

Chapter 2

Basic Concepts in Probability

In this chapter, some basic concepts required for the detection and estimation of signals are presented from the theory of probability, statistics and random processes. The reader is assumed to have some basic knowledge of the theory of probability. Some concepts on multivariate statistics and random vectors are also presented. For a more thorough treatise of the subject, readers should consult the noted reference books [73, 76].

2.1 Probability Distributions and Densities of Random Variables

Papoulis [73] defines a random variable x as a real function whose domain is the probability space S and such that the following is true:

- The set $\{x \leq X\}$ is an event for any real number X .
- The probability of the events $\{x = +\infty\}$ and $\{x = -\infty\}$ equals zero.

A random variable is called either a continuous random variable or a discrete random variable depending on whether it takes real values or discrete values, respectively. The *cumulative distribution function (cdf)* F_x of a random variable x at point $x = X_0$ is defined as the probability that $x \leq X_0$:

$$F_x(X_0) = P(x \leq X_0). \quad (2.1)$$

Changing the value of X_0 from $-\infty$ to ∞ in (2.1) gives the whole *cdf* for all values of x . From the definition of the random variable and *cdf*, it follows directly that $F_x(-\infty) = 0$ and $F_x(\infty) = 1$. A more widely used function is the *probability density function (pdf)*, which is defined for the continuous random variable x as the derivative of the cumulative distribution function:

$$p_x(x_0) = \left. \frac{dF_x(x)}{dx} \right|_{x=x_0} \quad (2.2)$$

The *pdf* is the probability that the variable has the value x . Often the probability density function is available or estimated in real applications. The cumulative distribution function can be computed from the *pdf* by the inverse relationship

$$F_x(X_0) = \int_{-\infty}^{X_0} p_x(\xi) d\xi. \quad (2.3)$$

One of the most widely used *pdfs* in signal processing is the normal distribution. It is characterized by two parameters, mean and variance. The normal distribution is given by

$$p_x(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \triangleq \mathcal{N}_x(\mu, \sigma^2) \quad (2.4)$$

where μ and σ^2 denotes the mean and the variance, respectively. Figure 2.1a shows the plot of the normal distribution for $\mu = 0$ and $\sigma = 1$ and the corresponding *cdf* is given in Fig. 2.1b. From the *cdf*, the probabilities of a random variable taking a value or a set of values can be obtained. For example, the probability of the random variable x taking the value less than or equal to 1.0 is found to be 0.84 from Fig. 2.1b, that is, $p(x \leq 1) = 0.84$.

There are several other distributions that are widely used in practice depending on the data collected and its statistical nature, including Bernoulli, Poisson, Rayleigh, uniform, Weibull distributions, to name few.

The mean (μ) and variance (σ^2) in (2.4) are also called the first and second moments and, in general, the k th moments are defined for continuous random variable as

$$\mu_k = E\{x^k\} = \int_{-\infty}^{+\infty} x^k p(x) dx \quad (2.5)$$

In the case of discrete random variable, the mean and the variances are computed as

$$\mu = \sum_{i=1}^N P(x_i) x_i \quad (2.6)$$

and

$$\sigma^2 = \sum_{i=1}^N P(x_i) (x_i - \mu)^2 \quad (2.7)$$

where N is the number of samples x_i , $i = \{1, 2, \dots, N\}$. If the random variable x_i is equally likely, then $P(x_i) = \frac{1}{N}$, then (2.6) becomes

