C

Review of Probability Theory

In this appendix we review some notions in probability theory that are used in Chapter 9 in connection with the Kalman filter. A good reference is Papoulis [1984].

C.1 MEAN AND VARIANCE

Given a random vector $z \in \mathbf{R}^n$, we denote by $f_z(\zeta)$ the *probability density function* (PDF) of z. The PDF represents the probability that z takes on a value within the differential region $d\zeta$ centered at ζ. Although the value of z may be unknown, it is quite common in many situations to have a good feel for its PDF.

The *expected value* of a function $g(z)$ of a random vector z is defined as

$$E\{g(z)\} = \int_{-\infty}^{\infty} g(\zeta) f_z(\zeta)\, d\zeta. \tag{C.1.1}$$

In particular, the *mean* or *expected value* of z is defined by

$$E\{z\} = \int_{-\infty}^{\infty} \zeta f_z(\zeta)\, d\zeta, \tag{C.1.2}$$

which we shall symbolize by $\bar{z}$ to economize on notation. Note that $\bar{z} \in \mathbf{R}^n$.

Note that the expectation operator is linear, so that, given two random variables x and z and two deterministic scalars a and b

$$\overline{ax + bz} = a\bar{x} + b\bar{z}. \tag{C.1.3}$$

The *covariance* of $z \in \mathbf{R}^n$ is given by

$$P_z = E\{(z - \bar{z})(z - \bar{z})^T\}. \tag{C.1.4}$$

Note that P_z is an $n \times n$ positive definite (constant) matrix. Using the linearity property, it is seen that

$$P_z = E\{zz^T\} - \bar{z}\bar{z}^T. \tag{C.1.5}$$

We call $E\{zz^T\}$ the *mean-square value* of z.

An important class of random vectors is characterized by the *Gaussian* or *normal* PDF

$$f_z(\zeta) = \frac{1}{\sqrt{(2\pi)^n|P_z|}} e^{-1/2(\zeta-\bar{z})^T P_z^{-1}(\zeta-\bar{z})}, \tag{C.1.6}$$

where in general $z \in \mathbf{R}^n$. In the scalar case $n = 1$ this reduces to the more familiar

$$f_z(\zeta) = \frac{1}{\sqrt{2\pi P_z}} e^{-(\zeta-\bar{z})^2/2P_z}. \tag{C.1.7}$$

Such random vectors take on values near the mean $\bar{z}$ with greatest probability, and have a decreasing probability of taking on values farther away from $\bar{z}$.

Many naturally occurring random variables are Gaussian. Increased importance is given to Gaussian PDF by *Central-Limit Theorem*, which states that the sum of a large number of random variables has approximately a Gaussian PDF, regardless of the distributions of the individual random variables.

C.2 TWO RANDOM VARIABLES

Given random vectors $z \in \mathbf{R}^n$, $x \in \mathbf{R}^m$ we denote by $f_{z,x}(\zeta, \xi)$ the *joint probability density function* of z and x. The joint PDF represents the probability that z takes on a value within the differential region $d\zeta$ centered at ζ and x takes on a value within the differential region $d\xi$ centered at ξ. An example of a joint PDF of two scalar random variables z_1 and z_2 is provided by (C.1.6) when $z = [z_1 \quad z_2]^T$.

The *expected value* of a function $g(z, x)$ of two random vectors z and x is defined as

$$E\{g(z, x))\} = \int_{-\infty}^{\infty} g(\zeta, \xi) f_{z,x}(\zeta, \xi)\, d\zeta d\xi. \tag{C.2.1}$$

The *cross-covariance* of two random variables $z \in \mathbf{R}^n$ and $x \in \mathbf{R}^m$ is defined as

$$P_{zx} = E\{(z - \bar{z})(x - \bar{x})^T\} \tag{C.2.2}$$

which is an $n \times m$ constant matrix.

The *conditional PDF* of x given z is given by

$$f_{x/z}(\xi/z = \zeta) = f_{zx}(\zeta, \xi)/f_z(\zeta). \tag{C.2.3}$$

The *conditional mean* of x given z is a random variable denoted by $\overline{x/z}$ and defined by the functional dependence

$$\overline{x/z=\zeta} = \int_{-\infty}^{\infty} \xi f_{x/z}(\xi/z=\zeta)\, d\xi. \tag{C.2.4}$$

Two random variables are said to be *independent* if

$$f_{z,x}(\zeta, \xi) = f_z(\zeta) f_x(\xi), \tag{C.2.5}$$

uncorrelated if

$$E\{zx^T\} = \overline{z}\overline{x}^T, \tag{C.2.6}$$

and *orthogonal* if

$$E\{zx^T\} = 0. \tag{C.2.7}$$

Independence implies uncorrelatedness. For normal random variables, these two properties are equivalent.

