APPENDIX B

Statistical Measures, Probability, and Random Processes

Paul D. Groves

This appendix provides a refresher on statistical measures, probability, and random processes to support the main body of the book. Section B.1 defines the mean, variance, and standard deviation as statistical measures of data. Section B.2 describes probability, including probability density, multiple variables, summed and differenced variables, and the radial error. Section B.3 describes some standard distributions, including the Gaussian and chi-square distributions. Section B.4 describes random processes, including white noise and Markov processes. Finally, Section B.5 describes hypothesis testing.

A number of textbooks [1–6] cover the material presented here in more detail.

B.1 Statistical Measures of Data

The mean, variance, and standard deviation are the most basic statistical measures. The *mean*, μ_x, and *variance*, σ_x^2, of a set of n observations, x_i, are defined as

$$\mu_x = \frac{1}{n}\sum_{i=1}^{n} x_i, \qquad \sigma_x^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu_x)^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \mu_x^2 . \tag{B.1}$$

They may also be written as $\mu(x)$ and $\sigma^2(x)$, respectively The *standard deviation* (SD), σ_x or $\sigma(x)$, is simply the square root of the variance, while the root mean square (RMS) is

$$r_x = \sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2} . \tag{B.2}$$

The mean and variance form part of a wider set of statistical measures called moments. The k^{th}-order moment, m_k, and central moment, c_k, are defined as

$$m_k(x) = \frac{1}{n}\sum_{i=1}^{n} x_i^k, \qquad c_k(x) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu_x)^k . \tag{B.3}$$

Note that c_1 is always zero. The mean is the first-order moment and the variance the second-order central moment. The RMS is the square root of the second-order moment. The third-order central moment is known as the skew.

Suppose the observations, x_i, are taken from a population with distribution X. Unbiased estimates of the mean, μ_X, and variance, σ_X^2, of the population's distribution are

$$\hat{\mu}_X = \mu_x, \qquad \hat{\sigma}_X^2 = \frac{n}{n-1}\sigma_x^2 . \tag{B.4}$$

The mean of a set of vector observations is simply the vector of the means of the components. The equivalent of the variance for a vector distribution is the covariance matrix, $\mathbf{C}$ ($\mathbf{P}$ may also be used). If $\mathbf{x}$ is an m-component vector, $\mathbf{C}_\mathbf{x}$ is an $m \times m$ matrix. A covariance matrix is symmetric with the diagonal elements corresponding to the variances of the components of the vector it describes and the off-diagonal elements giving the covariance of each pair of components. Where the components of the vector are uncorrelated, the off-diagonal elements of $\mathbf{C}$ will tend to zero as the number of observations increases.

The mean, $\boldsymbol{\mu}_\mathbf{x}$, and covariance, $\mathbf{C}_\mathbf{x}$, of a set of n observations, $\mathbf{x}_i$, are

$$\boldsymbol{\mu}_\mathbf{x} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i, \qquad \mathbf{C}_\mathbf{x} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_i - \boldsymbol{\mu}_\mathbf{x})(\mathbf{x}_i - \boldsymbol{\mu}_\mathbf{x})^\mathrm{T} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i^\mathrm{T} - \boldsymbol{\mu}_\mathbf{x}\boldsymbol{\mu}_\mathbf{x}^\mathrm{T} . \tag{B.5}$$

Higher-order moments can also describe vector data, noting that a k^{th}-order moment or central moment of a vector is a k^{th}-order tensor.

B.2 Probability

Consider an event, A, that occurs in some trials but not others. If n_A is the number of trials where event A occurs and n is the total number of trials, the probability of event A occurring in a trial may be written as

$$P(A) = \lim_{n\to\infty}\frac{n_A}{n} . \tag{B.6}$$

The probability of event A not occurring is $1 - P(A)$. Note that all probabilities are positive or zero.

Suppose the outcome of a trial is one of m mutually exclusive events, denoted A_1 to A_m. In any given trial, at least one and no more than one of these events will occur. Their probabilities thus obey the condition

$$\sum_{i=1}^{m} P(A_i) = 1 . \tag{B.7}$$

Consider a variable, x, that can take one of m discrete values, denoted x_1 to x_m. Suppose that x is taken from a population with distribution X, described by the discrete function p_X. The probability that x takes a given value, $P(x = x_i)$, is then $p_X(x_i)$. From above, the condition

$$\sum_{i=1}^{m} p_X(x_i) = \sum_{i=1}^{m} P(x = x_i) = 1 \tag{B.8}$$

must be met.

Consider now a variable, x, that is continuous. It can take an infinite number of values, each of which has an infinitesimal probability. Thus, the probability of x taking a discrete value is meaningless. Instead, its probability distribution is described by a *probability density function* (PDF) such that the probability of x lying between two values is the integral of the PDF between those values. A PDF must be positive or zero for all values of its argument. If x

is taken from a population with distribution X, described by the PDF f_X, the probability that it takes a value between x_- and x_+ is then

$$P(x_- \le x \le x_+) = \int_{x_-}^{x_+} f_X(x)dx \,. \tag{B.9}$$

As with discrete variables, the total probability of a continuous variable taking all values is always unity. Therefore, by analogy with (B.8), all PDFs obey the condition

$$\int_{-\infty}^{\infty} f(x)dx = 1 \,. \tag{B.10}$$

The PDF may also be considered a measure of the relative probability that a sample, x, from the distribution the PDF describes takes a certain value. Thus

$$\lim_{\delta x \to 0} \frac{P(a - \delta x \le x \le a + \delta x)}{P(b - \delta x \le x \le b + \delta x)} = \frac{f_X(a)}{f_X(b)} \,. \tag{B.11}$$

The *cumulative distribution function* (CDF) of x is defined as

$$F_X(X) = p(-\infty \le x \le X) = \int_{-\infty}^{X} f_X(x)dx \,. \tag{B.12}$$

Conversely, the PDF is the derivative of the CDF. Thus,

$$f_X(x) = \frac{dF_X(x)}{dx} \,. \tag{B.13}$$

Probability may be obtained directly from the CDF using

$$P(x_- \le x \le x_+) = F_X(x_+) - F_X(x_-) \,. \tag{B.14}$$

The median, $x_{0.5}$, can be obtained from the CDF by solving $F_X(x_{0.5}) = 0.5$. Similarly, the CDF can be obtained from a set of data by ordering the observations in increasing order. Thus, if x_1 is the smallest and x_n the largest of a set of n observations, the CDF is given by $F_x(x_i) = i/n$. Note that this is the CDF of the data set, not the population from which the data is taken.

The mean of a random variable is determined by applying the expectation operator, $\mathrm{E}(\,)$, to that variable. This is

$$\mu_x = \mathrm{E}(x) = \sum_{i=1}^{m} x_i p_X(x_i) \tag{B.15}$$

in the discrete case and

$$\mu_x = \mathrm{E}(x) = \int_{-\infty}^{\infty} x f_X(x)dx \tag{B.16}$$

in the continuous case. The mean of a parameter is sometimes referred to simply as the expectation of that parameter. However, the expectation operator may be applied to any function of a random variable with a known distribution. Thus, for the discrete case,

$$\mathrm{E}[g(x)] = \sum_{i=1}^{m} g(x_i) p_X(x_i), \tag{B.17}$$

while for the continuous case

$$\mathrm{E}[g(x)] = \int_{-\infty}^{\infty} g(x) f_X(x) dx \,. \tag{B.18}$$

The expectation operator can thus be used to determine the moments:

$$m_k(x) = \mathrm{E}\left(x^k\right), \qquad c_k(x) = \mathrm{E}\left[\left(x - \mu_x\right)^k\right]. \tag{B.19}$$

The variance of a random variable is thus

$$\sigma_x^2 = \mathrm{E}\left[\left(x - \mu_x\right)^2\right] = \mathrm{E}\left(x^2\right) - \mathrm{E}\left(x\right)^2, \tag{B.20}$$

giving

$$\sigma_x^2 = \sum_{i=1}^{m} x_i^2 p_X\left(x_i\right) - \mu_x^2 \tag{B.21}$$

in the discrete case and

$$\sigma_x^2 = \int_{-\infty}^{\infty} x^2 f_X(x) dx - \mu_x^2 \tag{B.22}$$

in the continuous case.

