ for general random vectors. If the random vectors are discrete, one defines the joint Probability Mass Function $P_{\underline{X}}(\underline{x})$ in a similar manner as we did for pairs of random variables. Random vectors are jointly continuous if there is a density $f_{\underline{X}}(\underline{x})$ such that the joint CDF can be written as

$$F_{\underline{X}}(\underline{x}) = \underset{\underline{a} \le \underline{x}}{\int \cdots \int} f_{\underline{X}}(a_1, \cdots, a_n)\, da_1 \cdots da_n.$$

While all of this is formally interesting, one seldom has enough information to compute the full multidimensional joint probability density of random vectors, unless one has extra structure. For instance, if the components of the random vector $\underline{X} = \begin{bmatrix} X_1 & X_2 & \cdots & X_n \end{bmatrix}^T$ are independent, then $f_{\underline{X}}(\underline{x}) = \prod_{k=1}^{n} f_{X_k}(x_k)$. However, it is much easier to compute statistics such as means, variances and covariances.

In this section, we focus on defining first- and second-order statistics for random vectors, and describe how they change as the random vectors undergo linear transformations. We show subsequently how to extend our analysis of pairs of jointly Gaussian random variables to Gaussian random vectors, where the full joint PDF can be defined in terms of first- and second order statistics.

Let $\underline{X}$ be a random vector with values in $\Re^n$. We assume random vectors are column vectors, so $\underline{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix}$. We define the mean of $\underline{X}$, or its expected value, as $\mathbb{E}[\underline{X}] = \begin{bmatrix} \mathbb{E}[X_1] \\ \mathbb{E}[X_2] \\ \vdots \\ \mathbb{E}[X_n] \end{bmatrix}$.

Since expectation is a linear operation, this is simply the vector of expected values, one for each random variable in the random vector $\underline{X}$. For pairs of random variables X, Y, this corresponds to stacking the individual expected values into a vector, as

$$\underline{X} = \begin{bmatrix} X \\ Y \end{bmatrix}; \qquad \mathbb{E}[\underline{X}] = \begin{bmatrix} \mathbb{E}[X] \\ \mathbb{E}[Y] \end{bmatrix}.$$

For pairs of random variables $\underline{X}$, we define the covariance matrix $\Sigma_{\underline{X}}$ as

$$\Sigma_{\underline{X}} = \begin{bmatrix} \mathsf{Var}[X] & \mathsf{Cov}[X,Y] \\ \mathsf{Cov}[X,Y] & \mathsf{Var}[Y] \end{bmatrix}.$$

Note that this is a symmetric matrix. We can write this covariance matrix as:

$$\begin{aligned} \Sigma_{\underline{X}} &= \begin{bmatrix} \mathbb{E}[(X-\mathbb{E}[X])^2] & \mathbb{E}[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])] \\ \mathbb{E}[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])] & \mathbb{E}[(Y-\mathbb{E}[Y])^2] \end{bmatrix} \\ &= \mathbb{E}\Big[\begin{bmatrix} (X-\mathbb{E}[X])^2 & (X-\mathbb{E}[X])(Y-\mathbb{E}[Y])] \\ (X-\mathbb{E}[X])(Y-\mathbb{E}[Y]) & (Y-\mathbb{E}[Y])^2 \end{bmatrix} \Big], \\ &= \mathbb{E}\Big[(\underline{X}-\mathbb{E}[\underline{X}])(\underline{X}-\mathbb{E}[\underline{X}])^T\Big] \end{aligned}$$

where $\underline{X}^T$ is the transpose of the column vector, resulting in a row vector. Hence, the covariance matrix is the expected value of the outer product between a column vector of dimension 2, and a row vector of dimension 2, resulting in a 2×2 matrix. Note that this is simply arranging the scalar statistics $\mathsf{Var}[X], \mathsf{Var}[Y], \mathsf{Cov}[X,Y]$ in a matrix form. We can generalize this to n-dimensional random vectors.

