Chapter 1
Random Functions

The basic filtering problem in signal processing is to operate on an observed signal to estimate a desired signal. The immediate difficulty is obvious: how to construct the filter when all one has is the observed signal. One might try some naïve approach like using a standard low-pass filter, but why should that produce a good result? The proper formulation of the problem, as laid down by Norbert Wiener in the 1930s, is to treat the observed and desired signals as random functions that are jointly probabilistically related, in which case one can find a filter that produces the best estimate of the desired random signal based on the observed random signal, where optimality is relative to some probabilistic error criterion. When an actual signal is observed, the optimal filter is applied. It makes no sense to enquire about the accuracy of the filter relative to any single observation, since if we knew the desired signal that led to the observation, a filter would not be needed. It would be like processing data in the absence of a criterion beyond the data itself and asking if the processing is beneficial. This would be a form of pre-scientific radical empiricism. As put by Hans Reichenbach (Reichenbach, 1971), "If knowledge is to reveal objective relations of physical objects, it must include reliable predictions. A radical empiricism, therefore, denies the possibility of knowledge." Since knowledge is our goal and optimal operator design is our subject, we begin by defining a random function and considering the basic properties of such functions, including the calculus of random functions.

A *random function*, or *random process*, is a family of random variables $\{X(\omega; t)\}$, t lying in some index set T, where, for each fixed t, the random variable $X(\omega; t)$ is defined on a sample space S ($\omega \in S$). For a fixed ω, $X(\omega; t)$ defines a function on the set T, and each of such functions is termed a *realization* of the random function. We focus on real-valued functions. If T is a subset of the real line $\Re$, then, for fixed ω, $X(\omega; t)$ is a signal, and the random function $\{X(\omega; t)\}$ is called a *random signal, stochastic process, or random time function.* Should a random process be defined only on the integers, it is sometimes called a *random time series.* In general, t can be a point in n-dimensional Euclidean space $\Re^n$, so that each realization is a

deterministic function of n variables. To simplify notation we usually write $X(t)$ to denote a random function, keeping in mind the underlying probability structure. In particular, if we fix t and let ω vary, then $X(t)$ is a random variable on the sample space. A specific realization will often be denoted by $x(t)$.

1.1 Moments

For each t, $X(t)$ is a random variable and has a probability distribution function $F(x; t) = P(X(t) \leq x)$, called a *first-order distribution*. For the random functions that concern us, $X(t)$ will possess a *first-order density* $f(x; t) = \frac{dF(x; t)}{dx}$, where the derivative might involve delta functions.

In practice it is common to index a random function by a random variable instead of elements in a sample space. Instead of considering the realizations to be dependent on observations coming from a sample space, it is more practical to suppose them to be chosen according to observations of a random variable. Since a random variable Z defined on a probability space induces a probability measure on the Borel field over the real line $\Re$, with the induced probability measure P_Z defined in terms of the original probability measure P by $P_Z(B) = P(Z \in B)$ for any event B, nothing is lost by indexing a random function by the values of a random variable.

For fixed t, the first-order distribution completely describes the behavior of the random variable; however, in general, we require the *nth-order probability distributions*

$$
\begin{aligned}
&F(x_1, x_2, \ldots, x_n; t_1, t_2, \ldots, t_n) \\
&\quad = P(X(t_1) \leq x_1, X(t_2) \leq x_2, \ldots, X(t_n) \leq x_n)
\end{aligned}
\tag{1.1}
$$

and the corresponding *nth-order densities* $f(x_1, x_2, \ldots, x_n; t_1, t_2, \ldots, t_n)$.

It is possible by integration to obtain the marginal densities from a given joint density. Hence, each nth-order density specifies the marginals for all subsets of $\{t_1, t_2, \ldots, t_n\}$. Now, suppose that we wish to give a complete characterization of the random function in terms of the various joint densities. If the point set is infinite, it is not generally possible to completely characterize the random function by knowing all finite joint densities; however, if the realizations are sufficiently well behaved, knowledge of the densities of all finite orders completely characterizes the random function — and we will always make this assumption. For practical manipulations, it is useful to be able to characterize a random function by means of the joint densities of some finite order.

If for each point set $\{t_1, t_2, \ldots, t_n\}$ the random variables $X(t_1)$, $X(t_2)$, ..., $X(t_n)$ are independent, then the random function is characterized by its first-order densities since

$$
f(x_1, x_2, \ldots, x_n; t_1, t_2, \ldots, t_n) = f(x_1; t_1)f(x_2; t_2) \cdots f(x_n; t_n).
\tag{1.2}
$$

An important class of random functions characterized by second-order distributions is the class of Gaussian random functions. A random function $X(t)$ is said to be *Gaussian*, or *normal*, if for any collection of n points $t_1, t_2, \ldots, t_n$, the random variables $X(t_1), X(t_2), \ldots, X(t_n)$ possess a multivariate Gaussian distribution. A multivariate Gaussian distribution is completely characterized by its mean vector and covariance matrix, and these are in turn determined from the first- and second-order densities of the variables.

A great deal of linear systems theory employs only second-order moment information. While mathematical tractability is gained, the loss is that there is no distinction between random functions possessing identical second-order moments. Linear filters have a natural dependency on second-order information — and therefore lack discrimination relative to random processes differing only at higher orders.

The *expectation* (*mean function*) of a random function $X(t)$ is the first-order moment

$$\mu_X(t) = E[X(t)] = \int_{-\infty}^{\infty} x f(x; t) dx. \tag{1.3}$$

Another function depending on $X(t)$ in isolation is the *variance function*:

$$\text{Var}[X(t)] = E[(X(t) - \mu_X(t))^2] = \int_{\infty}^{\infty} (x - \mu_X(t))^2 f(x; t) dx. \tag{1.4}$$

For fixed t, $\text{Var}[X(t)]$ is the variance of the random variable $X(t)$. The *standard deviation function* is defined by $\sigma_X(t) = \text{Var}[X(t)]^{\frac{1}{2}}$.