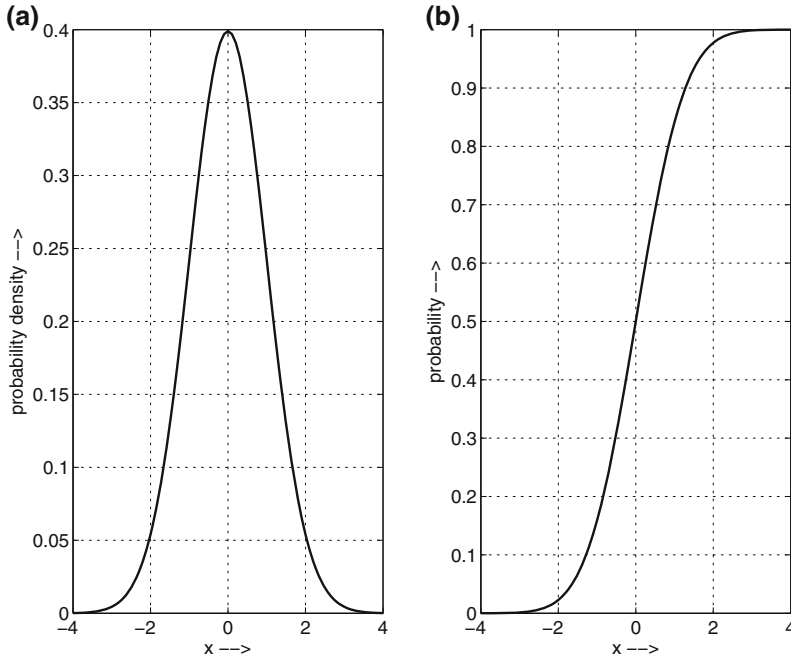


Fig. 2.1 **a** Normal distribution, **b** cumulative distribution

$$\mu = \frac{1}{N} \sum_{i=1}^N x_i \quad (2.8)$$

When the variance exists, then the standard deviation is given by σ . For a normal distribution with mean $\mu = 0$ and $\sigma = 1$, as shown in Fig. 2.1a, the probability that the random variable x lies between ± 1 , that is, $p(-1 \leq x \leq 1)$ is given by the area under the curve in Fig. 2.1a lying between $x = -1$ and $x = 1$ and it can be shown to be 68.2 % of the area under the curve and $p(-2 \leq x \leq 2)$ is 95.4 % of the area under the curve. This fact is often used in signal processing while selecting thresholds for various parameters.

Let $g(x)$ is a function of analog random variable x , then the expectation of $g(x)$ is denoted by $E[g(x)]$ and is defined by

$$E[g(x)] = \int_{-\infty}^{+\infty} g(x)p(x)dx \quad (2.9)$$

2.2 Bernoulli Distribution

Suppose an experiment is conducted with its outcome determined as “success” or “failure”. Let the probability of success is denoted by p , then the probability of failure is $(1 - p)$. An example of such an experiment is tossing of a coin, where if the coin toss results in “heads” it is called success and if the toss results in “tails” it is called failure. Other experiments could be the outcome of surgery, throwing a ball through hoops, the transmission of a disease, etc. Thus, the Bernoulli distribution is defined as

$$f(x) = p^x (1 - p)^{1-x}, \quad \text{for } x = 0, 1. \quad (2.10)$$

If “ n ” tosses of a coin are done, then the probability that “ k ” successive tosses result in heads is given by the probability mass function

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}. \quad (2.11)$$

2.3 Random Vectors

In general, more useful cases arise when there are multiple random variables. For example, in order to classify a target, one may consider several features that are pertinent to the target. These features can be denoted as a random vector

$$\mathbf{X} = (x_1, x_2, \dots, x_n)^T \quad (2.12)$$

where T denotes the transpose and n is the number of elements in $\mathbf{X}$. Just as in the case of a single variable, the *cdf* of a vector $\mathbf{X}$ is defined by

$$F_{\mathbf{X}}(\mathbf{X}_0) = P(\mathbf{X} \leq \mathbf{X}_0) \quad (2.13)$$

where $P(\cdot)$ denotes the probability and $\mathbf{X}_0$ is some constant vector. The multivariate probability density function $p_{\mathbf{X}}(\mathbf{X})$ of $\mathbf{X}$ is defined as the derivative of the cumulative distribution function $F_{\mathbf{X}}(\mathbf{X}_0)$ with respect to the components of $\mathbf{X}$:

$$p_{\mathbf{X}}(\mathbf{X}_0) = \frac{\partial}{\partial x_1} \frac{\partial}{\partial x_2} \dots \frac{\partial}{\partial x_n} F_{\mathbf{X}}(\mathbf{X}) \Big|_{\mathbf{X}=\mathbf{X}_0} \quad (2.14)$$