The *confidence level* is the probability, p, given by (B.9) that a random sample from a given distribution lies within the *confidence interval*, defined by the *confidence limits*, x_- and x_+. Where the PDF is symmetric about the mean, the confidence limits are typically expressed as $x_\pm = \mu_x \pm n\sigma_x$, where the relationship between n and p depends on the type of distribution. Where $f_X(x)$ is zero for negative x, the lower confidence limit, x_-, is usually set to zero.

B.2.1 Multiple Variables

The probability that two or more events happen simultaneously is the joint probability. Thus the joint probability that two events, A and B, will both happen is $P(A,B)$. This section focuses on continuous variables. For two continuous variables, x and y, the joint probability density function is $f_{X,Y}(x,y)$, normalized such that

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y) dxdy = 1 \,. \tag{B.23}$$

The joint probability that x takes a value between x_- and x_+ at the same time as y takes a value between y_- and y_+ is then

$$P\left(x_- \le x \le x_+, y_- \le y \le y_+\right) = \int_{y_-}^{y_+}\int_{x_-}^{x_+} f_{X,Y}(x,y) dxdy \,. \tag{B.24}$$

The PDFs describing the distributions of one of the two variables are known as the *marginal* or *unconditional* PDFs. They may be obtained from the joint PDF using

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y) dy, \qquad f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y) dx \,. \tag{B.25}$$

The joint cumulative distribution function is

$$F_{X,Y}(X,Y) = \int_{-\infty}^{Y}\int_{-\infty}^{X} f_{X,Y}(x,y) dxdy \,, \tag{B.26}$$

from which the joint probability may be obtained from

$$P(x_- \le x \le x_+, y_- \le y \le y_+) = F_{X,Y}(x_+, y_+) + F_{X,Y}(x_-, y_-) - F_{X,Y}(x_+, y_-) - F_{X,Y}(x_-, y_+) . \tag{B.27}$$

Conversely,

$$f_{X,Y}(x, y) = \frac{\partial^2 F_{X,Y}(x, y)}{\partial x \partial y} . \tag{B.28}$$

The above is readily extendable to three or more variables. It is also applicable to vectors, such as a position and/or velocity solution. Consider a vector of n continuous variables, $\mathbf{x} = (x_1, x_2, \cdots, x_n)$. The PDF may be written as $f_{\mathbf{X}}(\mathbf{x})$ and is normalized using

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty-\infty}^{\infty\infty} f_{\mathbf{X}}(\mathbf{x}) dx_1 dx_2 \cdots dx_n = 1 , \tag{B.29}$$

which may be abbreviated to

$$\int_{\Re^n} f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} = 1 . \tag{B.30}$$

The probability that **x** lies within an n-dimensional region, D, is given by

$$P(\mathbf{x} \in D) = \int_D f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} . \tag{B.31}$$

Marginal PDFs are obtained by integrating the joint PDF over all of the other components of **x**.

For the case where the variables are independent, the joint PDF or CDF is simply the product of the marginal PDFs or CDFs of the component variables. Thus, for two variables

$$\begin{aligned} f_{X,Y}(x, y) &= f_X(x) f_Y(y) \\ F_{X,Y}(X, Y) &= F_X(X) F_Y(Y) \end{aligned} . \tag{B.32}$$

However, in the general case, variables will be correlated. For example, the north and east position solutions of a navigation system will usually be derived from common sensor measurements, so will exhibit correlated errors.

Bayes rule may be applied to express the joint probability of two events as the marginal probability of one event, multiplied by the *conditional probability* of the other event, given the first event. Thus, for two events, A and B,

$$P(A, B) = P(A \mid B) P(B) = P(B \mid A) P(A) , \tag{B.33}$$

where $P(A \mid B)$ is conditional probability of event A, given event B and $P(B \mid A)$ is conditional probability of event B, given event A. If events A and B are independent, $P(A \mid B) = P(A)$ and $P(B \mid A) = P(B)$.

For two continuous variables, Bayes rule may be written as

$$f_{X,Y}(x, y) = f_{X|Y}(x \mid y) f_Y(y) = f_{Y|X}(y \mid x) f_X(x) , \tag{B.34}$$

where $f_{X|Y}$ and $f_{Y|X}$ are conditional PDFs. Substituting this into (B.25) shows that the conditional PDFs are normalized using

$$\int_{-\infty}^{\infty} f_{X|Y}(x \mid y) dx = 1, \qquad \int_{-\infty}^{\infty} f_{Y|X}(y \mid x) dy = 1 . \tag{B.35}$$

Application of the expectation operator is readily extendable to multiple variables. To determine the mean and variance of x, (B.25) is substituted into (B.16) and (B.22), giving

$$\begin{aligned} \mu_x = \mathrm{E}(x) &= \int_{-\infty}^{\infty} x f_X(x)dx \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f_{X,Y}(x,y)dxdy \\ \sigma_x^2 = \mathrm{E}\left(x^2\right) - \mathrm{E}(x)^2 &= \int_{-\infty}^{\infty} x^2 f_X(x)dx - \mu_x^2 \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^2 f_{X,Y}(x,y)dxdy - \mu_x^2 \end{aligned} . \tag{B.36}$$

Similarly, applying (B.4) enables these to be calculated in terms of $f_{X|Y}$ and f_Y. Where the joint probability is a function of other variables, these must be integrated across, either in the expectation calculation directly or via the calculation of the marginal PDF.

A useful application of the expectation operator in multiple variable systems is the covariance of two variables, defined as

$$\mathrm{cov}(x,y) = \mathrm{E}\left[\left(x-\mu_x\right)\left(y-\mu_y\right)\right] . \tag{B.37}$$

This is an example of a joint central moment and is useful for determining the extent to which two variables are correlated or independent. Where the variables are independent, their covariance is zero. For continuous variables, it is calculated using

$$\begin{aligned} \mathrm{cov}(x,y) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(x-\mu_x\right)\left(y-\mu_y\right) f_{X,Y}(x,y)dxdy \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy f_{X,Y}(x,y)dxdy - \mu_x\mu_y \end{aligned} . \tag{B.38}$$

Again, (B.34) may be applied to calculate the covariance in terms of conditional and marginal PDFs where required. Where the joint probability is a function of further variables, these must also be integrated over.

The correlation coefficient, $\rho_{x,y}$, is defined as

$$\rho_{x,y} = \frac{\mathrm{cov}(x,y)}{\sigma_x\sigma_y} . \tag{B.39}$$

This takes values between −1 and +1, such that $\rho_{x,y}$ is 0 when x and y is independent and ±1 when they are fully correlated or anti-correlated. In the latter case, one variable can be expressed only in terms of constants and the other variable, so there is no need to include both variables in the analysis.

Another application of the expectation is the cross-correlation, a non-central joint moment. This as defined as

$$\begin{aligned} R_{x,y} &= \mathrm{E}(xy) \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy f_{X,Y}(x,y)dxdy \end{aligned} . \tag{B.40}$$

Thus, the cross-correlation and covariance are related by

$$\mathrm{cov}(x,y) = R_{x,y} - \mu_x\mu_y . \tag{B.41}$$

Note that cross-correlation is applicable to deterministic variables, whereas covariance is not.

Returning to the continuous *n*-component vector variable, **x**, its mean is simply the vector of the means of its components, noting that the joint PDF must be integrated over every component of **x** for every component of the mean. Thus,

$$\begin{aligned}\boldsymbol{\mu}_{\mathbf{x}} = \mathrm{E}(\mathbf{x}) &= \int_{-\infty}^{\infty}\cdots\int_{-\infty-\infty}^{\infty\infty}\mathbf{x}f_{\mathbf{X}}(\mathbf{x})dx_1dx_2\cdots dx_n \\ &= \int_{\Re^n}\mathbf{x}f_{\mathbf{X}}(\mathbf{x})d\mathbf{x}\end{aligned}. \tag{B.42}$$

Similarly, the covariance matrix is given by

$$\begin{aligned}\mathbf{C}_{\mathbf{x}} = \mathrm{E}\left[(\mathbf{x}-\boldsymbol{\mu}_{\mathbf{x}})(\mathbf{x}-\boldsymbol{\mu}_{\mathbf{x}})^{\mathrm{T}}\right] &= \int_{\Re^n}(\mathbf{x}-\boldsymbol{\mu}_{\mathbf{x}})(\mathbf{x}-\boldsymbol{\mu}_{\mathbf{x}})^{\mathrm{T}}f_{\mathbf{X}}(\mathbf{x})d\mathbf{x} \\ &= \int_{\Re^n}\mathbf{x}\mathbf{x}^{\mathrm{T}}f_{\mathbf{X}}(\mathbf{x})d\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}}\boldsymbol{\mu}_{\mathbf{x}}{}^{\mathrm{T}}\end{aligned}, \tag{B.43}$$

where the diagonal elements are the variances of the components and the off-diagonal elements the covariances of each pair of components. As covariance matrices are symmetric, it is only necessary to separately calculate the upper or lower triangle of off-diagonal elements. Where the components of **x** are independent, $\mathbf{C}_{\mathbf{x}}$ will be a diagonal matrix.