For an n-dimensional random vector $\underline{X}$, the **covariance matrix** is an $n \times n$ matrix defined as

$$\Sigma_{\underline{X}} = \mathbb{E}\Big[(\underline{X}-\mathbb{E}[\underline{X}])(\underline{X}-\mathbb{E}[\underline{X}])^T\Big]$$

Using the linearity property of expectations, and multiplying the matrix, we get

$$\begin{aligned}
\Sigma_{\underline{X}} &= \mathbb{E}\Big[\underline{X}\underline{X}^T - \mathbb{E}[\underline{X}]\underline{X}^T - \underline{X}\mathbb{E}[\underline{X}]^T + \mathbb{E}[\underline{X}]\mathbb{E}[\underline{X}]^T\Big] \\
&= \mathbb{E}\Big[\underline{X}\underline{X}^T\Big] - \mathbb{E}\Big[\mathbb{E}[\underline{X}]\underline{X}^T\Big] - \mathbb{E}\Big[\underline{X}\mathbb{E}[\underline{X}]^T\Big] + \mathbb{E}\Big[\mathbb{E}[\underline{X}]\mathbb{E}[\underline{X}]^T\Big] \\
&= \mathbb{E}\Big[\underline{X}\underline{X}^T\Big] - \mathbb{E}[\underline{X}]\mathbb{E}\Big[\underline{X}^T\Big] - \mathbb{E}\Big[\underline{X}\Big]\mathbb{E}[\underline{X}]^T + \mathbb{E}[\underline{X}]\mathbb{E}[\underline{X}]^T \quad \text{(Take out constants from expectations)} \\
&= \mathbb{E}\Big[\underline{X}\underline{X}^T\Big] - \mathbb{E}[\underline{X}]\mathbb{E}[\underline{X}]^T \qquad \text{(add the last 3 terms, which are the same.)}
\end{aligned}$$

This is the generalization of the scalar identity $\mathsf{Cov}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$ to the vector case. As in the scalar case, the covariance matrix can be computed as the difference between the second moment matrix $\mathbb{E}\Big[\underline{X}\underline{X}^T\Big]$ and the outer product of the mean vectors $\mathbb{E}[\underline{X}]\mathbb{E}[\underline{X}]^T$.

Note that every element in the covariance matrix is either a variance of a random variable, or a covariance between two random variables. Specifically,

$$\Sigma_{\underline{X}} = . \begin{bmatrix} \mathsf{Var}[X_1] & \mathsf{Cov}[X_1, X_2] & \cdots & \mathsf{Cov}[X_1, X_n] \\ \mathsf{Cov}[X_1, X_2] & \mathsf{Var}[X_2] & \cdots & \mathsf{Cov}[X_2, X_n] \\ \vdots & \vdots & \ddots & \vdots \\ \mathsf{Cov}[X_1, X_n] & \mathsf{Cov}[X_2, X_n] & \cdots & \mathsf{Var}[X_n] \end{bmatrix}$$

Thus, the covariance matrix is a compilation of the second order statistics for the scalar components of the random vector $\underline{X}$.

The covariance matrix $\Sigma_{\underline{X}}$ has the following properties:

- It is a symmetric matrix.
- It is a positive semidefinite matrix: for any non-zero n-dimensional vector $\underline{a}$, the scalar defined by the matrix vector product $\underline{a}^T \Sigma_{\underline{X}} \underline{a} \geq 0$. See the appendix on linear algebra for details on what positive semi-definite means.
- The matrix $\Sigma_{\underline{X}}$ has all of its eigenvalues on the real line, and they are non-negative.
- The matrix $\Sigma_{\underline{X}}$ has n distinct eigenvectors, and each eigenvector is perpendicular to the others.

These properties will be useful in later chapters when we discuss problems of feature aggregation in data science problems. We briefly justify the most important property, that states that the covariance matrix must be positive semidefinite. Note the following: Given a random n-dimensional vector $\underline{X}$ and an n-dimensional constant vector $\underline{a}$, the random variable $Z = \underline{a}^T \underline{X}$ is a linear combination of the elements of X. If X were zero-mean, then $\mathbb{E}[Z] = \mathbb{E}[\underline{a}^T \underline{X}] = \underline{a}^T \mathbb{E}[X] = 0$. Thus, Z is also zero mean, with variance

$$\begin{aligned}
\mathsf{Var}[Z] = \mathbb{E}[Z^2] &= \mathbb{E}[\underline{a}^T \underline{X}\underline{X}^T \underline{a}] \quad \text{since } \underline{a}^T \underline{X} = \underline{X}^T \underline{a}, \\
&= \underline{a}^T \mathbb{E}[\underline{X}\underline{X}^T]\underline{a} = \underline{a}^T \Sigma_{\underline{X}} \underline{a} \geq 0
\end{aligned}$$

Thus, the positive semidefinite property follows because covariances of random variables are non-negative. Note how we carefully moved the constants $\underline{a}$ from the correct side of the expectation to keep the dimensions matching for the vector-matrix products.