2.3.1 Joint and Marginal Distributions

Let us assume that there are two different sensors, say, acoustic and seismic sensors. Let the feature vector for the acoustic sensor be denoted by the random vector

$\mathbf{X} = (x_1, x_2, \dots, x_n)^T$ of length ' n ' and the seismic sensor be denoted by the random vector $\mathbf{Y} = (y_1, y_2, \dots, y_m)^T$ of length ' m ', then the joint distribution function of two random vectors $\mathbf{X}$ and $\mathbf{Y}$ is given by

$$F_{\mathbf{X}, \mathbf{Y}}(\mathbf{X}_0, \mathbf{Y}_0) = P(\mathbf{X} \leq \mathbf{X}_0, \mathbf{Y} \leq \mathbf{Y}_0) \quad (2.15)$$

for some fixed vectors $\mathbf{X}_0$ and $\mathbf{Y}_0$. Let $\mathbf{Z} = (\mathbf{X}^T, \mathbf{Y}^T)$ be the super vector, then the joint density function $p_{\mathbf{Z}}(\mathbf{Z}) = p_{\mathbf{X}, \mathbf{Y}}(\mathbf{X}, \mathbf{Y})$ of $\mathbf{X}$ and $\mathbf{Y}$ is obtained by differentiating the joint distribution function $F_{\mathbf{X}, \mathbf{Y}}(\mathbf{X}, \mathbf{Y})$ as in (2.14) with respect to all the components of $\mathbf{X}$ and $\mathbf{Y}$. Hence,

$$F_{\mathbf{X}, \mathbf{Y}}(\mathbf{X}_0, \mathbf{Y}_0) = \int_{-\infty}^{\mathbf{X}_0} \int_{-\infty}^{\mathbf{Y}_0} p_{\mathbf{X}, \mathbf{Y}}(\eta, \xi) d\eta d\xi \quad (2.16)$$

The density functions with respect to a subset of $\mathbf{Z}$ are called the marginal density and are denoted by $p_{\mathbf{X}}(\mathbf{X})$ for $\mathbf{X}$ and $p_{\mathbf{Y}}(\mathbf{Y})$ for $\mathbf{Y}$ and given by

$$p_{\mathbf{X}}(\mathbf{X}) = \int_{-\infty}^{+\infty} p_{\mathbf{X}, \mathbf{Y}}(\mathbf{X}, \eta) d\eta \quad (2.17)$$

$$p_{\mathbf{Y}}(\mathbf{Y}) = \int_{-\infty}^{+\infty} p_{\mathbf{X}, \mathbf{Y}}(\xi, \mathbf{Y}) d\xi \quad (2.18)$$

2.3.2 Mean Vector and Correlation Matrix

Some of the general features of the target used for classification are the mean values of the individual features. Using (2.5)

$$\mathbf{m}_{\mathbf{X}} = \mu_{\mathbf{X}} = E[\mathbf{X}] = \int_{-\infty}^{+\infty} \mathbf{X} p_{\mathbf{X}}(\mathbf{X}) d\mathbf{X} \quad (2.19)$$

The mean vector $\mathbf{m}_{\mathbf{X}} = (m_{x_1}, m_{x_2}, \dots, m_{x_n})^T$ and the individual components are given by

$$m_{x_i} = E[x_i] = \int_{-\infty}^{+\infty} x_i p_{\mathbf{X}}(\mathbf{X}) d\mathbf{X} = \int_{-\infty}^{+\infty} x_i p_{x_i}(x_i) dx_i \quad (2.20)$$

where $p_{x_i}(x_i)$ is the marginal density of the i th component x_i of $\mathbf{X}$. Equation (2.20) is true due to the fact that the integrals over the other components of $\mathbf{X}$ reduce to unity by the definition of the marginal density given by (2.17) and (2.18).