The second-order *covariance function* of the random function $X(t)$ is defined by

$$\begin{aligned}
K_X(t, t') &= E[(X(t) - \mu_X(t))(X(t') - \mu_X(t'))] \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_X(t))(x' - \mu_X(t')) f(x, x'; t, t') dx dx'.
\end{aligned} \tag{1.5}$$

Letting $t = t'$ in the covariance function yields the *variance function*, $K_X(t, t) = \text{Var}[X(t)]$. It is seen directly from its definition that the covariance function is symmetric: $K_X(t, t') = K_X(t', t)$. The *correlation-coefficient function* is defined by

$$\rho_X(t, t') = \frac{K_X(t, t')}{\sigma_X(t)\sigma_X(t')}. \tag{1.6}$$

Clearly, $|\rho_X(t, t')| \leq 1$.

The *autocorrelation function* is defined by

$$R_X(t, t') = E[X(t)X(t')]. \tag{1.7}$$

A straightforward calculation yields

$$K_X(t, t') = R_X(t, t') - \mu_X(t)\mu_X(t'). \tag{1.8}$$

If the mean is identically zero, then the covariance and autocorrelation functions are identical.

Given two random functions $X(t)$ and $Y(s)$, the *cross-covariance function* is defined as the covariance between pairs of random variables, one from each of the two processes:

$$
\begin{aligned}
K_{XY}(t,s) &= E[(X(t) - \mu_X(t))(Y(s) - \mu_Y(s))] \\
&= \int_{-\infty}^{\infty} (x - \mu_X(t))(y - \mu_Y(s))f(x,y;t,s)dxdy,
\end{aligned}
\tag{1.9}
$$

where $f(x, y; t, s)$ is the joint density for $X(t)$ and $Y(s)$. If the cross-covariance function is identically zero, then the random functions are said to be *uncorrelated*; otherwise, they are said to be *correlated*. As with the covariance function, there is symmetry: $K_{XY}(t,s) = K_{YX}(s,t)$. The *cross-correlation coefficient* is defined by

$$
\rho_{XY}(t,s) = \frac{K_{XY}(t,s)}{\sigma_X(t)\sigma_Y(s)},
\tag{1.10}
$$

and $|\rho_{XY}(t, s)| \leq 1$. The *cross-correlation function* is defined by

$$
R_{XY}(t,s) = E[X(t)Y(s)]
\tag{1.11}
$$

and is related to the cross-covariance function by

$$
K_{XY}(t,s) = R_{XY}(t,s) - E[X(t)]E[Y(s)].
\tag{1.12}
$$

Example 1.1. Consider the random time function $X(Z; t) = I_{[Z, \infty)}(t)$, where Z is the standard normal variable and $I_{[Z, \infty)}(t)$, the indicator (characteristic) function for the random infinite interval $[Z, \infty)$, is defined by $I_{[Z, \infty)}(t) = 1$ if $t \in [Z, \infty)$ and $I_{[Z, \infty)}(t) = 0$ if $t \notin [Z, \infty)$. For each observation z of the random variable Z, $X(z; t)$ is a unit step function with a step at $t = z$. For fixed t, $X(t)$ is a Bernoulli variable with $X(t) = 1$ if $t \geq Z$ and $X(t) = 0$ if $t < Z$. The density of $X(t)$ is characterized by the probabilities

$$
P(X(t) = 1) = P(Z \leq t) = \frac{1}{2\pi} \int_{-\infty}^{t} e^{-z^2/2}dz = \Phi(t),
$$

where Φ denotes the probability distribution function of Z, and

$$
P(X(t) = 0) = P(Z > t) = 1 - \Phi(t).
$$

The mean function for the process $X(t)$ is given by $\mu_X(t) = P(X(t) = 1) = \Phi(t)$. To find the covariance for $X(t)$, we first find the autocorrelation $R_X(t, t')$, recognizing that $X(t)X(t')$ is a binomial random variable. If $t < t'$, then $P(t' \geq Z \mid t \geq Z) = 1$, so that

$$P(X(t)X(t') = 1) = P(t' \geq Z, t \geq Z)$$
$$= P(t' \geq Z | t \geq Z)P(t \geq Z)$$
$$= P(t \geq Z)$$
$$= \Phi(t).$$

A similar calculation shows that, for $t' \leq t$, $P(X(t)X(t') = 1) = \Phi(t')$. Consequently,

$$P(X(t)X(t') = 1) = \Phi(\min(t, t')).$$

Thus, the autocorrelation and covariance are given by

$$R_X(t, t') = E[X(t)X(t')] = \Phi(\min(t, t')),$$
$$K_X(t, t') = E[X(t)X(t')] - \mu_X(t)\mu_X(t')$$
$$= \Phi(\min(t, t')) - \Phi(t)\Phi(t').$$