Example 5.11
Suppose we have jointly continuous random variables $\underline{X} = [X_1, X_2, X_3]^T$, with joint probability density function

$$f_{\underline{X}}(\underline{x}) = \begin{cases} 6 & 0 \leq x_1 \leq x_2 \leq x_3 \leq 1, \\ 0 & \text{elsewhere.} \end{cases}$$

Compute the covariance matrix $\Sigma_{\underline{X}}$.

The range $R_{\underline{X}}$ of the density is shown on the right. We can see that it is an inverted triangular pyramid with base area 0.5 and height 1, so its volume is $\frac{1}{6}$, hence we use the constant 6 as the density in the range.

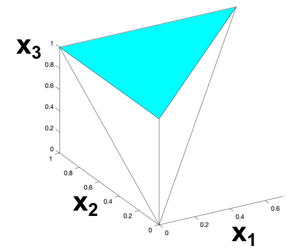

Figure 5.6: Range for Example.

We begin by computed the expected values:

$$\mathbb{E}[X_1] = \iiint_{\underline{x}\in R_{\underline{X}}} x_1 f_{\underline{X}}(\underline{x})\, d\underline{x} = \int_0^1 \Big(\int_0^{x_3} (\int_0^{x_2} 6x_1\, dx_1)\, dx_2\Big) dx_3 = \int_0^1 x_3^3\, dx_3 = \frac{1}{4}$$

$$\mathbb{E}[X_2] = \iiint_{\underline{x}\in R_{\underline{X}}} x_2 f_{\underline{X}}(\underline{x})\, d\underline{x} = \int_0^1 \Big(\int_0^{x_3} (\int_0^{x_2} 6x_2\, dx_1)\, dx_2\Big) dx_3 = \int_0^1 2x_3^3\, dx_3 = \frac{1}{2}$$

$$\mathbb{E}[X_3] = \iiint_{\underline{x}\in R_{\underline{X}}} x_3 f_{\underline{X}}(\underline{x})\, d\underline{x} = \int_0^1 \Big(\int_0^{x_3} (\int_0^{x_2} 6x_3\, dx_1)\, dx_2\Big) dx_3 = \int_0^1 3x_3^3\, dx_3 = \frac{3}{4}$$

Next, we compute the second moments:

$$\mathbb{E}[X_1^2] = \iiint_{\underline{x}\in R_{\underline{X}}} x_1^2 f_{\underline{X}}(\underline{x})\, d\underline{x} = \int_0^1 \Big(\int_0^{x_3} (\int_0^{x_2} 6x_1^2\, dx_1)\, dx_2\Big) dx_3 = \int_0^1 \frac{x_3^4}{2}\, dx_3 = \frac{1}{10}$$

$$\mathbb{E}[X_2^2] = \iiint_{\underline{x}\in R_{\underline{X}}} x_2^2 f_{\underline{X}}(\underline{x})\, d\underline{x} = \int_0^1 \Big(\int_0^{x_3} (\int_0^{x_2} 6x_2^2\, dx_1)\, dx_2\Big) dx_3 = \int_0^1 \frac{3x_3^4}{2}\, dx_3 = \frac{3}{10}$$

$$\mathbb{E}[X_3^2] = \iiint_{\underline{x}\in R_{\underline{X}}} x_3^2 f_{\underline{X}}(\underline{x})\, d\underline{x} = \int_0^1 \Big(\int_0^{x_3} (\int_0^{x_2} 6x_3^2\, dx_1)\, dx_2\Big) dx_3 = \int_0^1 3x_3^4\, dx_3 = \frac{3}{5}$$

