3.0 PROBABILITY, RANDOM VARIABLES AND RANDOM PROCESSES

3.1 Introduction

In this chapter we will review the concepts of probability, random variables and random processes. We begin by reviewing some of the definitions of probability. We then define random variables and density functions, and review some of the operations on random variables. We conclude by defining random processes and discussing some properties of random processes that we will need in our Kalman filter formulations.

3.2 Probability

3.2.1 Background

Suppose we roll a fair die. We say that the probability of a one appearing (the face of the die with one dot) is one-sixth. We arrived at this conclusion by considering that the die has six sides and that there is an equal chance that any one side will be facing up when the die lands. This approach to determining probability is termed the relative frequency approach and is suitable for defining probabilities associated with simple experiments. With this approach, we determine probabilities from the underlying experiment or process and deductive reasoning. For example, in the die example, we know that if we roll a die that one of six faces will show. We also assume that it is equally probable that any of the six faces will show. Given this information we thus deduce that the probability of any one face showing is one-sixth.

For more complex experiments, where the relative frequency approach becomes more difficult to apply, we can use the axiomatic approach to determine probabilities. The axiomatic approach is based on three probability axioms and uses them, and set theory, to determine probabilities associated with certain experiment outcomes.

There is a third approach to determining probabilities that is termed the experimental approach. In this approach probabilities are determined by conducting many experiment tries and analyzing the results. An example of the experimental approach is the method by which automotive fatality statistics are derived.

To proceed further we must define some terms. We will use the die example to do so. In the die example, the rolling of the die is termed an experiment and the appearance of a particular face is an outcome of the experiment. Thus, the die rolling experiment consists of six possible outcomes. If the experiment were the flipping of a coin there would be two outcomes: heads and tails. If we were to consider an experiment where we rolled two dice we could define either 36 outcomes or 11 outcomes, depending upon the how we decided to phrase the problem, and what probabilities we were trying to find. For the first case the outcomes would be “the face with one dot on die one and the face with one dot on die two”, “the face with one dot on die one and the face with two dots on die two”, and so forth. For the second case the outcomes would be “the sum of the dots on the two dice equals two”, “the sum of the dots on the two dice equals three”, and so forth.
3.2.2 Set Theory and the Axioms of Probability

When we discuss probabilities we talk about probabilities associated with events. An event is a collection of outcomes. Thus for the die problem, example events could be: “the face with one dot” or “the face with one dot or the face with two dots”. In the first case, the event consists of one outcome and in the second case the event consists of two outcomes. This idea of using events to represent collections of outcomes leads to the use of set theory since set theory gives us a means of “collecting objects”. The analogy here is that events are sets and outcomes are set elements.

There are two events that play special roles in probability theory. These are the certain event and the null event. The certain event contains all possible experiment outcomes and is analogous to the universe in set theory. It is denoted by $\mathcal{S}$. The null event contains no outcomes and is analogous to the empty set. It is denoted by $\mathcal{O}$.

In summary, we will use set theory to represent events and we associate probabilities with events. Thus, for the die experiment we can define an event $\mathcal{A}$ as

$$\mathcal{A} = \{\text{face with one dot or face with three dots or face with five dots}\} = \{f_1,f_3,f_5\}. \quad (3-1)$$

The probability of the event $\mathcal{A}$ occurring is

$$P(\mathcal{A}) = \frac{1}{2}. \quad (3-2)$$

We note that we could use set theory to write $\mathcal{A}$ as

$$\mathcal{A} = \mathcal{F}_1 \cup \mathcal{F}_3 \cup \mathcal{F}_5 = \{f_1\} \cup \{f_3\} \cup \{f_5\}. \quad (3-3)$$

which would allow us to write

$$P(\mathcal{A}) = P(\mathcal{F}_1 \cup \mathcal{F}_3 \cup \mathcal{F}_5). \quad (3-4)$$

This is where the axiomatic approach to probability becomes useful. It allows us to use set theory and derivations to compute complex probabilities from simple probabilities.