∎

For a sum, $X(t) + Y(t)$, of two random functions, linearity of expectation yields

$$\mu_{aX+bY}(t) = a\mu_X(t) + b\mu_Y(t) \tag{1.13}$$

for any real numbers a and b. For the covariance of a sum,

$$
\begin{aligned}
K_{X+Y}(t, t') = {}& E[(X(t) + Y(t))(X(t') + Y(t'))] \\
& - (\mu_X(t) + \mu_Y(t))(\mu_X(t') + \mu_Y(t')) \\
= {}& R_X(t, t') + R_Y(t, t') + R_{XY}(t, t') + R_{YX}(t, t') \\
& - \mu_X(t)\mu_X(t') - \mu_Y(t)\mu_Y(t') \\
& - \mu_X(t)\mu_Y(t') - \mu_Y(t)\mu_X(t') \\
= {}& K_X(t, t') + K_Y(t, t') + K_{XY}(t, t') + K_{YX}(t, t'),
\end{aligned}
\tag{1.14}
$$

where we have assumed that the relevant quantities exist. If $X(t)$ and $Y(t)$ are uncorrelated, then the cross-covariance is identically zero and the preceding identity reduces to the covariance of the sum being equal to the sum of the covariances:

$$K_{X+Y}(t, t') = K_X(t, t') + K_Y(t, t'). \tag{1.15}$$

The preceding relations generalize to sums of n random functions. Suppose that

$$W(t) = \sum_{j=1}^{n} X_j(t). \tag{1.16}$$

Then, assuming that the relevant quantities exist,

$$\mu_W(t) = \sum_{j=1}^{n} \mu_{X_j}(t), \tag{1.17}$$

$$K_W(t, t') = \sum_{i=1}^{n} \sum_{j=1}^{n} K_{X_i X_j}(t, t'). \tag{1.18}$$

Should the X_j be mutually uncorrelated, then all terms in the preceding summation for which $i \neq j$ are identically zero. Thus, the covariance reduces to

$$K_W(t, t') = \sum_{j=1}^{n} K_{X_j}(t, t'). \tag{1.19}$$

1.2 Calculus

Extending the calculus to random functions involves some subtlety because the difference quotient and Riemann sum defining the derivative and the integral, respectively, are random variables.

Random-process differentiation is made mathematically rigorous by defining the derivative of a random process via mean-square convergence. In general, $X_h(t)$ converges to $X(t)$ in the mean square (MS) if

$$\lim_{h \to 0} E[|X_h(t) - X(t)|^2] = 0. \tag{1.20}$$

For fixed h, $E[|X_h(t) - X(t)|^2]$ gives the mean-square distance between $X_h(t)$ and $X(t)$, so that MS convergence means that the distance between $X_h(t)$ and $X(t)$ is converging to 0.

The random function $X(t)$ is said to be *mean-square (MS) differentiable* and $X'(t)$ is the *mean-square derivative* of $X(t)$ at point t if

$$\lim_{h \to 0} E\left[\left|\frac{X(\omega; t+h) - X(\omega; t)}{h} - X'(\omega; t)\right|^2\right] = 0. \tag{1.21}$$

For fixed t and h, both the difference quotient and $X'(\omega; t)$ are functions of ω (random variables). The following theorem provides necessary and sufficient conditions for MS differentiability, and expressions for the mean and covariance functions of the derivative.

Theorem 1.1. The random function $X(t)$ is MS differentiable on the interval T if and only if the mean is differentiable on T and the covariance possesses a second-order mixed partial derivative with respect to u and v on T. In the case of differentiability,

$$\text{(i) } \mu_{X'}(t) = \frac{d}{dt}\mu_X(t), \tag{1.22}$$

$$\text{(ii) } K_{X'}(u.v) = \frac{\partial^2}{\partial u \partial v}K_X(u, v). \tag{1.23}$$

∎

The theory of MS differentiability extends to random functions of two variables: a partial derivative is an ordinary derivative applied to a function of two variables, with one of the variables being held fixed. The random function $Y(\omega; u, v)$ is the *MS partial derivative* of the random function $X(\omega; u, v)$ with respect to the variable u if

$$\lim_{h \to 0} E\left[\left|\frac{X(\omega; u + h, v) - X(\omega; u, v)}{h} - Y(\omega; u, v)\right|^2\right] = 0. \tag{1.24}$$

A two-dimensional analogue of Theorem 1.1 holds.

In ordinary calculus the integral of the time function $x(t)$ is defined as a limit of Riemann sums: for any partition $a = t_0 < t_1 < t_2 < \cdots < t_n = b$ of the interval $[a, b]$,

$$\int_a^b x(t)dt = \lim_{\|\Delta t_k\| \to 0} \sum_{k=1}^n x(t_k')\Delta t_k, \tag{1.25}$$

where $\Delta t_k = t_k - t_{k-1}$, $\|\Delta t_k\|$ is the maximum of the Δt_k over $k = 1, 2, \ldots, n$, t_k' is any point in the interval $[t_{k-1}, t_k]$, and the limit is taken to mean that the same value is obtained over all partitions, as long as $\|\Delta t_k\| \to 0$. The limit is, by definition, the value of the integral. The variable t is not restricted to a single dimension.

For a region of integration $T \subset \Re$, consider a disjoint collection of intervals I_k forming a partition $\Xi = \{I_k\}$ of T, meaning that $T = \cup_k I_k$. For each ω, we can form the Riemann sum corresponding to the realization of the random function $X(\omega; t)$ and the partition Ξ in a manner analogous to a deterministic function:

$$\Sigma_X(\omega; \Xi) = \sum_{k=1}^n X(\omega; t_k')\Delta(I_k), \tag{1.26}$$

where $t_k' \in I_k$. Letting $\|I_k\|$ be the maximum of the lengths, the limit can be taken over all partitions for which $\|I_k\| \to 0$ to give the integral

$$\int_T X(\omega; t)dt = \lim_{\Xi, \|I_k\| \to 0} \Sigma_X(\omega; \Xi). \tag{1.27}$$

Since the sum on the right-hand side of Eq. 1.26 depends not only on the partition, but also on the realization (on ω), the limit of Eq. 1.27 is a limit of random variables.