Finally, we compute the covariances between the components of $\underline{X}$ as

$$\mathbb{E}[X_1X_2] = \iiint_{\underline{x}\in R_{\underline{X}}} x_1x_2 f_{\underline{X}}(\underline{x})\, d\underline{x} = \int_0^1 \Big(\int_0^{x_3} (\int_0^{x_2} 6x_1x_2\, dx_1)\, dx_2\Big) dx_3 = \int_0^1 \frac{3x_3^4}{4}\, dx_3 = \frac{3}{20}$$

$$\mathbb{E}[X_1X_3] = \iiint_{\underline{x}\in R_{\underline{X}}} x_1x_3 f_{\underline{X}}(\underline{x})\, d\underline{x} = \int_0^1 \Big(\int_0^{x_3} (\int_0^{x_2} 6x_1x_3\, dx_1)\, dx_2\Big) dx_3 = \int_0^1 x_3^4\, dx_3 = \frac{1}{5}$$

$$\mathbb{E}[X_2X_3] = \iiint_{\underline{x}\in R_{\underline{X}}} x_2x_3 f_{\underline{X}}(\underline{x})\, d\underline{x} = \int_0^1 \Big(\int_0^{x_3} (\int_0^{x_2} 6x_2x_3\, dx_1)\, dx_2\Big) dx_3 = \int_0^1 2x_3^4\, dx_3 = \frac{2}{5}$$

Thus, the variances and covariances are given by:

$$\begin{aligned}
\mathsf{Var}[X_1] &= \mathbb{E}[X_1^2] - (\mathbb{E}[X_1])^2 = \frac{1}{10} - \frac{1}{16} = \frac{3}{80}\\
\mathsf{Var}[X_2] &= \mathbb{E}[X_2^2] - (\mathbb{E}[X_2])^2 = \frac{3}{10} - \frac{1}{4} = \frac{1}{20}\\
\mathsf{Var}[X_3] &= \mathbb{E}[X_3^2] - (\mathbb{E}[X_3])^2 = \frac{3}{5} - \frac{9}{16} = \frac{3}{80}\\
\mathsf{Cov}[X_1, X_2] &= \mathbb{E}[X_1X_2] - \mathbb{E}[X_1]\mathbb{E}[X_2] = \frac{3}{20} - \frac{1}{8} = \frac{1}{40}\\
\mathsf{Cov}[X_1, X_3] &= \mathbb{E}[X_1X_3] - \mathbb{E}[X_1]\mathbb{E}[X_3] = \frac{1}{5} - \frac{3}{16} = \frac{1}{80}\\
\mathsf{Cov}[X_2, X_3] &= \mathbb{E}[X_2X_3] - \mathbb{E}[X_2]\mathbb{E}[X_3] = \frac{2}{5} - \frac{3}{8} = \frac{1}{40}
\end{aligned}$$

The full covariance matrix is

$$\boldsymbol{\Sigma}_{\underline{X}} = \begin{bmatrix} 0.1 & 0.15 & 0.2\\ 0.15 & 0.3 & 0.4\\ 0.2 & 0.3 & 0.6\end{bmatrix} - \begin{bmatrix}0.25\\ 0.5\\ 0.75\end{bmatrix}\begin{bmatrix}0.25 & 0.5 & 0.75\end{bmatrix} = \begin{bmatrix}0.0375 & 0.0250 & 0.0125\\ 0.0250 & 0.0500 & 0.0125\\ 0.0125 & 0.0125 & 0.0375\end{bmatrix}$$

Assume we have a random n-dimensional vector $\underline{X}$ with mean $\underline{m}_{\underline{X}}$ and covariance $\Sigma_{\underline{X}}$. Define an affine transformation of $\underline{X}$ as follows: Let $\mathbf{A}$ be an $m \times n$ matrix, and $\underline{d}$ be an m-dimensional vector. The m-dimensional random vector $\underline{Y}$ is given by:

$$\underline{Y} = \mathbf{A}\underline{X} + \underline{d}.$$

We want to compute the first- and second-order statistics of $\underline{Y}$ based on knowing the statistics of $\underline{X}$.

It is easy to compute the mean using linearity of expectation:

$$\mathbb{E}[\underline{Y}] = \mathbb{E}[\mathbf{A}\underline{X} + \underline{d}] = \mathbb{E}[\mathbf{A}\underline{X}] + \mathbb{E}[\underline{d}] = \mathbf{A}\mathbb{E}[\underline{X}] + \underline{d} = \mathbf{A}\underline{m}_{\underline{X}} + \underline{d}$$

where we have pulled out constants from the expectations. Note that, since we are dealing with matrices and vectors, we move the constant matrix $\mathbf{A}$ out on the left side of the expectation, so that the dimensions of the matrices agree when doing matrix-vector multiplication.