The correlation matrix is

$$\begin{aligned}\mathbf{R}_{\mathbf{x}} = \mathrm{E}\left[\mathbf{x}\mathbf{x}^{\mathrm{T}}\right] &= \int_{\Re^n}\mathbf{x}\mathbf{x}^{\mathrm{T}}f_{\mathbf{X}}(\mathbf{x})d\mathbf{x} \\ &= \mathbf{C}_{\mathbf{x}} + \boldsymbol{\mu}_{\mathbf{x}}\boldsymbol{\mu}_{\mathbf{x}}{}^{\mathrm{T}}\end{aligned}. \tag{B.44}$$

Finally, where a vector, **x**, is reversibly transformable to another vector, **y**, the PDF may be transformed using

$$f_{\mathbf{Y}}(\mathbf{y}) = \left\|\frac{d\mathbf{x}}{\mathrm{d}\mathbf{y}}\right\| f_{\mathbf{X}}\left[\mathbf{x}(\mathbf{y})\right], \tag{B.45}$$

where ||**A**|| is the magnitude of the determinant of **A**, noting that PDFs must be positive (or zero) and that **x** and **y** must have the same number of components for the transformation to be reversible. Note also that all coordinate transformation matrices, $\mathbf{C}_{\alpha}^{\beta}$ (Section 2.2.2), have unit determinant. Transformations are used in the following sub-sections on summed and differenced variables and the radial error.

B.2.2 Summed and Differenced Variables

Summing of random variables is used in determining the overall effect of a series of errors on a measurement. Consider the sum, *s*, of two random variables, *x* and *y*. Thus,

$$s = x + y\,. \tag{B.46}$$

For the transformation to be reversible a second variable must be introduced. Therefore,

$$t = \tfrac{1}{2}(x - y) \tag{B.47}$$

is also defined. Applying (B.45), the joint PDF of *s* and *t* is

$$f_{S,T}(s,t) = f_{X,Y}\big(x(s,t), y(s,t)\big). \tag{B.48}$$

Substituting this in (B.25) and applying (B.46) and (B.47), the marginal PDF of the sum *s* is then

$$f_S(s) = \int_{-\infty}^{\infty} f_{X,Y}\left(\frac{s}{2}+t, \frac{s}{2}-t\right)dt\,. \tag{B.49}$$

From (B.36), the mean and variance of the summed measurement are

$$\begin{aligned}\mu_s &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} s f_{S,T}(s,t)\,dtds \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} s f_{X,Y}\left(\frac{s}{2}+t,\frac{s}{2}-t\right)dtds \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x+y) f_{X,Y}(x,y)\,dxdy\end{aligned}, \tag{B.50}$$

$$\begin{aligned}\sigma_s^2 &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} s^2 f_{S,T}(s,t)\,dtds - \mu_s^2 \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} s^2 f_{X,Y}\left(\frac{s}{2}+t,\frac{s}{2}-t\right)dtds - \mu_s^2 \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x+y)^2 f_{X,Y}(x,y)\,dxdy - \mu_s^2\end{aligned}. \tag{B.51}$$

Applying (B.36) again and also (B.38) gives

$$\begin{aligned}\mu_s &= \mu_x + \mu_y \\ \sigma_s^2 &= \sigma_x^2 + \sigma_y^2 + 2\,\mathrm{cov}(x,y)\end{aligned}, \tag{B.52}$$

noting that the covariance is zero in cases where x and y are independent. By extension, the mean and variance of the sum of n random variables, x_1 to x_n are

$$\begin{aligned}\mu_s &= \sum_{i=1}^{n} \mu_i \\ \sigma_s^2 &= \sum_{i=1}^{n} \sigma_i^2 + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} \mathrm{cov}(x_i, x_j)\end{aligned}. \tag{B.53}$$

Differencing pairs of ranging measurements involving different transmitters and a common receiver or different receivers and a common transmitter is a feature of many positioning techniques. It enables accuracy to be improved by canceling out common-mode errors between measurements. Therefore, in analyzing the statistical properties of differenced variables, it is important to consider the constituent variables jointly.

Consider a differenced measurement, defined as

$$z = x - y\,. \tag{B.54}$$

Defining also an average,

$$a = \tfrac{1}{2}(x+y) \tag{B.55}$$

and applying (B.45), the joint PDF of z and a is

$$f_{Z,A}(z,a) = f_{X,Y}\big(x(z,a), y(z,a)\big). \tag{B.56}$$

Substituting this in (B.25) and applying (B.54) and (B.55), the marginal PDF of z is then

$$f_Z(z) = \int_{-\infty}^{\infty} f_{X,Y}\left(a+\frac{z}{2}, a-\frac{z}{2}\right)da\,. \tag{B.57}$$

From (B.36), the mean and variance of the differenced measurement are

$$\begin{aligned}\mu_z &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} z f_{Z,A}(z,a)\,da\,dz \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} z f_{X,Y}\left(a+\frac{z}{2}, a-\frac{z}{2}\right)da\,dt\,, \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-y) f_{X,Y}(x,y)\,dx\,dy\end{aligned} \tag{B.58}$$

$$\begin{aligned}\sigma_z^2 &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} z^2 f_{Z,A}(z,a)\,da\,dz - \mu_z^2 \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} z^2 f_{X,Y}\left(a+\frac{z}{2}, a-\frac{z}{2}\right)da\,dz - \mu_z^2\,. \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-y)^2 f_{X,Y}(x,y)\,dx\,dy - \mu_z^2\end{aligned} \tag{B.59}$$

Applying (B.36) again and also (B.38) gives

$$\begin{aligned}\mu_z &= \mu_x - \mu_y \\ \sigma_z^2 &= \sigma_x^2 + \sigma_y^2 - 2\,\mathrm{cov}(x,y)\end{aligned}\,. \tag{B.60}$$

B.2.3 Radial Error

Navigation system performance is often expressed in terms of a 2D or 3D radial error rather than the individual Cartesian components. However, this can sometimes be misleading. To illustrate this, Figures B.1 and B.2 each compare three different types of error ellipse. The first ellipse, a circle, is derived from a radial measure, the second from separate north and east variances, σ_N^2 and σ_E^2, while the third is derived from both variances and their covariance, $\mathrm{cov}(N,E)$. If the error distribution is modeled as bivariate Gaussian (Section B.3.2), the third error ellipse represents a best estimate, while the others represent successive approximations to this.

Figure B.1 shows that where the north and east variances are similar and the correlation coefficient, $\rho_{N,E}$, is small, the error circle derived from the radial measure is a good approximation. Figure B.2 shows the case where the best estimate error ellipse is elongated, so the error circle derived from the radial measure is a very poor approximation (the error ellipse derived only from σ_N^2 and σ_E^2 is also poor here).

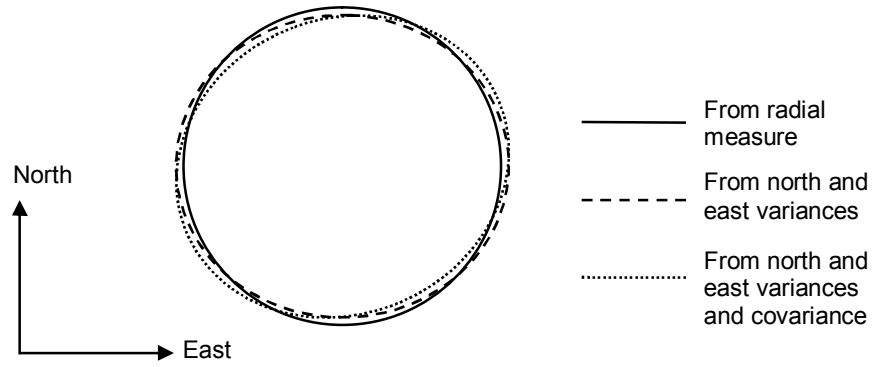

Figure B.1 Comparison of error ellipses where $\sigma_E = 1.1\sigma_N$ and $\rho_{N,E} = 0.1$.