A random function $X(\omega; t)$ is said to be *mean-square integrable* and possess integral $\int_T X(\omega; t)dt$, itself a random variable, if and only if

$$\lim_{\Xi, \|I_k\| \to 0} E[|I - \Sigma_X(\omega, \Xi)|^2] = 0, \tag{1.28}$$

where the limit is taken over all partitions $\Xi = \{I_k\}$ for which $\|I_k\| \to 0$. The integral of a random function can often be obtained by integration of realizations; however, MS integrability depends on the limit of Eq. 1.28.

The basic mean-square integrability theorem concerns integrands of the form $g(t, s)X(s)$, where $g(t, s)$ is a deterministic function of two variables. The resulting integral is a random function of t and is not dependent on dimension.

Theorem 1.2. If the integral

$$Y(t) = \int_T g(t, s)X(s)ds \tag{1.29}$$

exists in the MS sense, then

$$\text{(i) } \mu_Y(t) = \int_T g(t, s)\mu_X(s)ds, \tag{1.30}$$

$$\text{(ii) } K_Y(t, t') = \int_T \int_T g(t, s)g(t', s')K_X(s, s')dsds'. \tag{1.31}$$

Conversely, if the deterministic integrals in (i) and (ii) exist, then the integral defining $Y(t)$ exists in the MS sense. ∎

Writing out (i) and (ii) in terms of the definitions of the mean and covariance shows that they state that integration and expectation can be interchanged. According to the theorem, these interchanges are justified if and only if there is MS integrability.

If we let $g(t, s)$ be a function of only s in Eq. 1.29, then the stochastic integral is just a random variable Y and Eq. 1.31 gives the variance of Y. Since the variance is nonnegative, the integral is nonnegative:

$$\int_T \int_T K_X(s, s')g(s)g(s')dsds' \geq 0. \tag{1.32}$$

The inequality holds for any domain T and any function $g(s)$ for which the integral exists. A function of two variables for which this is true is said to be

nonnegative definite. The requirement of nonnegative definiteness for a covariance function constrains the class of symmetric functions that can serve as covariance functions.

1.3 Three Fundamental Processes

This section discusses three random functions that are very important in applications and theory. All involve generalized functions.

1.3.1 Poisson process

The one-dimensional Poisson model is mathematically described in terms of points arriving randomly in time and letting $X(t)$ count the number of points arriving in the interval $[0, t]$. Three assumptions are postulated:

(i) The numbers of arrivals in any finite set of non-overlapping intervals are independent.
(ii) The probability of exactly one arrival in an interval of length t is $\lambda t + o(t)$.
(iii) The probability of two or more arrivals in an interval of length t is $o(t)$.

The parameter λ is constant over all t intervals, and $o(t)$ represents any function $g(t)$ for which $\lim_{t \to 0} g(t)/t = 0$. Condition (ii) says that, for infinitesimal t, the probability of exactly one arrival in an interval of length t is λt plus a quantity very small in comparison to t, and condition (iii) says that, for infinitesimal t, the probability of two or more arrivals in an interval of length t is very small in comparison to t. The random time points are called *Poisson points*, and each realization of the *Poisson process* corresponds to a set of time points resulting from a particular observation of the arrival process. It can be proven that $X(t)$ possesses a Poisson density with mean and variance equal to λt, namely,

$$P(X(t) = k) = e^{-\lambda t} \frac{(\lambda t)^k}{k!} \tag{1.33}$$

for $k = 0, 1, 2, \ldots$.

It follows from assumption (i) that the Poisson process has independent increments: if $t < t' < u < u'$, then $X(u') - X(u)$ and $X(t') - X(t)$ are independent. Using this independence, we find the covariance function. If $t < t'$, then the autocorrelation is obtained as

$$
\begin{aligned}
E[X(t)X(t')] &= E[X(t)^2 + X(t)(X(t') - X(t))] \\
&= E[X(t)^2] - E[X(t)]^2 + E[X(t)]E[X(t')] \\
&= \mathrm{Var}[X(t)] + E[X(t)]E[X(t')] \\
&= \lambda t + \lambda^2 t t'.
\end{aligned}
\tag{1.34}
$$

Hence, for $t < t'$,

$$K_X(t, t') = \text{Var}[X(t)] = \lambda t. \tag{1.35}$$

Interchanging the roles of t and t' yields

$$K_X(t, t') = \text{Var}[X(\min(t, t'))] = \lambda \min(t, t') \tag{1.36}$$

for all t and t'.

Theorem 1.1 states that MS differentiability depends on existence of the mixed partial derivative $\partial^2 K_X(u, v)/\partial u \partial v$. The Poisson process is not MS differentiable, but its covariance function possesses the generalized mixed partial derivative

$$\frac{\partial^2 K_X(t, t')}{\partial t \partial t'} = \lambda \delta(t - t'). \tag{1.37}$$

Although we have defined differentiability in terms of MS convergence, it is possible to give meaning to a generalized differentiability for which differentiability of the process is related to the generalized mixed partial derivative of the covariance and the derivative of the mean; that is, Theorem 1.1 can be applied in a generalized sense.

If a random process has covariance function $\lambda \delta(t - t')$ and constant mean λ (which is the derivative of the mean of the Poisson process), then we refer to the process as the *generalized derivative of the Poisson process*. Proceeding heuristically, the Poisson process has step-function realizations, where steps occur at points in time randomly selected according to a Poisson density with parameter λ. For any realization, say $x(t)$, with steps at $t_1, t_2, \ldots$, the usual generalized derivative is given by

$$x'(t) = \sum_{k=1}^{\infty} \delta(t - t_k). \tag{1.38}$$

If we assume that the derivative of the Poisson process consists of the process whose realizations agree with the derivatives of the realizations of the Poisson process itself, then the derivative process is given by

$$X'(t) = \sum_{k=1}^{\infty} \delta(t - Z_k), \tag{1.39}$$

where the Z_k form a sequence of random Poisson points.