To compute the covariance matrix of $\underline{Y}$, we subtract the mean from both sides, to get:

$$\underline{Y} - \mathbb{E}[\underline{Y}] = \mathbf{A}\underline{X} + \underline{d} - \mathbf{A}\underline{m}_{\underline{X}} - \underline{d} = \mathbf{A}(\underline{X} - \underline{m}_{\underline{X}})$$

Using the definition of covariance, we compute it as follows:

$$\begin{aligned}
\Sigma_{\underline{Y}} &= \mathbb{E}\left[(\underline{Y} - \mathbb{E}[\underline{Y}])(\underline{Y} - \mathbb{E}[\underline{Y}])^T\right] = \mathbb{E}\left[\mathbf{A}(\underline{X} - \underline{m}_{\underline{X}})\left(\mathbf{A}(\underline{X} - \underline{m}_{\underline{X}})\right)^T\right] \\
&= \mathbb{E}\left[\mathbf{A}(\underline{X} - \underline{m}_{\underline{X}})(\underline{X} - \underline{m}_{\underline{X}})^T\mathbf{A}^T\right] \\
&= \mathbf{A}\mathbb{E}\left[(\underline{X} - \underline{m}_{\underline{X}})(\underline{X} - \underline{m}_{\underline{X}})^T\right]\mathbf{A}^T \\
&= \mathbf{A}\Sigma_{\underline{X}}\mathbf{A}^T
\end{aligned}$$

This is the generalization of the scalar scaling law for covariances, where if $Y = aX$, then $\mathsf{Var}[Y] = a^2\mathsf{Var}[X]$. The extension to vectors is careful to keep the order of the scaling by $\mathbf{A}$ and $\mathbf{A}^T$ to keep the dimensions of the resulting matrix correct.

Example 5.12

Let's revisit the example of 5.7. We have a pair of random variables X, Y with first- and second-order statistics $\mathbb{E}[X] = \mathbb{E}[Y] = 0, \mathsf{Var}[X] = \mathsf{Var}[Y] = \frac{1}{4}, \mathsf{Cov}[X, Y] = 0$.

Let's form this into a vector $\underline{X} = \begin{bmatrix} X \\ Y \end{bmatrix}$. The mean vector $\underline{m}_{\underline{X}} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, and the resulting covariance matrix is

$$\Sigma_{\underline{X}} = \begin{bmatrix} \frac{1}{4} & 0 \\ 0 & \frac{1}{4} \end{bmatrix}.$$

Define two new variables defined as $U = 3X + 2Y + 1, V = 2X - 3Y - 1$. Define the vector $\underline{W} = \begin{bmatrix} U \\ V \end{bmatrix}$. We can write the transformation from $\underline{X}$ to $\underline{W}$ as:

$$\underline{W} = \begin{bmatrix} 3 & 2 \\ 2 & -3 \end{bmatrix} \underline{X} + \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

Then, the first- and second-order statistics of $\underline{W}$ are:

$$\mathbb{E}[\underline{W}] = \begin{bmatrix} 3 & 2 \\ 2 & -3 \end{bmatrix} \mathbb{E}[\underline{X}] + \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

$$\begin{aligned}
\Sigma_{\underline{W}} &= \begin{bmatrix} 3 & 2 \\ 2 & -3 \end{bmatrix} \Sigma_{\underline{X}} \begin{bmatrix} 3 & 2 \\ 2 & -3 \end{bmatrix}^T = \begin{bmatrix} 3 & 2 \\ 2 & -3 \end{bmatrix} \begin{bmatrix} \frac{1}{4} & 0 \\ 0 & \frac{1}{4} \end{bmatrix} \begin{bmatrix} 3 & 2 \\ 2 & -3 \end{bmatrix} \\
&= \begin{bmatrix} 3 & 2 \\ 2 & -3 \end{bmatrix} \begin{bmatrix} \frac{3}{4} & \frac{1}{2} \\ \frac{1}{2} & -\frac{3}{4} \end{bmatrix} = \begin{bmatrix} \frac{13}{4} & 0 \\ 0 & \frac{13}{4} \end{bmatrix}
\end{aligned}$$

which says that $\mathsf{Var}[U] = \frac{13}{4}, \mathsf{Var}[V] = \frac{13}{4}, \mathsf{Cov}[U, V] = 0$. These are the same answers we saw in Example 5.7.