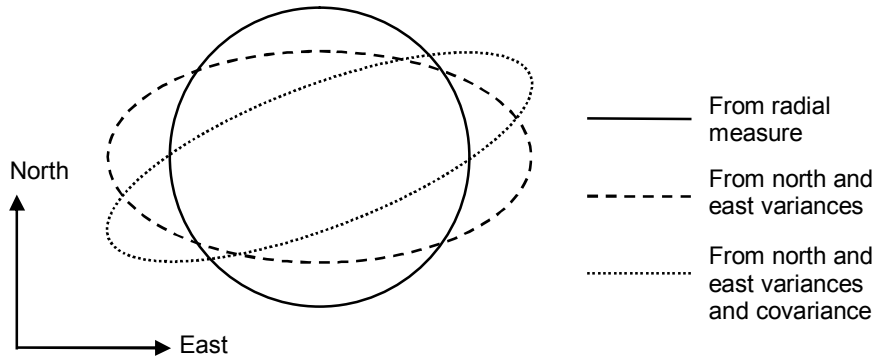

Figure B.2 Comparison of error ellipses where $\sigma_E = 2\sigma_N$ and $\rho_{N,E} = 0.7$.

Elongated error distributions often occur in real navigation systems. Poor signal geometries can occur using radio navigation, depending on location (and also time for GNSS), while dead-reckoning errors are dependent on the host vehicle trajectory and maneuvers. Where position error information is used for collision avoidance, the choice of a poor measure for that position error can compromise safety. Hence, 2D radial errors are not recommended as a description of real-time navigation performance. However, they can be useful to describe the average navigation performance over a range of locations, times, and/or trajectories.

The 3D radial error rarely provides a useful performance measure in navigation. This is because most navigation technologies exhibit quite different performance in the vertical axis than in the horizontal axes. This applies to attitude as much as it does to position and velocity. The discussion below will therefore focus on the 2D radial error.

A radial variable, r, may be obtained from two variables, x and y using

$$r = \sqrt{x^2 + y^2}\,, \tag{B.61}$$

where the positive root is always assumed. For the transformation to be reversible, an azimuth, ψ, is also defined. This is

$$\psi = \arctan_2(y, x), \tag{B.62}$$

noting that this is limited to the range 0 to 2π or $-\pi$ to $+\pi$. Thus, the original variables may be recovered using

$$\begin{aligned} x &= r\cos\psi \\ y &= r\sin\psi \end{aligned}. \tag{B.63}$$

Note that x and y normally represent quantities along orthogonal axes, but they need not be independent.

From (B.45), the joint PDF of r and ψ is given by

$$f_{R,\Psi}(r,\psi) = r f_{X,Y}\big(x(r,\psi), y(r,\psi)\big). \tag{B.64}$$

Substituting this in (B.25) and applying (B.63), the PDF of the radial variable alone is given by

$$f_R(r) = r\int_0^{2\pi} f_{X,Y}\big(r\cos\psi, r\sin\psi\big)d\psi\,. \tag{B.65}$$

Where x and y represent errors with zero-mean distributions, r is the radial error. A number of different statistical measures of that radial error may be computed [7]. From (B.38), the mean radial error, μ_r, is

$$\begin{aligned}\mu_r = \mathrm{E}(r) &= \int_0^\infty r f_R(r)\,dr \\ &= \int_0^\infty \int_0^{2\pi} r^2 f_{X,Y}(r\cos\psi, r\sin\psi)\,d\psi\,dr \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \sqrt{x^2+y^2}\, f_{X,Y}(x,y)\,dx\,dy\end{aligned} \tag{B.66}$$

Note that the standard deviation of the radial error, σ_r, is not a useful measure for quantifying radial error.

The circular error probable (CEP) is the radius, $r_{0.5}$, of a circle enclosing half the error distribution. A circle of radius $r_{0.95}$, enclosing 95% of the error distribution is also commonly used. These are calculated using

$$\int_0^{r_\alpha} f_R(r)\,dr = \alpha\,. \tag{B.67}$$

The 3D equivalent of the CEP is the spherical error probable (SEP).

Another type of measure commonly used is the distance root-mean-squared (DRMS). This is simply the root sum of the variances of two orthogonal axes of the Cartesian position error distribution, for example

$$r_D = \sqrt{\sigma_x^2 + \sigma_y^2}\,. \tag{B.68}$$

Note that the DRMS varies according to which axes are selected; sometimes the major and minor axes of the error ellipse are used. Twice the DRMS, known as 2DRMS, is also commonly used. Section B.3.2 compares radial error measures for the case where $f_{X,Y}$ is a bivariate Gaussian distribution with independent components of equal variance.

B.3 Standard Distributions

Many random (and partially random) variables conform to, or may be approximated to, one of a number of standard distributions whose properties are well known. The simplest of these is the uniform distribution, while the Gaussian (or normal) and chi-square distributions are important in navigation. These are described below. Other continuous distributions include the beta, Cauchy, exponential, gamma, Laplace, Maxwell, Nakagami-m, Rayleigh, and Weibull distributions. Discrete distributions include the Bernoulli, binomial, geometric, negative binomial, Poisson, and trinomial distributions.

B.3.1 Continuous Uniform Distribution

The scalar continuous uniform distribution has the PDF

$$f_U(x) = \begin{cases} 0 & x < a \\ \dfrac{1}{b-a} & a \le x \le b \\ 0 & b < x \end{cases}. \tag{B.69}$$

Its mean and variance are

$$\mu_x = \tfrac{1}{2}(a+b), \qquad \sigma_x^2 = \tfrac{1}{12}(b-a)^2\,. \tag{B.70}$$

A multi-variate uniform distribution always has independent variables. Thus, its PDF is simply the product of those of its constituents and its covariance matrix is diagonal. The joint PDF of two or more correlated variables with uniform distributions is not uniform.

B.3.2 Gaussian Distribution

The *Gaussian* or *Normal* distribution of a continuous scalar, illustrated in Figure B.3, has PDF

$$f_G(x) = \frac{1}{\sigma_x \sqrt{2\pi}} \exp\left[-\frac{(x-\mu_x)^2}{2\sigma_x^2} \right]. \tag{B.71}$$

Samples from this distribution are denoted $N(\mu,\sigma)$. The CDF, F_G, is listed in mathematical tables and available as a function in many software packages. As the Gaussian distribution is symmetric, $F_G[-(x-\mu_x)] = 1 - F_G[(x-\mu_x)]$. Note that any linear combination of Gaussian-distributed variables also has a Gaussian distribution.

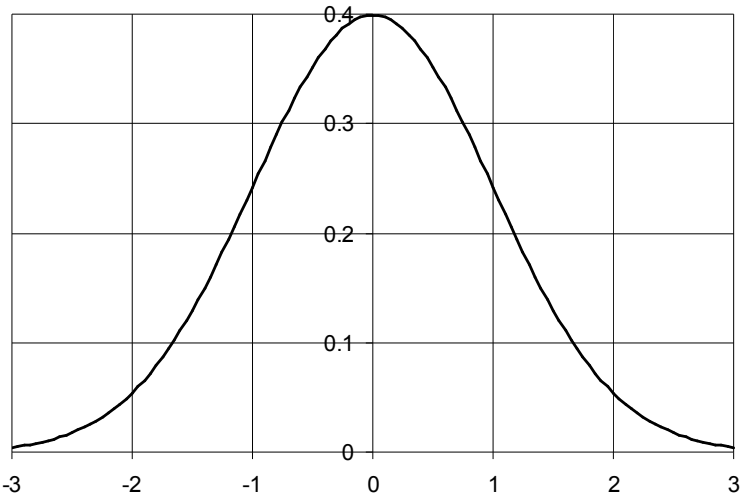

Figure B.3 PDF of a zero-mean unit-variance Gaussian distribution.

The Gaussian distribution is important for a number of reasons. The central limit theorem states that the sum of independent random variables from an arbitrary distribution tends towards Gaussian as the number of samples increases. However, of most relevance here, a Kalman filter models all sources of random error with Gaussian distributions. Table B.1 gives confidence levels and limits for a zero mean Gaussian distribution.