The axioms of probability are as follows

1. $P(\mathcal{S}) = 1$, 
2. $P(\mathcal{A}) \geq 0$ and 
3. If $\mathcal{A}$ and $\mathcal{B}$ are mutually exclusive then

$$P(\mathcal{A} \cup \mathcal{B}) = P(\mathcal{A}) + P(\mathcal{B}).$$

Recall that two sets (events) are mutually exclusive if $\mathcal{A} \cap \mathcal{B} = \mathcal{O}$. Stated another way, two events are mutually exclusive if the appearance of one
precludes the appearance of another. Thus if $\mathcal{A} = \{f1\}$ and $\mathcal{B} = \{f2\}$ then $\mathcal{A}$ and $\mathcal{B}$ are mutually exclusive because if the roll of the die results in the face with one dot showing, the face with two dots can’t be showing, or is precluded from occurring.

As indicated earlier, the use of the above axioms and set theory allows us to derive many other probability relations. Two of the more important ones are

$$P(\mathcal{O}) = 0 \quad (3-5)$$

and

$$P(\mathcal{A} \cup \mathcal{B}) = P(\mathcal{A}) + P(\mathcal{B}) - P(\mathcal{A} \cap \mathcal{B}). \quad (3-6)$$

Another important probability relation concerns the intersection of two events. This relation is usually stated as a definition since it cannot be derived from set theory and the axioms of probability. The probability relation states that if two events $\mathcal{A}$ and $\mathcal{B}$ are independent then

$$P(\mathcal{A} \cap \mathcal{B}) = P(\mathcal{A})P(\mathcal{B}). \quad (3-7)$$

We say that two events are independent if the occurrence of one event doesn’t give any information about, or influence the occurrence of, the other event. For example, if we consider the experiment of rolling a pair of dice, the event consisting of a one for die one is independent of the event consisting of a three for die two. This is because we assume that what we roll on the first die will not influence what we roll on the second die. It should be noted that two events can’t be mutually exclusive and independent, unless one of the events is the null event.

For those who are interested in the finer points of probability theory and independent events, it should be noted that one can’t really discuss independent events on a single experiment. They can only be defined on combined experiments. The reader is referred to Papoulis (Papoulis – 1991) or some other text on probability theory for further discussions of this area.

### 3.3 Random Variables

#### 3.3.1 Background

In the discussions of the previous section we referred to outcomes of the die experiment as “face with one dot”, etc. If we were discussing the die experiment less formally, we would refer to “the rolling of a one”. What we have done is assign numbers to the experiment outcomes. When we assign numbers to experiment outcomes we define a random variable for the experiment. Aside from notational convenience, there is another reason for defining random variables for experiments: our mathematics background. Specifically, we have developed a wealth of mathematical tools for manipulating numbers. We have also devoted a great deal of time to relating physical processes to numbers. Unfortunately, we don’t have a wealth of mathematical tools for sets and we only have limited ability to relate sets and elements to physical processes. For this reason, we use random variables to take probability theory out of the
context of set theory and put it in the context of mathematics that we know how to work with.

3.3.2 Definition of a Random Variable

We formally define a random variable as follows: For every experimental outcome, \( \xi \), we assign a number \( x(\xi) \). With this we have defined a function \( x(\xi) \), which is a mapping from a set of \( \xi \)'s to a set of numbers. This function is termed a random variable. In all cases that are of concern to us, we define \( x(\xi) \) as the numerical value of the outcome, \( \xi \), i.e., \( x(\xi) = \xi \). An example of this is the die problem we just discussed. Another example is the voltage across a resistor or the range to a target. It should be noted that we are not required to use such a convention (Papoulis – 1991). However, it is convenient.

In the future we will not explicitly indicate the dependency of the random variable on the experiment outcome. That is, we will replace \( x(\xi) \) by \( x \). We do this for two reasons. One is convenience and the other is to force us to stop explicitly thinking about sets and think in terms of numbers. This is somewhat subtle but is essential to preventing confusion when studying random variable theory.

There are two basic types of random variables: discrete and continuous. Discrete random variables can have only a finite number of values and continuous random variables can take on a continuum of values (i.e., there are an infinite number of them). An example of a discrete random variable is the random variable associated with the die experiment. An example of a continuous random variable is the voltage across a resistor.