Since each realization of $X'(t)$ consists of a train of pulses, the process is called the *Poisson impulse process*. If we now apply the generalized form of Theorem 1.1 that was alluded to previously, given that the mean and covariance of the Poisson process are λt and $\lambda \min(t, t')$, respectively, we conclude that the mean and covariance of the Poisson impulse process are given by $\mu_{X'}(t) = \lambda$ and $K_{X'}(t, t') = \lambda \delta(t - t')$.

The Poisson process is generated by random Poisson points, which are often said to model complete randomness, meaning that intuitively they model a "uniform" distribution of points across the infinite interval $[0, \infty)$. For application, a basic proposition states that the time distribution governing the arrival of the kth Poisson point following a given Poisson point is governed by a gamma distribution with $\alpha = k$ and $\beta = 1/\lambda$ (λt being the mean of the Poisson process). In particular, for $k = 1$, the inter-arrival time is governed by an exponential distribution with mean $1/\lambda$.

Up to this point we have based our discussion of the Poisson process on the conditions defining a Poisson arrival process, a key consequence being that the process possesses independent increments. In general, a random process $X(t)$, $t \geq 0$, is said to have *independent increments* if $X(0) = 0$ and, for any $t_1 < t_2 < \cdots < t_n$, the random variables $X(t_2) - X(t_1), X(t_3) - X(t_2), \ldots, X(t_n) - X(t_{n-1})$ are independent. The process has *stationary independent increments* if $X(t + r) - X(t' + r)$ is identically distributed to $X(t) - X(t')$ for any t, t', and r. When the increments are stationary, the increment distribution depends only on the length of time, $t - t'$, not the specific points in time, $t + r$ and $t' + r$. As defined via the arrival model, the Poisson process has stationary independent increments.

Axiomatically, we define a process $X(t)$ to be a *Poisson process* with *mean rate* λ if

P1. $X(t)$ has values in $\{0, 1, 2, \ldots\}$.
P2. $X(t)$ has stationary independent increments.
P3. For $s < t$, $X(t) - X(s)$ has a Poisson distribution with mean $\lambda(t - s)$.

The axiomatic formulation P1 through P3 completely captures the arrival model.

By generalizing the axioms P1 through P3, we can arrive at a definition of Poisson points in space. Consider points randomly distributed in Euclidean space $\mathfrak{R}^n$ and let $N(D)$ denote the number of points in a domain D. The points are said to be *distributed in accordance with a Poisson process with mean rate* λ if

1. For any disjoint domains $D_1, D_2, \ldots, D_r$, the counts $N(D_1), N(D_2), \ldots, N(D_r)$ are independent random variables.
2. For any domain D of finite volume, $N(D)$ possesses a Poisson distribution with mean $\lambda v(D)$, where $v(D)$ denotes the volume (measure) of D.

1.3.2 White noise

A zero-mean random function $X(k)$ defined on a discrete domain (taken to be the positive integers) is called *discrete white noise* if $X(k)$ and $X(j)$ are uncorrelated for $k \neq j$. If $X(k)$ is discrete white noise, then its covariance is given by

$$K_X(k, j) = E[X(k)X(j)] = \text{Var}[X(k)]\delta_{kj}, \tag{1.40}$$

where $\delta_{kj} = 1$ if $k = j$, and $\delta_{kj} = 0$ if $k \neq j$. For any function g defined on the integers and for all k,

$$\sum_{i=1}^{\infty} K_X(k, i)g(i) = \text{Var}[X(k)]g(k). \tag{1.41}$$

If there were a similar process in the continuous setting and $X(t)$ were such a random function defined over domain T, then the preceding equation would take the form

$$\int_T K_X(t, t')g(t')dt' = I(t)g(t), \tag{1.42}$$

where $I(t)$ is a function of t that plays the role played by $\text{Var}[X(k)]$ in Eq. 1.41. If we set

$$K_X(t, t') = I(t)\delta(t - t'), \tag{1.43}$$

then we obtain Eq. 1.42. Hence, any zero-mean random function having a covariance of the form $I(t)\delta(t - t')$ is called *continuous white noise*. White noise plays key roles in canonical representation, noise modeling, and design of optimal filters.

Continuous white noise does not exist from a standard mathematical perspective and requires generalized functions for a rigorous definition. We shall manipulate white noise formally under the assumption that the manipulations are justified in the context of generalized functions. It follows from the covariance function $I(t)\delta(t - t')$ that continuous white noise processes have infinite variance (set $t = t'$) and uncorrelated variables. $I(t)$ is called the *intensity* of the white noise process.

For an approximation of continuous white noise in one dimension, there exists a normal, zero-mean, stochastic process $X(t)$ having covariance function $K_X(t, t') = e^{-b|t-t'|}$, for $b > 0$. For very large values of b, this covariance function behaves approximately like the covariance of white noise.

1.3.3 Wiener process

Suppose that a particle, starting at the origin (in $\Re$) moves a unit length to the right or left. Movements are taken independently, and for each movement the probabilities of going right or left are p and $q = 1 - p$, respectively. Let $X(n)$ be the number of units the particle has moved to the right after n movements. $X(n)$ is called the one-dimensional *random walk*. The range of $X(n)$ is the set of integers $\{-n, -n + 2, \ldots, n - 2, n\}$.