Table B.1 Confidence Levels and Symmetrical Limits for a Zero Mean Gaussian Distribution

Confidence Level	*Symmetrical Confidence Limits* (*Standard Deviations,* σ)
50%	±0.675
68.2%	±1
90%	±1.645
95%	±1.960
95.45%	±2
99%	±2.576
99.73%	±3
99.9%	±3.291
99.99%	±3.891
99.9937%	±4
99.999%	±4.417

The combined PDF of a vector, **x**, comprising *n* continuous variables with Gaussian distributions is that of the *n*-dimensional Gaussian distribution:

$$f_{G,n}(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}|\mathbf{C}_\mathbf{x}|^{1/2}} \exp\left[-\tfrac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_\mathbf{x})^{\mathrm{T}} \mathbf{C}_\mathbf{x}^{-1}(\mathbf{x}-\boldsymbol{\mu}_\mathbf{x})\right], \tag{B.72}$$

where $\boldsymbol{\mu}_\mathbf{x}$ and $\mathbf{C}_\mathbf{x}$ are, the mean and covariance, defined in (B.42) and (B.43), respectively.

The components of any Gaussian-distributed vector, **x**, may be expressed as a linear combination of the components of another vector, **y**, of the same size, such that the components of **y** are mutually independent. The transformation is

$$\begin{aligned} \mathbf{y} &= \mathbf{T}_x^y \mathbf{x} \\ \boldsymbol{\mu}_\mathbf{y} &= \mathbf{T}_x^y \boldsymbol{\mu}_\mathbf{x} \\ \mathbf{C}_\mathbf{y} &= \mathbf{T}_x^y \mathbf{C}_\mathbf{x} \mathbf{T}_x^{y\,\mathrm{T}} \end{aligned}, \tag{B.73}$$

where the transformation matrix, $\mathbf{T}_x^y$, is selected so that $\mathbf{C}_\mathbf{y}$ is diagonal. Matrix diagonalization is described in Section A.6 of Appendix A. The PDF of **y** is then

$$f_{G,n}(\mathbf{y}) = \prod_{i=1}^{n} \left\{ \frac{1}{\sigma_{y,i}\sqrt{2\pi}} \exp\left[-\frac{(y_i - \mu_{y,i})^2}{2\sigma_{y,i}^2} \right] \right\}. \tag{B.74}$$

Consider the radial error (Section B.2.3) of a bivariate Gaussian distribution where the components are independent, zero mean, and have equal variance, $\sigma_{x,y}^2$. From (B.74), the PDF is

$$f_{G,2}(x,y) = \frac{1}{2\pi\sigma_{x,y}^2} \exp\left[-\frac{x^2+y^2}{2\sigma_{x,y}^2} \right]. \tag{B.75}$$

Applying (B.61) to (B.65), the PDF of the radial error is then

$$f_R(r) = \begin{cases} \dfrac{r}{\sigma_{x,y}^2} \exp\left[-\dfrac{r^2}{2\sigma_{x,y}^2} \right] & 0 \le r \\ 0 & r < 0 \end{cases}, \tag{B.76}$$

while, from (B.12), the CDF is

$$F_R(R) = \begin{cases} 1 - \exp\left[-\dfrac{R^2}{2\sigma_{x,y}^2} \right] & 0 \le R \\ 0 & R < 0 \end{cases}. \tag{B.77}$$

From (B.66), the mean radial error is

$$\mu_r = \sqrt{\frac{\pi}{2}}\sigma_{x,y}, \tag{B.78}$$

while, from (B.67), the CEP, $r_{0.5}$, and 95% radial error, $r_{0.95}$, are given by

$$r_\alpha = \sqrt{2\ln\left(\frac{1}{1-\alpha}\right)}\sigma_{x,y}. \tag{B.79}$$

Finally, from (B.68), the DRMS and 2DRMS are simply $\sqrt{2}\sigma_{x,y}$ and $2\sqrt{2}\sigma_{x,y}$, respectively. Applying (B.77), Table B.2 gives the confidence levels (i.e. probability enclosed) and confidence limits for each measure [2].

Table B.2 Confidence Levels and Radial Limits for a Bivariate Gaussian Distribution with Independent Equal-Variance Components

Measure	*Confidence Level*	*Symmetrical Confidence Limits (Single-Axis SDs, $\sigma_{x,y}$)*
Single axis SD, $\sigma_{x,y}$	39.3%	1
CEP, $r_{0.5}$	50%	1.177
Mean radial error, μ_r	54.4%	1.253
DRMS, r_D	63.2%	1.414
95% error radius, $r_{0.95}$	95%	2.448
2DRMS, $2r_D$	98.2%	2.828

B.3.3 Chi-square Distribution

The chi-square statistic can be thought of as a measure of the degree to which a set of observations fit a Gaussian distribution. It is defined as

$$\chi_n^2 = (\mathbf{x} - \boldsymbol{\mu}_\mathbf{x})^{\mathrm{T}} \mathbf{C}_\mathbf{x}^{-1} (\mathbf{x} - \boldsymbol{\mu}_\mathbf{x}). \tag{B.80}$$

where $\mathbf{x}$ is the set of observations, $\boldsymbol{\mu}_\mathbf{x}$ is their mean, and $\mathbf{C}_\mathbf{x}$ is their covariance. Where the observations, x_i, are independent, this simplifies to

$$\chi_n^2 = \sum_{i=1}^{n} \frac{(x_i - \mu_{x,i})^2}{\sigma_{x,i}^2}. \tag{B.81}$$

The number of components of $\mathbf{x}$, n, is the number of degrees of freedom of the chi-square distribution. Its PDF is

$$f_{\chi^2}(\chi_n^2, n) = \frac{(1/2)^{n/2}}{\Gamma(n/2)} (\chi_n^2)^{(n/2-1)} \exp(-\tfrac{1}{2}\chi_n^2), \tag{B.82}$$

where the gamma function, Γ, is

$$\Gamma(n/2) = \int_0^\infty \alpha^{n/2-1} e^{-\alpha} d\alpha. \tag{B.83}$$

Figure B.4 shows the PDF and Table B.3 gives the confidence limits for chi-square distributions with 1 to 4 degrees of freedom.

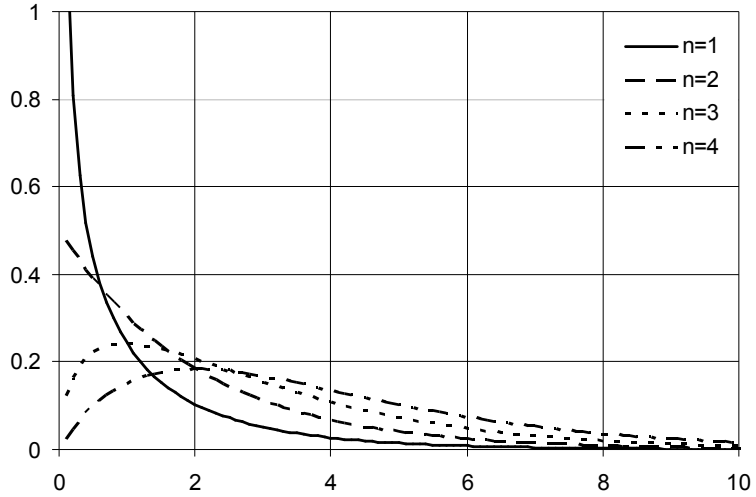

Figure B.4 PDF of chi-square distributions with 1 to 4 degrees of freedom.

Table B.3 Upper Confidence Limits (Lower Confidence Limit is 0) for Chi-Square Distributions with *n* Degrees of Freedom

	Upper Confidence Limit			
Confidence Level	$n = 1$	$n = 2$	$n = 3$	$n = 4$
90%	2.71	4.61	6.25	7.78
95%	3.84	5.99	7.81	9.49
99%	6.64	9.21	11.34	13.28
99.9%	10.83	13.75	16.27	18.47

B.3.4 Mixture Distributions

Consider the case where the distribution of a random variable depends on another event. For example, a position error may have one distribution when a fault has occurred and another distribution otherwise. Similarly, a pedestrian step length distribution will depend on whether the subject is walking, running, or jogging. In this case, the total distribution of the random variable may be expressed as the sum of each conditional distribution, multiplied by the probability of the condition. This is known as a mixture distribution and is different from the summed variable distribution described in Section B.2.2.

Consider two mutually exclusive events, A and B (i.e. $P(A)+P(B)=1$), If x is a continuous random variable with a PDF that depends on the event, the overall PDF may be expressed as the following mixture

$$f_X(x) = f_{X|A}(x)P(A) + f_{X|B}(x)P(B), \tag{B.84}$$

where $f_{X|A}$ and $f_{X|B}$ are the conditional PDFs. If these describe Gaussian distributions then the overall distribution of x is a Gaussian mixture. However, the different conditional PDFs don't have to describe the same type of distribution.