3.3.3 Distribution and Density Functions

We define the probability

\[
P(\{x \leq x\}) = F_x(x)
\]

as the *distribution function* of the random variable \( x \). In the above \( \{x \leq x\} = \{x(\xi) \leq x\} \) is the event consisting of outcomes, \( \xi \), such that the random variable associated with those outcomes, \( x(\xi) \), is less than, or equal to, the number \( x \). The distribution function of the random variable \( x \) is the probability associated with this event. This is a long-winded word description that we don’t use. In its place we say that the distribution function of \( x \) is the probability that \( x \) is less than, or equal to, \( x \). When we make the latter statement we should realize that it is a “shorthand” notation for the longer, more formal and correct definition. In most cases we drop the subscript \( x \) and denote the distribution function as \( F(x) \).

We define the *density function* of the random variable, \( x \), as
\[ f_x(x) = f(x) = \frac{dF(x)}{dx}. \]  \hspace{1cm} (3-9) 

We can relate the density function back to the distribution function, and probability, by

\[ P\{x \leq x'\} = F(x) = \int_{-\infty}^{x} f(\alpha) d\alpha. \]  \hspace{1cm} (3-10) 

The probability \( P\{x_1 < x \leq x_2\} \) is given by

\[ P\{x_1 < x \leq x_2\} = \int_{x_1}^{x_2} f(\alpha) d\alpha. \]  \hspace{1cm} (3-11) 

For discrete random variables, \( f(x) \) consists of a series of impulse functions (Dirac delta functions) located at the values that the random variables can take on. The weight on each impulse is equal to the probability associated with the random variable. By this we mean that the weight is the probability associated with the event containing all outcomes for which \( x = x_k \). In equation form

\[ f(x) = \sum_{k} P\{x = x_k\} \delta(x - x_k). \]  \hspace{1cm} (3-12) 

As an example, the density function associated with the die experiment is

\[ f(x) = \sum_{k=1}^{6} \frac{1}{6} \delta(x - k). \]  \hspace{1cm} (3-13) 

The density function of a continuous random variable is a continuous function of \( x \).

Random variables are usually referenced by the density function associated with them. Thus a Gaussian random variable is a random variable whose density function is the Gaussian function,

\[ f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}. \]  \hspace{1cm} (3-14) 

We can derive two very important properties of the density function from Equations 3-10 and 3-11. Specifically:

- \( f(x) \geq 0 \ \forall x \) and
- \( \int_{-\infty}^{\infty} f(x) dx = 1. \)

### 3.3.4 Mean, Variance and Moments of Random Variables

Rather than work directly with random variables and their density functions we often work with functions of random variables. We do this for
much the same reason we transitioned from set theory and probabilities to random variables and density functions – our ability (or inability) to perform the mathematics. In particular, we don’t have the mathematical tools to work directly with random variables and/or density functions. Actually, we do have the tools but, as you will recall from random variable theory, they are primitive and difficult to use.

The particular functions that we are interested in are termed the various moments of the random variable, \( x \). The first of these is the first moment of \( x \), which is also called the mean. It is defined as

\[
\eta = E\{x\} = \int_{-\infty}^{\infty} xf(x) \, dx.
\]

With this, the reader can show that a Gaussian random variable with the density function in Equation 3-12 has a mean of \( \eta \).

Two important moments of \( x \) are its second moment and its second central moment. The first is termed the mean squared value of \( x \) and the second is termed the variance of \( x \). The square root of the variance is termed the standard deviation of \( x \). The mean squared value and variance are given by

\[
E\{x^2\} = \int_{-\infty}^{\infty} x^2 f(x) \, dx
\]

and

\[
\sigma^2 = E\{(x-\eta)^2\} = \int_{-\infty}^{\infty} (x-\eta)^2 f(x) \, dx,
\]

respectively.

### 3.3.5 Multiple Random Variables

For some experiments we can define several related or unrelated random variables. For example, if two dice are used in the experiment, separate random variables, \( x \) and \( y \), can be assigned to each die. For the case of two random variables we define joint distribution and joint density functions as

\[
F(x, y) = P\{\{x \leq x\} \cap \{y \leq y\}\}
\]

and

\[
f(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y},
\]

respectively.

An example of the density function for two, jointly Gaussian random variables is

\[
f(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-r^2}} \exp\left[\frac{-1}{2(1-r^2)} \left( \frac{(x-\eta_x)^2}{\sigma_x^2} - \frac{2r(x-\eta_x)(y-\eta_y)}{\sigma_x\sigma_y} + \frac{(y-\eta_y)^2}{\sigma_y^2} \right) \right],
\]

© M. Budge, Jr – Jan 2018
where the definitions of the \( \eta \)'s and \( \sigma \)'s should be obvious and \( r \) is the
\textit{correlation coefficient} defined by
\[
r = \frac{C(x, y)}{\sigma_x \sigma_y}
\]
(3-21)
\( C(x, y) \) is the \textit{covariance} between the random variables \( x \) and \( y \) and is defined
below.