As shown in Fig. 2.1, the variance gives a measure of the spread of the data or the random variable. Similarly, covariance of two random variables give the measure how two variables vary together. The covariance of two random variables is defined as expectation

$$C_{x_i, x_j} = \text{cov}(x_i, x_j) = E[(x_i - \mu_i)(x_j - \mu_j)] \quad (2.21)$$

where μ_i and μ_j denotes the mean values of random variable x_i and x_j , respectively. The covariance matrix $\text{cov}(x_i, x_j)$ is a 2×2 matrix whose ij th element is $E[(x_i - \mu_i)(x_j - \mu_j)]$. Note that the diagonal elements are the variances of the individual random variables $E[(x_i - \mu_i)^2]$. The off-diagonal elements $E[(x_i - \mu_i)(x_j - \mu_j)]$, $i \neq j$, are called the covariance of x_i and x_j .

If the random vector Y is a linear transform of another vector X

$$Y = AX$$

then

$$E[Y] = AE[X] \quad (2.22)$$

and

$$\text{cov}(Y) = A \text{cov}(X) A^T \quad (2.23)$$

Another statistical parameter that is closely related to the covariance is the correlation and it is given by the second moment

$$r_{ij} = E[x_i x_j] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x_i x_j p_{\mathbf{X}}(\mathbf{X}) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x_i x_j p_{x_i, x_j}(x_i, x_j) dx_j dx_i \quad (2.24)$$

If $\mathbf{X} = (x_1, x_2, \dots, x_n)^T$, then the *correlation matrix* is given by

$$R_{\mathbf{X}} = E[\mathbf{X}\mathbf{X}^T] \quad (2.25)$$

and the component r_{ij} in i th row and j th column of matrix $R_{\mathbf{X}}$ is the correlation between x_i and x_j . When $\mathbf{m}_{\mathbf{X}} = 0$ the correlation and covariance matrices become identical, that is, $R_{\mathbf{X}} = C_{\mathbf{X}}$.

Uncorrelatedness and Whiteness: The random vectors X_1 and X_2 are said to be uncorrelated if the cross-correlation matrix C_{X_1, X_2} is a zero matrix:

$$C_{X_1, X_2} = \text{cov}(X_1, X_2) = E[(X_1 - m_1)(X_2 - m_2)] = 0 \quad (2.26)$$

It can be shown that for uncorrelated random vectors, the correlation matrix is

$$R_{X_1, X_2} = E[X_1 X_2^T] = E[X_1] E[X_2^T] = m_{X_1} m_{X_2}^T \quad (2.27)$$

In particular if the random vectors have zero mean and unit covariance then they are said to be white, that is,

$$C_{X_1, X_2} = R_{X_1, X_2} = I; \quad m_{X_1} = m_{X_2} = 0. \quad (2.28)$$

2.3.3 Independence of Variables

The independence of random variables is a very important concept that is often invoked in signal processing to simplify matters. Two random variables x and y are said to be independent if

$$P(x \leq x_0, y \leq y_0) = P(x \leq x_0) P(y \leq y_0) \quad (2.29)$$

Similarly, multiple random variables are said to be independent if

$$P(x \leq x_0, y \leq y_0, \dots, z \leq z_0) = P(x \leq x_0) P(y \leq y_0) \dots P(z \leq z_0) \quad (2.30)$$

2.3.4 Conditional Probabilities and Bayes Theorem

Let X and Y be two random events with probabilities $P(X) > 0$ and $P(Y) > 0$, then the probability of X assuming Y , is defined as

$$P(X|Y) = \frac{P(XY)}{P(Y)} \quad (2.31)$$

Rewriting the above equation, we get the product rule

$$P(A, B) = P(A|B) P(B). \quad (2.32)$$

We can extend this for three variables:

$$P(A, B, C) = P(A|B, C) P(B|C) P(C). \quad (2.33)$$

In general, the above equation can be extended to n variables to obtain the chain rule

$$P(A_1, A_2, \dots, A_n) = P(A_1|A_2, \dots, A_n) P(A_2|A_3, \dots, A_n) \dots P(A_{n-1}|A_n) P(A_n) \quad (2.34)$$