C.3 RANDOM PROCESSES

If the random vector is a time function it is called a *random process*, symbolized as $z(t)$. Then, the PDF may also be time varying and we write $f_z(\zeta, t)$. In this situation, the expected value and covariance matrix are also functions of time, so we write $\overline{z}(t)$ and $P_z(t)$. In discrete time, we write z_k, $\overline{z}_k$, $P_z(k)$ and so on.

Many random processes $z(t)$ of interest to us have a time-invariant PDF. These are *stationary* processes and, even though they are random time functions, they have a constant mean and covariance.

To characterize the relation between two random processes $z(t)$ and $x(t)$ we employ the joint PDF $f_{zx}(\zeta, \xi, t_1, t_2)$, which represents the probability that $(z(t_1), x(t_2))$ is within the differential area $d\zeta \; x \; d\xi$ centered at (ζ, ξ). We shall usually assume that the processes $z(t)$, $x(t)$ are *jointly stationary*, that is, the joint PDF is not a function of both times t_1 and t_2, but depends only on the difference $(t_1 - t_2)$.

In the stationary case, the expected value of the function of two variables $g(z, x)$ is defined as

$$E\{g(z(t_1), x(t_2))\} = \int_{-\infty}^{\infty} g(\zeta, \xi) f_{z,x}(\zeta, \xi, t_1 - t_2)\, d\zeta d\xi. \tag{C.3.1}$$

In particular, the *cross-correlation function* is defined by

$$R_{zx}(\tau) = E\{z(t + \tau)x^T(t)\}. \tag{C.3.2}$$

The cross-correlation function of two nonstationary processes is defined as

$$R_{zx}(t, \tau) = E\{z(t)x^T(\tau)\}. \tag{C.3.3}$$

Considering $z(t_1)$ and $z(t_2)$ as two jointly distributed random stationary processes, we may define the *autocorrelation function* of $z(t)$ as

$$R_z(\tau) = E\{z(t + \tau)z^T(t)\} \tag{C.3.4}$$

The autocorrelation function gives us some important information about the random process $z(t)$. For instance

$$\text{trace}[R_z(0)] = \text{trace}[E\{z(t)z^T(t)\}] = E\{\|z(t)\|^2\}$$

is equal to the total energy in the process $z(t)$. (In writing this equation recall that, for any compatible matrices M and N, trace(MN) = trace(NM).)

If

$$R_{zx}(\tau) = 0 \tag{C.3.5}$$

we call $z(t)$ and $x(t)$ *orthogonal*. If

$$R_z(\tau) = P\delta(\tau) \tag{C.3.6}$$

where P is a constant matrix and $\delta(t)$ is the dirac delta, then $z(t)$ is orthogonal to $z(t + \tau)$ for any $\tau \neq 0$. What this means is that the value of the process $z(t)$ at one time t is unrelated to its value at another time $\tau \neq t$. Such a process is called *white noise*. An example is the thermal noise in an electric circuit, which is due to the thermal agitation of the electrons in the resistors.

In the discrete-time case, for these quantities we write $R_{zx}(k)$, and so on.

C.4 SPECTRAL DENSITY AND LINEAR SYSTEMS

For a stationary random process $x(t)$, the *spectral density* is defined as

$$\phi_x(s) = \mathbf{L}(R_x(\tau)), \tag{C.4.1}$$

that is, the Laplace transform of the autocorrelation function. In discrete time

$$\phi_x(z) = \mathbf{Z}(R_x(k)), \tag{C.4.2}$$

the Z-transform of the autocorrelation function. The *cross-spectral density* $\phi_{xy}(s)$ of two processes $x(t)$ and $y(t)$ is likewise defined in terms of the cross-correlation function.

Given $u(t) \in \mathbf{R}^m$ and $y(t) \in \mathbf{R}^p$ the input and output respectively of a linear system with impulse response $h(t)$ and transfer function $H(s)$, we have

$$\begin{aligned}
R_{yu}(t) &= h(t)*R_u(t) \\
R_y(t) &= h(t)*R_u(t)*h^T(-t) \\
\phi_{yu}(s) &= H(s)\phi_u(s) \\
\phi_y(s) &= H(s)\phi_u(s)H^T(-s),
\end{aligned} \tag{C.4.3}$$

with * denoting convolution.

According to (C.3.6), for white noise $P\delta(0)$ is the covariance of $z(t)$, which is unbounded. Since

$$\phi_z(s) = P, \tag{C.4.4}$$

we call P a *spectral density matrix*. It is sometimes loosely referred to as a covariance matrix. In the discrete case, however,

$$R_z(k) = P\delta_k, \tag{C.4.5}$$

with δ_k the Kronecker delta, which is bounded. Therefore, P is a covariance matrix.

Applied Optimal Control & Estimation

Digital Design & Implementation

Frank L. Lewis

The University of Texas at Arlington

1992

TEXAS INSTRUMENTS

Prentice Hall, Englewood Cliffs, New Jersey 07632