If two random variables, \( x \) and \( y \), are \textit{independent} we can write their joint
density as
\[
f(x, y) = f(x) f(y).
\]
(3-22)
The reverse also holds. If \textit{Equation 3-22} is valid then \( x \) and \( y \) are independent.

We can relate the \textit{marginal} densities of \( x \) and \( y \) to their joint densities by
the equations
\[
f(x) = \int_{-\infty}^{\infty} f(x, y) dy
\]
(3-23)
and
\[
f(y) = \int_{-\infty}^{\infty} f(x, y) dx.
\]
(3-24)

As with single random variables, we use \textit{joint moments} of joint random
variables in lieu of using the actual random variables, marginal densities and
joint densities. Again, the reason for this is our depth of mathematical tools.
The two specific moments we use for two random variables are their \textit{first joint
moment} and their \textit{first, joint, central moment}. The former we term the \textit{correlation}
and the latter we term their \textit{covariance}. The correlation is given by
\[
R(x, y) = E\{xy\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xyf(x, y) dxdy
\]
(3-25)
and the covariance is given by
\[
C(x, y) = E\{(x-\eta_x)(y-\eta_y)\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x-\eta_x)(y-\eta_y)f(x, y) dxdy.
\]
(3-26)

Some properties and terms associated with correlation and covariance
are as follows.

- If \( R(x, y)=0 \) then \( x \) and \( y \) are \textit{orthogonal}.
- If \( C(x, y)=0 \) then \( x \) and \( y \) are \textit{uncorrelated}.
- If \( x \) and/or \( y \) are zero mean then \( R(x, y)=C(x, y) \).
- If \( x \) and \( y \) are independent, they are also uncorrelated. The reverse is
  not true.
• If x and y are uncorrelated and jointly Gaussian, they are also independent.

3.4 Random Processes

3.4.1 Introduction

In the previous section we discussed the concept of random variables wherein we assigned a number to experiment outcomes. In some experiments it is more appropriate to assign time functions to experiment outcomes. Examples of this might be where the experiment outcome is the selection of a generator or observation of the voltage at the output of a receiver. In the first case, we might assign sinusoidal time functions (see Figure 3-1) to each experiment outcome. The sinusoids might be of a different amplitude, frequency and relative phase for each experiment outcome (generator selection). For the second case we might assign time functions such as those of Figure 3-2.

Figure 3-1. Two samples of a random process – generator outputs
If we assign a time function, \( x(t, \xi) \), to each experiment outcome, \( \xi \), the collection of time functions is termed a random process or a stochastic process. As we did with random variables, we will not explicitly denote the dependency on \( \xi \) and write \( x(t, \xi) \) as \( x(t) \).

We note that \( x(t, \xi) \) can represent four things:

1. A family of time functions for \( t \) and \( \xi \) variable. This is a stochastic process.
2. A single time function for \( t \) variable and \( \xi \) fixed.
3. A random variable for \( t \) fixed and \( \xi \) variable.
4. A number for \( t \) and \( \xi \) fixed.

We use representations 2 and 3 to help us devise ways to deal with random processes. Specifically we will use random variable theory and time-function analysis techniques. Indeed, since \( x(t, \xi) \) is a random variable for a specific value of \( t \) we can use the density functions to characterize it. Thus, for a time \( t \), the distribution function of \( x(t, \xi) \) is

\[
F(x, t) = P\left(\{x(t, \xi) \leq x\}\right)
\]

and its density is

\[
f(x, t) = \frac{\partial F(x, t)}{\partial x}
\]

We include the variable \( t \) in the expressions for the distribution and density functions to denote the fact that the characteristics (i.e., density function) of the

Figure 3-2. Two samples of random processes – receiver outputs
random variable \( x(t, \xi) \) can change with time. \( F(x, t) \) and \( f(x, t) \) are termed the first-order distribution and density of the random process.