More generally, if the PDF of x depends on a discrete random variable, y, it may be expressed as

$$f_X(x) = \sum_i f_{X|i}(x)p_Y(y_i), \tag{B.85}$$

where $f_{X|i}(x)$ is the PDF of x under the condition that $y = y_i$. The mean and variance of the mixture are

$$\mu_x = \sum_i p_Y(y_i)\mu_{x|i}, \tag{B.86}$$

$$\sigma_x^2 = \sum_i p_Y(y_i)\left(\sigma_{x|i}^2 + \mu_{x|i}^2\right) - \mu_x^2, \tag{B.87}$$

where $\mu_{x|i}$ and $\sigma_{x|i}^2$ are, respectively, the mean and variance of the conditional PDF $f_{X|i}$.

B.4 Random Processes

A *random process*, also known as a *stochastic process* or random signal, is a random variable that changes continuously with time, t. A *random sequence* is a random variable that changes at discrete time intervals. Note that a random process is often approximated to a random sequence in discrete-time processing.

The temporal properties of a random process are described by comparing the values of that process at different times. Suppose z is a continuous random process, taken from a population with distribution Z. Its PDF is $f_Z(z,t)$. Two continuous random variables may then be defined as

$$x = z(t), \qquad y = z(t+\tau). \tag{B.88}$$

Their PDFs are then

$$f_X(x) = f_Z(x,t), \qquad f_Y(y) = f_Z(y,t+\tau). \tag{B.89}$$

A joint PDF, $f_{X,Y}$, may then be formed as described in Section B.2.1.

From (B.37), the autocovariance of z is defined as

$$\begin{aligned} \operatorname{cov}\big(z(t),z(t+\tau)\big) &= \operatorname{cov}(x,y) \\ &= \mathrm{E}\big[(x-\mu_x)(y-\mu_y)\big] \\ &= \mathrm{E}\big[(z(t)-\mu_z(t))(z(t+\tau)-\mu_z(t+\tau)\big] \end{aligned}, \tag{B.90}$$

while, from (B.40), the autocorrelation is

$$\begin{aligned} R_z(t,t+\tau) &= R_{x,y} \\ &= \mathrm{E}(xy) \\ &= \mathrm{E}\big[z(t)z(t+\tau)\big] \end{aligned}. \tag{B.91}$$

The power spectral density (PSD) is the Fourier transform of the autocorrelation. In terms of angular frequency, ω, the PSD is

$$S_z^{\omega}(\omega,t) = \int_{-\infty}^{\infty} R_z(t,t+\tau)e^{-j\omega\tau}d\tau, \tag{B.92}$$

where $j = \sqrt{-1}$. Conversely,

$$R_z(t,t+\tau) = \frac{1}{2\pi}\int_{-\infty}^{\infty} S_z^{\omega}(\omega,t)e^{j\omega\tau}d\omega. \tag{B.93}$$

PSD is an even function of frequency, so $S_z^{\omega}(-\omega,t) = S_z^{\omega}(\omega,t)$.

If a random signal has amplitude, a, the power of that signal is given by

$$P(t) = R_a(t,t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} S_a^{\omega}(\omega,t)d\omega = \frac{1}{2\pi}\int_{-\infty}^{\infty} S_a^{f}(f,t)df, \tag{B.94}$$

where $\omega = 2\pi f$ and $S_a^f(f,t) = 2\pi S_a^{\omega}(\omega,t)$, noting that superscripts f and ω are used to denote frequency and angular frequency PSD, respectively. Thus, the power is obtained by integrating the PSD of the amplitude over the frequency spectrum.

A random process or sequence is *strict-sense stationary* (SSS) where its PDF does not vary with time. It is *wide-sense stationary* (WSS) where its mean is time-independent and its autocorrelation and PSD depend only on the time interval, τ, not the time, t. An SSS process is always WSS, but not vice versa. For a stationary process, z, the autocorrelation may be written simply as $R_z(\tau)$. It has the properties $R_z(\tau) \le R_z(0)$ and $R_z(-\tau) = R_z(\tau)$. This ensures that the PSD is always real.

An *ergodic* process has a time average that is the same as the ensemble mean, i.e.

$$\lim_{\tau\to\infty}\frac{1}{\tau}\int_{t}^{t+\tau} z(t')dt' = \mathrm{E}\big(z(t)\big) = \int_{-\infty}^{\infty} z(t)f_Z(z,t)dz. \tag{B.95}$$

An ergodic process must be WSS, but not vice versa.

B.4.1 Multiple Random Processes

The correlation of two separate random processes, x and y, may be described by the cross-correlation, defined as

$$R_{x,y}(t,t+\tau) = \mathrm{E}\big[x(t)y(t+\tau)\big]. \tag{B.96}$$

Similarly, the cross spectral density (CSD) is

$$S_{x,y}^{\omega}(\omega,t)=\int_{-\infty}^{\infty}R_{x,y}(t,t+\tau)e^{-j\omega\tau}d\tau \tag{B.97}$$

and

$$R_{x,y}(t,t+\tau)=\frac{1}{2\pi}\int_{-\infty}^{\infty}S_{x,y}^{\omega}(\omega,t)e^{j\omega\tau}d\omega\,. \tag{B.98}$$

Where both processes are WSS, the notational convention

$$R_{x,y}(\tau)=\mathrm{E}\left[x(t)y(t+\tau)\right],\qquad R_{y,x}(\tau)=\mathrm{E}\left[x(t+\tau)y(t)\right] \tag{B.99}$$

is adopted, noting that $R_{x,y}(-\tau)=R_{y,x}(\tau)$. Similarly, $S_{x,y}^{\omega}(-\omega)=S_{x,y}^{\omega}(\omega)$.

For a vector of random processes, **x**, the correlation matrix is

$$\mathbf{R}_{\mathbf{x}}(t,t+\tau)=\mathrm{E}\left[\mathbf{x}(t)\mathbf{x}^{\mathrm{T}}(t+\tau)\right], \tag{B.100}$$

where the diagonal elements comprise the autocorrelations of each component of **x** and the off-diagonal elements comprise the cross-correlations of each pair of components of **x**.

The PSD matrix is

$$\mathbf{S}_{\mathbf{x}}^{\omega}(\omega,t)=\int_{-\infty}^{\infty}\mathbf{R}_{\mathbf{x}}(t,t+\tau)e^{-j\omega\tau}d\tau\,. \tag{B.101}$$

Similarly, the diagonal elements are the PSDs of each component of **x** and the off-diagonal elements are the CSDs.

B.4.2 White Noise

A continuous *white noise process* is a random process that has a time-constant and frequency-independent power spectral density. The term originates from white light, which occupies the whole visible spectrum. The Fourier transform of a constant is a Dirac delta function, $\delta(\tau)$, which is infinite when its argument is zero and zero otherwise. Note that the units of a Dirac delta function are the inverse of those of its argument (a Kronecker delta function is unitless). The autocorrelation function of a white noise process, *w*, may be expressed as

$$R_w(t_1,t_2)=\mathrm{E}\left[w(t_1)w(t_2)\right]=S_w^{f1}\delta\left(t_1-t_2\right), \tag{B.102}$$

where $S_w^{f1}=2S_w^{f}$ is the single-sided PSD. As R_w is zero for $t_1\neq t_2$, the white noise process must be zero mean, so the autocorrelation is the same as the autocovariance. Thus, for $t_1=t_2$, the autocorrelation is the variance.

$$R_w(t_1,t_2)=\begin{matrix}\sigma_w^2 & t_1=t_2\\ 0 & t_1\neq t_2\end{matrix}\,. \tag{B.103}$$

A true continuous white noise process is strictly a theoretical concept. A constant PSD across all frequencies requires an infinite bandwidth. From (B.94), an infinite power is thus required to transmit a white noise signal while, from (B.93), the variance would be infinite. Therefore, continuous white noise processes never arise in the real world. Noise processes that are not white are known as colored noise.

A bandlimited white noise process has a PSD which is constant over a certain frequency range and drops down to zero outside this range. An example is the thermal noise interference on a radio signal. At frequencies within the constant PSD band, the noise process may be treated as white, while at frequencies outside this band, it must be treated as colored. It is common for a noise process to be white at low frequencies and colored at high frequencies,

particularly after conversion to baseband. In this case, the process may be treated as white when considering timescales greatly in excess of the process's correlation time.