To find the density for $X(n)$, let Y be the binomial random variable for n trials with probability p of success. For $X(n) = x$, there must be $(n + x)/2$ movements to the right and $(n - x)/2$ movements to the left. Hence, $X(n) = x$ if and only if $Y = (n + x)/2$. Hence,

$$f_{X(n)}(x) = P\left(Y = \frac{n+x}{2}\right) = \binom{n}{\frac{n+x}{2}} p^{(n+x)/2} q^{(n-x)/2}. \qquad (1.44)$$

Owing to the relationship between $X(n)$ and Y, the moments of $X(n)$ can be evaluated in terms of moments of $Y = (n + X(n))/2$. Specifically, $X(n) = 2Y - n$ and

$$E[X(n)^m] = \sum_{y=0}^{n} (2y - n)^m P(Y = y). \qquad (1.45)$$

Since Y is binomial, $E[Y] = np$, and $E[Y^2] = npq + n^2 p^2$. Letting $m = 1$ and $m = 2$ yields

$$E[X(n)] = 2np - n, \qquad (1.46)$$

$$E[X(n)^2] = 4(npq + n^2 p^2) + (1 - 4p)n^2. \qquad (1.47)$$

When movements to the right and left are equiprobable, so that $p = q = \frac{1}{2}$, $E[X(n)] = 0$ and $\mathrm{Var}[X(n)] = E[X(n)^2] = n$. Since the process has stationary independent increments, the covariance argument applied to the Poisson process also applies here. Thus,

$$K_X(n, n') = \mathrm{Var}[X(\min(n, n'))] = \min(n, n'). \qquad (1.48)$$

In the random-walk analysis the time interval between movements is 1; however, the analysis can be adapted to any finite time interval. Moreover, instead of restricting the motion to a single dimension, it could be analyzed in two dimensions, where the random walker can choose between four directions: left, right, up, and down. Taking a limiting situation, one can imagine an infinitesimal particle being continually acted upon by forces from its environment. The motion of such a particle can appear spasmodic, and under suitable phenomenological conditions such motion is referred to as *Brownian motion*.

Suppose that a particle is experiencing Brownian motion and $X(t)$ is its displacement in a single dimension from its original initial position. Assuming the particle's motion results from a multitude of molecular impacts lacking regularity, it is reasonable to suppose that $X(t)$ has independent increments. Moreover, assuming that the nature of the particle, the medium, and the relationship between the particle and its medium remain stable, and that the displacement over any time interval depends only on the elapsed time, not on the moment the time period commences, it is reasonable to postulate stationarity of the increments. Next, suppose that for any fixed t, $X(t)$ is normally distributed with mean zero, the latter reflecting forces acting on the particle without directional bias.

In accordance with the foregoing considerations, a *Wiener process* $X(t)$, $t \geq 0$, is defined to be a random function satisfying the following conditions:

1. $X(0) = 0$.
2. $X(t)$ has stationary independent increments.
3. $E[X(t)] = 0$.
4. For any fixed t, $X(t)$ is normally distributed.

Based on these conditions, one can verify that, for $0 \leq t' < t$, the increment $X(t) - X(t')$ has mean zero and variance $\sigma^2|t - t'|$, where σ^2 is a parameter to be empirically determined. In particular, $\text{Var}[X(t)] = \sigma^2 t$ for $t \geq 0$. As for the covariance, based on independent increments, the same argument used for the random walk process can be applied to obtain

$$K_X(t, t') = \text{Var}[X(\min(t, t'))] = \sigma^2 \min(t, t'). \tag{1.49}$$

Up to a multiplicative constant, the Poisson and Wiener processes have the same covariance. If we take the generalized mixed partial derivative of the Wiener process covariance, we obtain, as in the Poisson case, a constant times a delta function:

$$\frac{\partial^2 K_X(t, t')}{\partial t \partial t'} = \sigma^2 \delta(t - t'). \tag{1.50}$$

Since $\mu_X(t) \equiv 0$, $d\mu_X(t)/dt \equiv 0$. Consequently, via the generalized version of Theorem 1.1, $K_{X'}(t, t') = \sigma^2 \delta(t - t')$, and the derivative of the Wiener process is white noise.

1.4 Stationarity

In general, the nth-order probability distributions of a random function at two different sets of time points need not have any particular relation to each other. This section discusses two situations in which they do. First, the covariance function of a random process $X(t)$ is generally a function of two variables; however, in some cases it is a function of the difference between the variables. A stronger relation occurs when the nth-order probability distribution itself is invariant under a translation of the time point set.

If the covariance function of the random function $X(t)$ can be written as

$$K_X(t, t') = k_X(\tau), \tag{1.51}$$

where $\tau = t - t'$ (scalar or vector) and $X(t)$ has a constant mean μ_X, then $X(t)$ is said to be *wide-sense (WS) stationary*. Its variance function is constant:

$$\text{Var}[X(t)] = K_X(t, t) = k_X(t - t) = k_X(0). \tag{1.52}$$

Owing to the symmetry of $K_X(t, t')$, the covariance function of $X(t)$ is an even function,

$$k_X(-\tau) = k_X(t' - t) = k_X(t - t') = k_X(\tau). \tag{1.53}$$

Hence, $k_X(\tau) = k_X(|\tau|)$. The correlation coefficient reduces to a function of τ:

$$\rho_X(\tau) = \frac{k_X(\tau)}{k_X(0)}. \tag{1.54}$$

Since $|\rho_X(\tau)| \leq 1$, $|k_X(\tau)| \leq k_X(0)$. The autocorrelation is also a function of τ:

$$R_X(t, t') = k_X(\tau) + \mu_X^2 = r_X(\tau). \tag{1.55}$$

A random function $X(t)$ is WS stationary if and only if its covariance function is *translation invariant*, which means that, for any increment h,

$$K_X(t + h, t' + h) = K_X(t, t'). \tag{1.56}$$

To see this, suppose that $X(t)$ is WS stationary. Then

$$K_X(t + h, t' + h) = k_X(t + h - (t' + h)) = k_X(t - t') = K_X(t, t'). \tag{1.57}$$

Conversely, if the covariance function is translation invariant, then

$$K_X(t, t') = K_X(t - t', t' - t') = K_X(t - t', 0), \tag{1.58}$$

which is a function of $t - t'$.