If we consider two instants of time, \( t_a \) and \( t_b \), we can describe the characteristics of the random variables at those times in terms of their first-order densities and the joint density of the random variables at the two times; specifically \( f(x_a, x_b, t_a, t_b) \). The joint density is termed the second-order density of the random process.

3.4.2 Autocorrelation, Autocovariance, Crosscorrelation, Crosscovariance

As with random variables, we find it easier to use the moments of the random process rather than the random process or its various densities. To this end, we define the mean of a random process as

\[
\eta(t) = E[x(t)] = \int_{-\infty}^{\infty} x f(x, t) \, dx.
\]

Note that the mean is, in general, a function of time.

We define the autocorrelation and autocovariance of a random process as

\[
R(t_a, t_b) = E\{x(t_a) x(t_b)\}
\]

and

\[
C(t_a, t_b) = E\{(x(t_a) - \eta(t_a))(x(t_b) - \eta(t_b))\}
\]

respectively. The autocorrelation of a random process is equivalent to the correlation between two random variables and the autocovariance of a random process is equivalent to the covariance between two random variables. Note that the autocorrelation and autocovariance describe the relation between two random variables that result by considering samples of a single random process at two times. The fact that we are considering two random variables leads to the use of correlation and covariance. We use the prefix “auto” to denote the fact that the random variables are derived from a single random process.

If we have two random processes, \( x(t) \) and \( y(t) \), we define their crosscorrelation and crosscovariance as

\[
R_{xy}(t_a, t_b) = E\{x(t_a) y(t_b)\}
\]

and

\[
C_{xy}(t_a, t_b) = E\{(x(t_a) - \eta_x(t_a))(y(t_b) - \eta_y(t_b))\}
\]

respectively. We use the prefix “cross” to denote that the two random variables at the two times are now derived from two, different random processes.

At this point we want to briefly delineate some properties of random processes, autocorrelations, aurocovariances, crosscorrelations and crosscovariances.
First we note that if \( x(t) \) is zero mean then, \( C(t_a, t_b) = R(t_a, t_b) \) \( \forall t_a, t_b \).

Also, if \( x(t) \) and/or \( y(t) \) are zero mean then \( C_{xy}(t_a, t_b) = R_{xy}(t_a, t_b) \) \( \forall t_a, t_b \).

If \( t_a = t_b \) then \( C(t_a, t_a) = \sigma^2(t_a) \), the variance of \( x(t) \) (evaluated at \( t = t_a \)).

If \( C_{xy}(t_a, t_b) = 0 \) \( \forall t_a, t_b \) we say that the random processes \( x(t) \) and \( y(t) \) are uncorrelated.

If the crosscorrelation of two random processes is zero for all times then the random processes are orthogonal.

If \( C(t_a, t_b) = 0 \) \( \forall t_a \neq t_b \) we say that the random process \( x(t) \) is white (we will discuss this again shortly).

If the random variables \( x(t_a) \) and \( y(t_b) \) are independent for all \( t_a \) and \( t_b \) then the processes, \( x(t) \) and \( y(t) \) are independent.

If two processes are independent they are also uncorrelated. The reverse is not normally true. However, if the processes are jointly Gaussian and uncorrelated, they are also independent.

Two processes are jointly Gaussian if all orders (first, second, etc.) of their individual density functions are Gaussian functions and if all orders of their joint density functions are Gaussian.

We say that a stochastic process, \( x(t) \), is \textit{strict sense stationary} if the statistics of \( x(t) \) are independent of a shift in the time origin. The implications of this are that the first order density of \( x(t) \) is independent of \( t \) and all higher order density functions of \( x(t) \) depend only upon relative time. The statements in the above sentence can be written in equation form as: \( f(x, t) = f(x) \), \( f(x_a, x_b, t_a, t_b) = f(x_a, x_b, t_a - t_b) \), \( f(x_a, x_b, x_c, t_a, t_b, t_c) = f(x_a, x_b, x_c, t_a - t_b, t_a - t_c) \) and so forth. As a result of the above we can write the first and second moments as

\[
\eta(t) = E[x(t)] = \int_{-\infty}^{\infty} x f(x, t) dx = \int_{-\infty}^{\infty} x f(x) dx = \eta, \quad (3-34)
\]

\[
R(t + \tau, t) = R(\tau) \quad \text{and} \quad (3-35)
\]

\[
C(t + \tau, t) = C(\tau). \quad (3-36)
\]

We say that two random processes, \( x(t) \) and \( y(t) \) are \textit{jointly strict sense stationary} if each is strict sense stationary and if their joint statistics are independent of a shift in the time origin. If two random processes are jointly strict sense stationary we can write
\( R_{xy}(t + \tau, t) = R_{xy}(\tau) \) and
\( C_{xy}(t + \tau, t) = C_{xy}(\tau) \).  