The following PSD and autocorrelation are an example of a bandlimited white noise process, x:

$$S_x^{\omega}(\omega) = \begin{array}{ll} S_0^{\omega} & |\omega| \le \pi B \\ 0 & |\omega| > \pi B \end{array}, \tag{B.104}$$

$$R_x(\tau) = BS_0^{\omega} \frac{\sin(\pi B\tau)}{\pi B\tau} = BS_0^{\omega}\mathrm{sinc}(\pi B\tau), \tag{B.105}$$

where B is the double-sided noise bandwidth. In a practical system, the white region of the PSD will be less clearly defined due to filtering limitations.

A *white noise sequence* is a discrete-time sequence of mutually-uncorrelated random variables from a zero mean distribution. Samples have the property

$$\mathrm{cov}(w_i, w_j) = \mathrm{E}(w_i w_j) = \begin{array}{ll} \sigma_w^2 & i = j \\ 0 & i \ne j \end{array}. \tag{B.106}$$

Samples from a bandlimited white noise process may be treated as a white noise sequence, provided the sampling rate is much less than the double-sided noise bandwidth. This also applies to the time average or the integral, provided the time interval is much greater than the reciprocal of the noise bandwidth, B.

Consider the sum, s, of n samples, w_1 to w_n, from a white noise sequence of constant variance, σ_w^2. From (B.53) and (B.106), the variance of s is

$$\sigma_s^2 = n\sigma_w^2. \tag{B.107}$$

The sum of a white noise sequence is known as a *random walk* sequence. A random walk has zero mean, but its variance increases linearly with the number of steps (where the noise variance is constant). The terms random noise and random walk are sometimes used interchangeably because the latter is the integral of the former. For example, acceleration random noise and velocity random walk are the same phenomenon. Where the white noise samples have a Gaussian distribution, the random walk sequence may be described as a *Wiener* or *Brownian-motion* sequence.

Considering now the average, a, of the same n samples, w_1 to w_n. This has variance

$$\sigma_a^2 = \frac{\sigma_s^2}{n^2} = \frac{\sigma_w^2}{n}. \tag{B.108}$$

Sums and averages of white noise sequences will themselves form white noise sequences if each is the sum or average of an independent set of samples from a white noise sequence. Otherwise, they will be correlated.

Consider two sequences of samples, s and z, obtained by integrating the same bandlimited white noise process, w, over different sampling intervals, respectively, τ_s and τ_z, where $\tau_s, \tau_z >> 1/B$. Thus,

$$s_i = \int_{(i-1)\tau_s}^{i\tau_s} w(t)dt, \qquad z_i = \int_{(i-1)\tau_z}^{i\tau_z} w(t)dt. \tag{B.109}$$

From (B.107), the ratio of the variances may then be expressed as

$$\frac{\sigma_s^2}{\sigma_z^2} = \frac{\tau_s}{\tau_z}. \tag{B.110}$$

Similarly, consider two further sequences of samples, a and b, obtained by averaging w over intervals, τ_a and τ_b, respectively, where, again, $\tau_a, \tau_b >> 1/B$. Thus,

$$a_i = \frac{1}{\tau_a} \int_{(i-1)\tau_a}^{i\tau_a} w(t)dt, \qquad b_i = \frac{1}{\tau_b} \int_{(i-1)\tau_b}^{i\tau_b} w(t)dt. \tag{B.111}$$

From (B.108),

$$\frac{\sigma_a^2}{\sigma_b^2} = \frac{\tau_b}{\tau_a}. \tag{B.112}$$

From (B.109) and (B.111), it can be seen that where unit time intervals are used, integrating or averaging gives the same result. The variance of these samples can thus be expressed as a single-sided PSD, S_w^{f1}, multiplied or divided by unit time, as appropriate. Therefore, from (B.110) and (B.112) the variances of white noise samples integrated or averaged over a general time interval may be expressed in terms of this PSD using

$$\sigma_s^2 = S_w^{f1}\tau_s, \tag{B.113}$$

$$\sigma_a^2 = \frac{S_w^{f1}}{\tau_a}, \tag{B.114}$$

provided the integration or averaging interval, τ_s or τ_a, significantly exceeds the reciprocal of the white noise bandwidth. As sampling intervals can vary according to processing, manufacturers often specify noise sources using the PSD, S_w^{f1}, or its root, rather than the variance or SD. Where a bandlimited white noise process has a Gaussian distribution, its integral is known as a Wiener process.

It is also useful in navigation to consider the double and triple integrals of a bandlimited white noise process:

$$D = \int_0^{\tau_D}\int_0^{t_1} w(t_2)dt_2dt_1, \qquad T = \int_0^{\tau_T}\int_0^{t_1}\int_0^{t_2} w(t_3)dt_3dt_2dt_1. \tag{B.115}$$

These have variances

$$\sigma_D^2 = \tfrac{1}{3}\tau_D^3 S_w^{f1}, \qquad \sigma_T^2 = \tfrac{1}{5}\tau_T^5 S_w^{f1}, \tag{B.116}$$

where, again, $\tau_D, \tau_T >> 1/B$.

B.4.3 Markov Processes

A first-order *Markov process* is a continuous-time random process, $x(t)$, the future value of which depends only on its present value and a white noise process, $w(t)$; it does not depend on previous values of x. It may be described by

$$\dot{x}(t) = f\big(x(t),t\big) + g\big(w(t),t\big), \tag{B.117}$$

where f and g are arbitrary functions (f is not the PDF, here). Note that if the Markov process is stationary, f and g are functions of, respectively, x and w, only. This is readily extendible to a vector process, $\mathbf{x}(t)$:

$$\dot{\mathbf{x}}(t) = \mathbf{f}\big(\mathbf{x}(t),t\big) + \mathbf{g}\big(\mathbf{w}(t),t\big), \tag{B.118}$$

noting that a component of $\dot{\mathbf{x}}$ may be a function of any components of $\mathbf{x}$ and $\mathbf{w}$.

Similarly, a first-order *Markov sequence* is a discrete-time sequence, x_k, which depends only on its previous value, x_{k-1}, and a white noise sequence, w_{k-1}; it does not depend on earlier values of x. It may be described by

$$x_k = \phi_{k-1}(x_{k-1}) + \gamma_{k-1}(w_{k-1}), \tag{B.119}$$

where ϕ and γ are arbitrary functions. If the Markov sequence is stationary, ϕ and γ are functions of, respectively, x and w, only. This is also extendable to a vector sequence, $\mathbf{x}_k$:

$$\mathbf{x}_k = \boldsymbol{\varphi}_{k-1}(\mathbf{x}_{k-1}) + \boldsymbol{\gamma}_{k-1}(\mathbf{w}_{k-1}). \tag{B.120}$$

Higher-order versions of the Markov process and sequence also exist. An n^{th}-order Markov process is described by a differential equation to n^{th} order in time. An n^{th}-order Markov sequence, x_k is a function of x_{k-n} to x_{k-2} as well as x_{k-1}.

A *Gauss-Markov* process imposes the additional constraint that the differential equation describing it is linear. A first-order Gauss-Markov process is described by

$$\dot{x}(t) = f(t)x(t) + g(t)w(t) \tag{B.121}$$

in the scalar case and

$$\dot{\mathbf{x}}(t) = \mathbf{F}(t)\mathbf{x}(t) + \mathbf{G}(t)\mathbf{w}(t) \tag{B.122}$$

in the vector case, where the scalars f and g and matrices $\mathbf{F}$ and $\mathbf{G}$ are not functions of x or w. The Gauss-Markov process is so-named because if the white noise process and the initial value of the Markov process have Gaussian distributions then the Markov process at later times will also have a Gaussian distribution, enabling its PDF to be represented with only two parameters, its mean and variance. Note that, where $f(t) = 0$, (B.121) describes a random walk process.