2.4 Cross Correlation

Cross correlation is a widely used concept in signal processing to determine the time difference between two signals recorded by two microphones in an array of microphones—often used to estimate the direction of arrival angle of the signal. The correlation between two signals $f(t)$ and $g(t)$ is defined as

$$r_{fg} = (f \star g)(t) \triangleq \int_{-\infty}^{\infty} f^*(\tau)g(t + \tau) d\tau, \quad (2.35)$$

where $f^*(t)$ is the complex conjugate of $f(t)$. In signal processing, the cross correlation is a measure of similarity between two waveforms (or signals) as a function of time lag. For discrete signals the cross correlation is given by

$$r_{fg} = (f \star g)[n] \triangleq \sum_{m=-\infty}^{m=\infty} f^*[m]g[n + m]. \quad (2.36)$$

The following example shows cross correlation of two signals and estimation of time delay between the two signals.

Example: Let $f[n]$ is the original signal and $g[n]$ is the delayed and attenuated signal without any noise, as shown in Fig. 2.2a, b, respectively. Each signal is of $N_s = 2048$ samples long. Their cross correlation r_{fg} is shown in (c) with cross correlation lag. The signals are sampled at a sampling rate of $f_s = 20,000$ samples/s. In order to compute the time delay between the two signals find the index k at which r_{fg} is

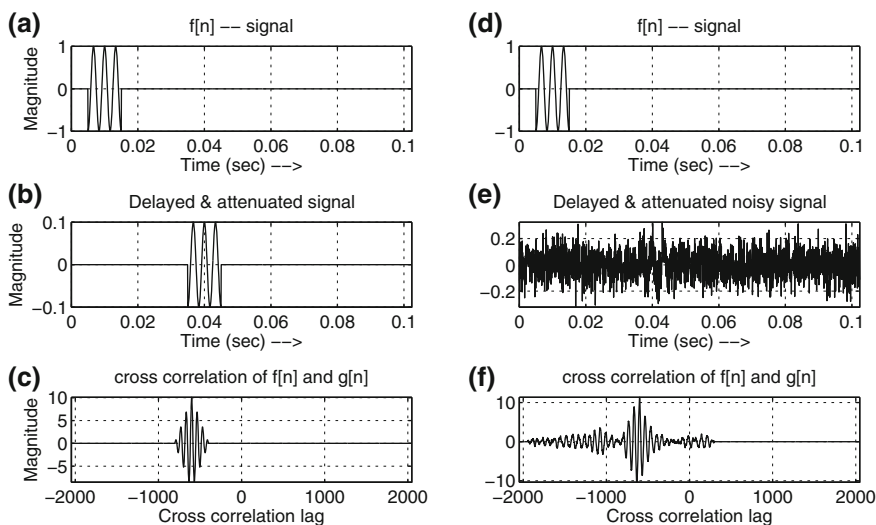


Fig. 2.2 Cross correlation of signals

maximum, then $(N_s - k)/f_s$ gives the time delay between the signals f and g . For the signals shown in Fig. 2.2a, b, the delay is estimated to be 0.03 s. A new signal $z(t) = g(t) + n(t)$ is generated by adding noise $n(t)$ to the signal $g(t)$ and is shown in Fig. 2.2e. The cross correlation r_{fz} is shown in Fig. 2.2f and the time delay is computed. The time delay is found to be 0.03 s. In spite of the signal $z(t)$ being very noisy, we are able to correctly estimate the time delay between the two signals using the cross correlation technique.

Battlefield Acoustics

Damarla, T.

2015, XV, 262 p. 158 illus., 90 illus. in color., Hardcover

ISBN: 978-3-319-16035-1