Example 5.13
Let's revisit Example 5.9. Assume that X, Y are correlated random variables, such that $\mathbb{E}[X] = \mathbb{E}[Y] = 1, \text{Var}[X] = 1, \text{Var}[Y] = 1$ and $\text{Cov}[X,Y] = 0.5$. Let $\underline{X} = \begin{bmatrix} X \\ Y \end{bmatrix}$. Then,

$$\mathbb{E}[\underline{X}] = \begin{bmatrix} 1 \\ 1 \end{bmatrix}; \qquad \Sigma_{\underline{X}} = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}.$$

Define derived random variables $A = 2X - 3, B = X - 2Y$. Let $\underline{W} = \begin{bmatrix} A \\ B \end{bmatrix}$. Then,

$$\underline{W} = \begin{bmatrix} 2 & 0 \\ 1 & -2 \end{bmatrix} \underline{X} + \begin{bmatrix} -3 & 0 \end{bmatrix}.$$

Using this equation, we obtain

$$\mathbb{E}[\underline{W}] = \begin{bmatrix} 2 & 0 \\ 1 & -2 \end{bmatrix} \mathbb{E}[\underline{X}] + \begin{bmatrix} -3 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 1 & -2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \begin{bmatrix} -3 \\ 0 \end{bmatrix} = \begin{bmatrix} -1 \\ -2 \end{bmatrix}$$
$$\Sigma_{\underline{W}} = \begin{bmatrix} 2 & 0 \\ 1 & -2 \end{bmatrix} \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 1 & -2 \end{bmatrix}^T = \begin{bmatrix} 2 & 0 \\ 1 & -2 \end{bmatrix} \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 0 & -2 \end{bmatrix}$$
$$= \begin{bmatrix} 2 & 0 \\ 1 & -2 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 1 & -1.5 \end{bmatrix} = \begin{bmatrix} 4 & 0 \\ 0 & 3 \end{bmatrix}$$

and thus $\text{Var}[A] = 4, \text{Var}[B] = 3$, and $\text{Cov}[A,B] = 0$, agreeing with the results from Example 5.9.

As the examples illustrate, the use of random vectors enables us to recover the same first- and second-order statistics for the random variables when we analyze them individually as pairs of random variables. The advantage of the vector notation is that it scales nicely to compute statistics for random vectors of dimension greater than 2, exploiting simple results from linear algebra.

5.5.1 Gaussian random vectors

A special case of random vectors is what are termed **Gaussian random vectors**. For pairs of jointly Gaussian random variables X, Y, their joint PDF is completely characterized by the first- and second-order statistics. Extending this to random vectors of dimension greater than two is straightforward, as we will show below.

We define a jointly Gaussian random vector as a generalization of what we did with pairs of random variables. First, we define n independent standard Gaussian random variables $Z_i \sim \mathcal{N}(0,1)$. We define the vector

$$\underline{Z} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix}.$$

Then, an n-dimensional random vector $\underline{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix}$ is defined to be a *Gaussian random vector* (or equivalently, $\{X_1, \ldots, X_n\}$ are defined to be a set of jointly Gaussian random variables) if

$$\underline{X} = \mathbf{A}\underline{Z} + \underline{b}$$

for some $n \times n$ matrix $\mathbf{A}$ and some n-dimensional vector $\underline{b}$.

Note that $\underline{Z}$ has mean $\underline{0}$, with covariance matrix as the $n \times n$ identity matrix $\mathbf{I}_n$. Hence, $\mathbb{E}[\underline{X}] = \mathbf{A}\underline{0} + \underline{b} = \underline{b}$. Furthermore, the covariance matrix of $\underline{X}$ is

$$\Sigma_{\underline{X}} = \mathbf{A}\Sigma_{\underline{Z}}\mathbf{A}^T = \mathbf{A}\mathbf{A}^T.$$

For a Gaussian random vector $\underline{Z}$ to be jointly continuous, the transformation $\mathbf{A}$ must be invertible. This means that the resulting covariance $\Sigma_{\underline{X}}$ is invertible. We focus only on jointly continuous random Gaussian random variables in this text.

An equivalent definition is that $\underline{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix}$ is a Gaussian random vector if, for all constant vectors $\underline{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$, the random variable $Z = \sum_{k=1}^n a_k X_k$ is a Gaussian random variable. Note that $Z = \underline{a}^T \underline{X}$ in vector notation. As noted before, it is *not enough* that each entry X_i is marginally a Gaussian random variable for the vector to be a Gaussian random vector! *All* linear combinations of the entries must also be Gaussian. The converse, however is true: the entries of a Gaussian random vector are individually Gaussian random variables.