A Gauss-Markov process is sometimes known as an exponentially-correlated Markov process because, where $f = -\alpha$ is a negative constant, (B.121) has the solution

$$x(t+\tau) = \exp(-\alpha|\tau|)x(t) + \int_t^{t+\tau} g(t')w(t')dt' \tag{B.123}$$

and the autocorrelation is

$$R_x(t, t+\tau) = \exp(-\alpha|\tau|)\sigma_x^2(t). \tag{B.124}$$

Similarly, in a Gauss-Markov sequence, x_k is a linear function of its previous values and the white noise sequence. A first-order Gauss-Markov sequence is described by

$$x_k = \phi_{k-1}x_{k-1} + \gamma_{k-1}w_{k-1} \tag{B.125}$$

in the scalar case and

$$\mathbf{x}_k = \boldsymbol{\Phi}_{k-1}\mathbf{x}_{k-1} + \boldsymbol{\Gamma}_{k-1}\mathbf{w}_{k-1} \tag{B.126}$$

in the vector case, where the scalars ϕ_{k-1} and γ_{k-1} and the matrices $\boldsymbol{\Phi}_{k-1}$ and $\boldsymbol{\Gamma}_{k-1}$ are not functions of x_{k-1} and w_{k-1}. In the random walk case, $\phi_{k-1} = 1$.

B.5 Hypothesis Testing

In many navigation applications, there is a need to use measurement data, which has a time-evolving nature and is subject to noise and bias, to answer a Boolean question. Examples are:

- Has a fault occurred?
- Has a user entered or left a particular region of space?
- Is a pedestrian walking or running?

The answers to these questions are hypotheses.

One way of determining Boolean (or discrete) information from continuous data is to perform a hypothesis test. A test statistic, s, is computed from the measurement data and compared with a threshold, T. Depending on the outcome, one of two hypotheses, denoted H_0 and H_1, is assumed. H_0 is known as the null hypothesis and represents the more likely outcome of a test, for example the no-fault hypothesis.

Two errors can occur in a hypothesis test. A type 1 error is that the test outcome is H_1 when H_0 is true. The probability of this, α, is called the *significance level*. Thus,

$$\begin{aligned} P(\text{Outcome} = H_1 \mid \text{Truth} = H_0) &= \alpha \\ P(\text{Outcome} = H_0 \mid \text{Truth} = H_0) &= 1-\alpha \end{aligned}, \tag{B.127}$$

where H_1 denotes a fault hypothesis and H_0 the no-fault hypothesis, α is the probability of false alarm, p_{fa}. A type 2 error is that the test outcome is H_0 when H_1 is true. The probability of this, β, is called the *operating characteristic*, while $1-\beta$ is called the *power of the test*. Thus,

$$\begin{aligned} P(\text{Outcome} = H_0 \mid \text{Truth} = H_1) &= \beta \\ P(\text{Outcome} = H_1 \mid \text{Truth} = H_1) &= 1-\beta \end{aligned}, \tag{B.128}$$

where H_1 denotes a fault hypothesis and H_0 the no-fault hypothesis, β is the probability of missed detection, p_{md}, and $1-\beta$ is the detection probability, p_d.

Both α and β are functions of the threshold, T. Where the PDFs of the test statistic for each hypothesis, $f_{S|H0}$ and $f_{S|H1}$, are known, they may be used to compute α and β. For example, for the hypothesis test

$$\text{Outcome} = \begin{matrix} H_0 & |s| \le T \\ H_1 & |s| > T \end{matrix}, \tag{B.129}$$

the significance level is

$$\begin{aligned} \alpha &= \int_{-\infty}^{-T} f_{S|H0}(s \mid H_0)ds + \int_{T}^{\infty} f_{S|H0}(s \mid H_0)ds \\ &= 1 - \int_{-T}^{T} f_{S|H0}(s \mid H_0)ds \end{aligned} \tag{B.130}$$

and the operating characteristic is

$$\begin{aligned} \beta &= \int_{-T}^{T} f_{S|H1}(s \mid H_1)ds \\ &= 1 - \int_{-\infty}^{-T} f_{S|H1}(s \mid H_1)ds - \int_{T}^{\infty} f_{S|H1}(s \mid H_1)ds \end{aligned}. \tag{B.131}$$

Figure B.5 illustrates this. Note that a change in threshold that reduces α will increase β and vice versa.

The test threshold, T, may be set to give a particular value of α, in which case, β becomes a function of T and $f_{S|H1}$. Alternatively, it may be set to give a particular value of β, in which case, α becomes a function of T and $f_{S|H0}$. Another option is to set T to the value that minimizes $\alpha+\beta$ or $\alpha\beta$. Whether it is more important to minimize α or β depends on the application. For example, for a safety-critical fault detection test, it is more important to minimize the probability of missed detection than the probability of false alarm.

For tests where the PDFs vary, the threshold may be varied to maintain a constant α or β or meet some other criterion. Sometimes with a variable PDF hypothesis test, it may not be possible to set a threshold which gives acceptable values of both α and β, in which case the test should be declared unavailable.

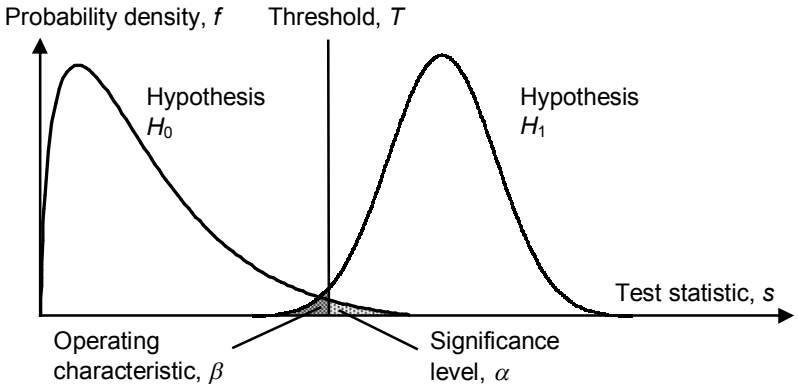

Figure B.5 Significance level and operating characteristic of a hypothesis test (in this case, s is always positive).

For applications without a strict time limit, α and β may generally be reduced by collecting more measurement data from which to calculate the test statistic. A number of approaches to variable-time hypothesis testing may be taken. In one approach, two thresholds are set, one defined by α and the other by β. If the test statistic falls between the two thresholds, neither hypothesis is assumed and more data is collected. As the amount of data used to calculate the test statistic increases, its PDFs will sharpen, enabling the thresholds to be brought closer together until one or other hypothesis may be assumed with the confidence levels defined by α and β.

Alternatively, Bayes theorem, (B.33), may be used to determine the probability of each hypothesis given the data:

$$\begin{aligned} P(H_0 \mid s) &= \frac{f_{S|H0}(s \mid H_0)P(H_0)}{f_{S|H0}(s \mid H_0)P(H_0) + f_{S|H1}(s \mid H_1)P(H_1)} \\ P(H_1 \mid s) &= \frac{f_{S|H1}(s \mid H_1)P(H_1)}{f_{S|H0}(s \mid H_0)P(H_0) + f_{S|H1}(s \mid H_1)P(H_1)} \end{aligned} . \tag{B.132}$$

Confidence levels, T_0 and T_1, are set such that H_0 is assumed if $P(H_0 \mid s) \geq T_0$, H_1 is assumed if $P(H_1 \mid s) \geq T_1$, and more data collected otherwise ($T_0 + T_1 \leq 1$).

All forms of hypothesis testing are easily extendable to multiple hypotheses and vector test statistics. The solution space of the test statistic, $\mathbf{s}$, is divided into non-overlapping regions, one for each hypothesis, H_0 to H_n, and, optionally, regions indicating that more data is required. Each non-null hypothesis, H_i, has its own operating characteristic, β_i. For a vector test statistic, significance levels, operating characteristics, and probabilities must be calculated using the joint PDF of $\mathbf{s}$.

References

[1] Brown, R. G. and P. Y. C. Hwang , *Introduction to Random Signals and Applied Kalman Filtering*, 3rd ed., New York: Wiley, 1997.

[2] Farrell, J. A., *Aided Navigation: GPS with High Rate Sensors*, New York: McGraw Hill, 2008.

[3] Gelb, A. (Ed.), *Applied Optimal Estimation*, Cambridge, MA: MIT Press, 1974.

[4] Grimmett, G. R. and D. R. Stirzaker, *Probability and Random Processes*, Second Edition, Oxford, UK: Oxford University Press, 1992.

[5] Papoulis, A. and S. U. Pillai, *Probability, Random Variables and Stochastic Processes*, 4thd ed., New York: McGraw Hill, 2002.

[6] Simon, D., *Optimal State Estimation*, New York: Wiley, 2006.

[7] Leenhouts, P. P., "On the Computation of Bi-Normal Radial Error," *Navigation: JION*, Vol. 32, No. 1, 1985, pp. 16-28.