Example 1.2. Let $Y(t)$ be the Poisson process with mean λt and r be a positive constant. The *Poisson increment process* is defined by

$$X(t) = Y(t + r) - Y(t).$$

According to the Poisson model, $X(t)$ counts the number of points in $[t, t + r]$, and

$$\mu_X(t) = E[Y(t + r)] - E[Y(t)] = \lambda r.$$

For the covariance $K_X(t, t')$, there are two cases: $|t - t'| > r$ and $|t - t'| \leq r$. If $|t - t'| > r$, then the intervals determined by t and t' are nonoverlapping and, owing to independent increments, $X(t)$ and $X(t')$ are independent, and their covariance is 0. Suppose that $|t - t'| \leq r$. First consider the case where $t < t'$. Then $t < t' < t + r < t' + r$, and we can apply the result of Eq. 1.34 together with the observation that, because the process counts the number of points in $[t, t + r]$, its mean is λr. Thus,

$$K_X(t, t') = E[(Y(t' + r) - Y(t'))(Y(t + r) - Y(t))] - E[X(t')]E[X(t)]$$
$$= E[Y(t' + r)Y(t + r)] - E[Y(t' + r)Y(t)] - E[Y(t + r)Y(t')]$$
$$+ E[Y(t')Y(t)] - \lambda^2 r^2$$
$$= \lambda(t + r) + \lambda^2(t' + r)(t + r) - \lambda t - \lambda^2 t(t' + r) - \lambda t'$$
$$- \lambda^2 t'(t + r) + \lambda t + \lambda^2 tt' - \lambda^2 r^2$$
$$= \lambda[r - (t' - t)].$$

Owing to symmetry, interchanging the roles of t and t' ($t' \leq t$) yields

$$K_X(t, t') = \lambda(r - |t - t'|)$$

when $|t - t'| \leq r$. Hence, $X(t)$ is WS stationary with

$$k_X(\tau) = \begin{cases} \lambda(r - |\tau|), & \text{if } |\tau| \leq r \\ 0, & \text{if } |\tau| > r \end{cases}.$$ ∎

A stronger form of stationarity concerns higher-order probabilistic information. The random function $X(t)$ is said to be *strict-sense stationary (SS stationary)* if, for any points $t_1, t_2, \ldots, t_n$, and for any increment h, its nth-order distribution function satisfies the relation

$$F(x_1, x_2, \ldots, x_n; t_1 + h, t_2 + h, \ldots, t_n + h) = F(x_1, x_2, \ldots, x_n; t_1, t_2, \ldots, t_n). \quad (1.59)$$

In terms of the nth-order density,

$$f(x_1, x_2, \ldots, x_n; t_1 + h, t_2 + h, \ldots, t_n + h) = f(x_1, x_2, \ldots, x_n; t_1, t_2, \ldots, t_n). \quad (1.60)$$

Given any finite set of random variables from the random function, a spatial translation of each by a constant h results in a collection of random variables whose multivariate distribution is identical to that of the original collection. From a probabilistic perspective, the new collection is indistinguishable from the first. If we define the random vector

$$\mathbf{X}(t_1, t_2, \ldots, t_n) = \begin{pmatrix} X(t_1) \\ X(t_2) \\ \vdots \\ X(t_n) \end{pmatrix}, \quad (1.61)$$

then $\mathbf{X}(t_1 + h, t_2 + h, \ldots, t_n + h)$ is identically distributed to $\mathbf{X}(t_1, t_2, \ldots, t_n)$.

If we restrict Eq. 1.59 to a single point, then it becomes $F(x; t + h) = F(x; t)$. $X(t + h)$ is identically distributed to $X(t)$, and therefore the mean at $t + h$ must equal the mean at t for all h, which implies that the mean must be constant.

Now consider two points such that

$$F(x_1, x_2; t_1 + h, t_2 + h) = F(x_1, x_2; t_1, t_2). \tag{1.62}$$

$\mathbf{X}(t_1 + h, t_2 + h)$ is identically distributed to $\mathbf{X}(t_1, t_2)$. Hence,

$$K_X(t_1 + h, t_2 + h) = K_X(t_1, t_2). \tag{1.63}$$

Therefore, the covariance function is translation invariant, and $X(t)$ is WS stationary. In summary, SS stationarity implies WS stationarity.

For a Gaussian random function, SS and WS stationarity are equivalent: since a Gaussian process is completely described by its first- and second-order moments and these are translation invariant for a WS stationary process, the higher-order probability distribution functions must also be translation invariant.