If $\underline{X}$ had mean $\underline{m}_{\underline{X}}$ and covariance $\Sigma_{\underline{X}}$, then $Z = \underline{a}^T \underline{X}$ is a scalar Gaussian random variable with mean $\mathbb{E}[Z] = \underline{a}^T \underline{m}_{\underline{X}}$ and variance $\underline{a}^T \Sigma_{\underline{X}} \underline{a}$.

A jointly continuous Gaussian random vector $\underline{X}$ have a probability density function that is completely described by its mean $\underline{m}_{\underline{X}}$ and covariance $\Sigma_{\underline{X}}$. We use the notation $\underline{X} \sim N(\underline{m}_{\underline{X}}, \Sigma_{\underline{X}})$ to denote this density. We can write the joint PDF of $\underline{X}$ as

$$f_{\underline{X}}(\underline{x}) = \frac{1}{\sqrt{(2\pi)^n \det(\Sigma_{\underline{X}})}} e^{-\frac{1}{2}(\underline{x} - \underline{m}_{\underline{X}})^T (\Sigma_{\underline{X}})^{-1} (\underline{x} - \underline{m}_{\underline{X}})}.$$

An important property of pairs of jointly Gaussian random variables X, Y is that they are independent if and only if $\mathsf{Cov}[X, Y] = 0$. For Gaussian random vectors, the components $X_1, X_2, \ldots, X_n$ are mutually independent if and only if $\mathsf{Cov}[X_i, X_j] = 0$ for all $i, j \in 1, \ldots, n, i \neq j$. What this means is that the covariance matrix $\Sigma_{\underline{X}}$ is diagonal, with zeros in all the non-diagonal entries. For independent random vectors, the covariance matrix is

$$\Sigma_{\underline{X}} = \begin{bmatrix} \mathsf{Var}[X_1] & 0 & \cdots & 0 \\ 0 & \mathsf{Var}[X_2] & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathsf{Var}[X_n] \end{bmatrix}$$

In this special case,

$$\Sigma_{\underline{X}}^{-1} = \begin{bmatrix} \frac{1}{\mathsf{Var}[X_1]} & 0 & \cdots & 0 \\ 0 & \frac{1}{\mathsf{Var}[X_2]} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\mathsf{Var}[X_n]} \end{bmatrix}$$

and the joint probability density factors as

$$f_{\underline{X}}(\underline{x}) = \prod_{k=1}^{n} \frac{1}{\sqrt{2\pi \mathsf{Var}[X_k]}} e^{-\frac{(x_k - m_k)^2}{2\mathsf{Var}[X_k]}},$$

which shows the equivalence between independence and having a diagonal covariance matrix.

Example 5.14
Let's revisit Example 5.12 where X, Y are jointly Gaussian random variables with first- and second-order statistics $\mathbb{E}[X] = \mathbb{E}[Y] = 1, \mathsf{Var}[X] = 1, \mathsf{Var}[Y] = 1$ and $\mathsf{Cov}[X, Y] = 0.5$. Let $\underline{X} = \begin{bmatrix} X \\ Y \end{bmatrix}$. Then,

$$\mathbb{E}[\underline{X}] = \begin{bmatrix} 1 \\ 1 \end{bmatrix}; \qquad \Sigma_{\underline{X}} = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}.$$

Define $\underline{W}$ as

$$\underline{W} = \begin{bmatrix} 2 & 0 \\ 1 & -2 \end{bmatrix} \underline{X} + \begin{bmatrix} -3 & 0 \end{bmatrix}.$$

Then, from Example 5.12, we know

$$\mathbb{E}[\underline{W}] = \underline{m}_{\underline{W}} = \begin{bmatrix} -1 \\ -2 \end{bmatrix}; \qquad \Sigma_{\underline{W}} = \begin{bmatrix} 4 & 0 \\ 0 & 3 \end{bmatrix}.$$

which implies that the two components of $\underline{W}$ are uncorrelated, and hence, mutually independent. The joint density of $\underline{W}$ is Gaussian, and given by

$$f_{\underline{W}}(\underline{w}) = \left(\frac{1}{\sqrt{8\pi}} e^{-\frac{(w_1+1)^2}{8}} \right) \left(\frac{1}{\sqrt{6\pi}} e^{-\frac{(w_2+2)^2}{6}} \right),$$

which shows the factored form.