(3-37)

(3-38)

The requirement for strict sense stationarity is usually more restrictive than necessary. Thus, we define a weaker condition by saying that a random process is \textit{wide sense stationary} if Equations 3-34, 3-35 and 3-36 are valid for that random process. We say that two processes, \( x(t) \) and \( y(t) \), are \textit{jointly wide sense stationary} if they are each wide sense stationary and Equations 3-37 and 3-38 are valid for the two random processes.

As a final definition, we define \textit{white noise} or a \textit{white random process}. A random process, \( x(t) \), is white if
\[
E \{ (x(t) - \eta(t))^2 \} = \sigma^2(t) \delta(\tau)
\]

(3-39)

where \( \delta(\tau) \) is the Dirac delta function.

The above discussions have considered continuous-time systems. The concepts in the previous discussions can be extended to discrete-time systems by replacing the time variables by multiples of some sample period, \( T \). That is, \( t = kT \), \( t_a = k_a T \), \( t + \tau = kT + mT \), and so forth where \( k \), \( k_a \) and \( m \) are integers. When we study sampled data systems we often drop the \( T \) variable and simply use the integers as arguments of the various variables. With this we can write the various means, autocorrelations, etc. of Equations 3-28 through 3-38 as

\[
\eta(k) = E[x(k)] = \int_{-\infty}^{\infty} xf(x,k)dx,
\]

(3-40)

\[
R(k_a, k_b) = E\{x(k_a)x(k_b)\},
\]

(3-41)

\[
C(k_a, k_b) = E\{(x(k_a) - \eta(k_a))(x(k_b) - \eta(k_b))\},
\]

(3-42)

\[
R_{xy}(k_a, k_b) = E\{x(k_a)y(k_b)\},
\]

(3-43)

\[
C_{xy}(k_a, k_b) = E\{(x(k_a) - \eta_x(k_a))(y(k_b) - \eta_y(k_b))\}
\]

(3-44)

and similarly for the remainder.

For white, discrete time random processes we write
\[
C(k + m, k) = E\{(x(k + m) - \eta(k + m))(x(k) - \eta(k))\} = \sigma^2(k)\delta_s(m)
\]

(3-45)

where \( \delta_s(m) \) is the Kronecker delta function and is defined as
\[
\delta_s(m) = \begin{cases} 
1 & \text{if } m = 0 \\
0 & \text{if } m \neq 0.
\end{cases}
\]

(3-46)

\[\text{3.5 Vector Random Processes}\]
When we work with Kalman filters we will generally work with state variables. As such, we will be concerned with vector random processes. That is, collections of random processes that we represent using vectors. Because of this we need to define some vector forms of the mean, autocorrelation, autocovariance, etc. equations that we defined in the previous pages.

To start we represent the mean of a continuous time, vector random process as

$$\eta(t) = \begin{bmatrix} \eta_1(t) \\ \eta_2(t) \\ \vdots \\ \eta_n(t) \end{bmatrix} = E \{ x(t) \} = \begin{bmatrix} E \{ x_1(t) \} \\ E \{ x_2(t) \} \\ \vdots \\ E \{ x_n(t) \} \end{bmatrix}. \quad (3-47)$$

In a similar fashion we can write

$$R(t_a, t_b) = E \{ x(t_a) x^T(t_b) \} = \begin{bmatrix} E \{ x_1(t_a)x_1(t_b) \} & E \{ x_1(t_a)x_2(t_b) \} & \cdots & E \{ x_1(t_a)x_n(t_b) \} \\ E \{ x_2(t_a)x_1(t_b) \} & E \{ x_2(t_a)x_2(t_b) \} & \cdots & E \{ x_2(t_a)x_n(t_b) \} \\ \vdots & \vdots & \ddots & \vdots \\ E \{ x_n(t_a)x_1(t_b) \} & E \{ x_n(t_a)x_2(t_b) \} & \cdots & E \{ x_n(t_a)x_n(t_b) \} \end{bmatrix} \quad (3-48)$$

The $T$ superscript notation in Equation 3-48 denotes the transpose operation. It will be noted that the last matrix of Equation 3-85 contains both autocorrelation and crosscorrelation terms. As a result, we term this matrix the autocorrelation matrix.

We will often need to discuss the crosscorrelation between two, different, vector random processes. This is again a matrix, which we will term the crosscorrelation matrix. This matrix has the form

$$R_{xy}(t_a, t_b) = E \{ x(t_a)y^T(t_b) \} = \begin{bmatrix} E \{ x_1(t_a)y_1(t_b) \} & E \{ x_1(t_a)y_2(t_b) \} & \cdots & E \{ x_1(t_a)y_m(t_b) \} \\ E \{ x_2(t_a)y_1(t_b) \} & E \{ x_2(t_a)y_2(t_b) \} & \cdots & E \{ x_2(t_a)y_m(t_b) \} \\ \vdots & \vdots & \ddots & \vdots \\ E \{ x_n(t_a)y_1(t_b) \} & E \{ x_n(t_a)y_2(t_b) \} & \cdots & E \{ x_n(t_a)y_m(t_b) \} \end{bmatrix}$$
\[
\begin{bmatrix}
R_{x_1y_1}(t_a,t_b) & R_{x_1y_2}(t_a,t_b) & \cdots & R_{x_1y_n}(t_a,t_b) \\
R_{x_2y_1}(t_a,t_b) & R_{x_2y_2}(t_a,t_b) & \cdots & R_{x_2y_n}(t_a,t_b) \\
\vdots & \vdots & \ddots & \vdots \\
R_{x_ny_1}(t_a,t_b) & R_{x_ny_2}(t_a,t_b) & \cdots & R_{x_ny_n}(t_a,t_b)
\end{bmatrix}
\]  

(3-49)

It will be noted that this equation contains only crosscorrelation terms. It will also be noted that, while the autocorrelation matrix is always square, the crosscorrelation matrix is an \( n \times m \) matrix and is most often not square.

We define the autocovariance matrix by the equation

\[
C(t_a,t_b) = E\left\{ (x(t_a) - \eta(t_a))(x(t_b) - \eta(t_b))^T \right\}.
\]  

(3-50)

If we let \( t_a = t_b = t \) we obtain the covariance matrix

\[
C(t,t) = E\left\{ (x(t) - \eta(t))(x(t) - \eta(t))^T \right\} = P(t).
\]  

(3-51)

Finally, we write the crosscovariance matrix as

\[
C_{xy}(t_a,t_b) = E\left\{ (x(t_a) - \eta_x(t_a))(y(t_b) - \eta_y(t_b))^T \right\}
\]  

(3-52)

We can define the sampled data versions of Equations 3-47 through 3-52 by replacing the time variable, \( t \), with the stage variable, \( k \).

Some notes on vector random processes

- Two vector random processes are uncorrelated if, and only if,
  \( C_{xy}(t_a,t_b) = 0 \ \forall t_a, t_b \).

- A continuous-time, vector random process is white if, and only if,
  \( C(t + \tau,t) = P(t)\delta(\tau) \).

- A discrete-time, vector random process is white if, and only if,
  \( C(k + m,k) = P(k)\delta_s(m) \).

- A vector random process is Gaussian if all the elements of the vector are jointly Gaussian.
3.6 Problems

1. Derive the relations given by Equations 3-5 and 3-6.

2. Use Equations 3-10 and 3-11 to show that \( f(x) \geq 0 \ \forall x \) and \( \int_{-\infty}^{\infty} f(x) \, dx = 1 \).

3. Show that the Gaussian density of Equation 3-12 has a mean of \( \eta \).

4. Show that \( R(x, y) = C(x, y) \) if \( \eta_x = 0 \) and/or \( \eta_y = 0 \).

5. Show that two independent random variables are also uncorrelated.

6. Show that two uncorrelated, jointly Gaussian random variables are also independent.

7. Show that two independent random processes are also uncorrelated.

8. Show that two uncorrelated, jointly Gaussian random processes are also independent.

9. Show that \( R_{xy}^T(k_a, k_b) = R_{yx}(k_b, k_a) \).

10. Show that \( P(k) \) is a non-negative definite matrix.