1.5 Linear Systems

If Ψ is a linear operator on a class of random functions, then, by superposition,

$$\Psi(a_1 X_1 + a_2 X_2) = a_1 \Psi(X_1) + a_2 \Psi(X_2). \tag{1.64}$$

For $Y = \Psi(X)$, we desire $\mu_Y(s)$ and $K_Y(s, s')$ in terms of $\mu_X(t)$ and $K_X(t, t')$, respectively. Schematically, we would like to find operations to complete (on the bottom horizontal arrows) the following commutative diagrams involving the expectation and covariance:

$$
\begin{array}{ccc}
X(t) & \xrightarrow{\ \Psi\ } & Y(s) \\
\downarrow{\scriptstyle E} & & \downarrow{\scriptstyle E} \\
\mu_X(t) & \longrightarrow & \mu_Y(s)
\end{array}
\tag{1.65}
$$

$$
\begin{array}{ccc}
X(t) & \xrightarrow{\ \Psi\ } & Y(s) \\
\downarrow{\scriptstyle \mathrm{Cov}} & & \downarrow{\scriptstyle \mathrm{Cov}} \\
K_X(t, t') & \longrightarrow & K_Y(s, s')
\end{array}
\tag{1.66}
$$

We have considered the cases where the operators are differentiation and integration in Theorems 1.1 and 1.2, respectively, a key point being interchange of the linear operator and the expectation, $E[\Psi(X)] = \Psi[E(X)]$. In terms of the relation $Y(s) = \Psi(X)(s)$, the interchange can be written as $\mu_Y(s) = \Psi(\mu_X)(s)$, and thus commutativity is achieved in the diagram of Eq. 1.65 with Ψ on the bottom arrow. Although interchange of expectation and a linear operator is not always valid, it is valid in practical situations, and henceforth we assume conditions to be such that it is justified.

For the moment, let us focus on linear systems operating on deterministic functions, in particular, integral operators defined in terms of a weighting function $g(s, t)$ via

$$y(s) = \int_T g(s, t)x(t)dt, \tag{1.67}$$

where $x(t)$ belongs to some linear space of functions and the variables can be scalars or vectors. Whereas $x(t)$ is defined over T, the output function $y(s)$ is defined over some set of values $s \in S$, where S need not equal T.

In the discrete sense,

$$y(n) = \sum_{k=-\infty}^{\infty} g(n, k)x(k) = \int_{-\infty}^{\infty} g(n, t)x(t)dt, \tag{1.68}$$

where

$$x(t) = \sum_{k=-\infty}^{\infty} x(k)\delta(t - k). \tag{1.69}$$

If Ψ is a linear operator on a linear function space L, the functions $x_1(t), x_2(t), \ldots, x_n(t)$ lie in L,

$$x(t) = \sum_{k=1}^{n} a_k x_k(t), \tag{1.70}$$

and $y(s) = \Psi(x)(s)$, then

$$y(s) = \sum_{k=1}^{n} a_k y_k(s), \tag{1.71}$$

where $y_k(s) = \Psi(x_k)(s)$ for $k = 1, 2, \ldots, n$. Superposition applies to finite sums of input functions; should a sum be infinite, and even converge, interchanging summation with the operator may not be valid, or, to achieve validity, the procedure might have to be interpreted in some specialized sense. When the functions involved are "well-behaved," such interchange can often take place.

More generally, if a function $x(t)$ is represented as an integral,

$$x(t) = \int_U a(u)Q(t, u)du, \tag{1.72}$$

and Ψ is a linear operator such that for each fixed u, $Q(t, u)$ is in the domain of Ψ, can we interchange the order of integration and application of Ψ and write

$$y(s) = \Psi(x)(s) = \int_U a(u)[\Psi_t(Q(t, u))](s)du, \tag{1.73}$$

where the subscript t of Ψ_t denotes that Ψ is applied relative to the variable t (for fixed u)? Validity depends on the function class involved. Two points are germane: (1) conditions can be imposed on the function class to make the interchange valid; (2) interchange facilitates the use of weighting functions to represent linear system laws, and the suggestiveness of such representations makes interchange of laws and integrals, at least in a formal manner, invaluable. Consequently, we will apply superposition freely to functions defined in terms of weighting functions, recognizing that for finite sums (or for weighting functions that are finite sums of delta functions), the application is mathematically rigorous.

If Ψ and $x(t)$ are defined by Eqs. 1.67 and 1.72, respectively, then, by superposition,

$$
\begin{aligned}
y(s) &= \int_T \int_U g(s,t)a(u)Q(t,u)du\,dt \\
&= \int_U a(u)\left(\int_T g(s,t)Q(t,u)dt\right)du.
\end{aligned}
\tag{1.74}
$$

Combining this with Eq. 1.73 shows that

$$
\Psi_t Q(t,u)(s) = \int_T g(s,t)Q(t,u)dt.
\tag{1.75}
$$

Consider representation of a function via an integral with a delta function kernel:

$$
x(t) = \int_{-\infty}^{\infty} x(u)\delta(t-u)du,
\tag{1.76}
$$

where, for notational convenience only, we have employed functions of a single variable. Application of Ψ to $x(t)$ yields the output

$$
y(s) = \int_{-\infty}^{\infty} x(u)\Psi_t\delta(t-u)(s)du.
\tag{1.77}
$$

For random-function inputs, an operator defining the bottom arrow in the commuting diagram of Eq. 1.66 provides a formulation of the output covariance of a linear system in terms of the input covariance. To avoid cumbersome notation, two conventions will be adopted. First, equations may be shortened by not including the variable s subsequent to the operation. Although this practice will result in equations with the variable s on the left and no explicitly stated variable s on the right, no confusion should result if one keeps the meaning of the operations in mind. A second convention will be

omission of parentheses when the meaning of the operations is obvious. For instance, we may write ΨX instead of $\Psi(X)$.

For the centered random functions X_0 and Y_0, the identity $E\Psi = \Psi E$ yields

$$
\begin{aligned}
Y_0(s) &= Y(s) - \mu_Y(s) \\
&= \Psi[X(t) - \mu_X(t)](s) \\
&= \Psi[X_0(t)](s).
\end{aligned} \tag{1.78}
$$

Consequently,

$$
\begin{aligned}
K_Y(s, s') &= E[Y_0(s)Y_0(s')] \\
&= E[\Psi_t X_0(t)\Psi_{t'}X_0(t')] \\
&= E[\Psi_t\Psi_{t'}X_0(t)X_0(t')] \\
&= \Psi_t\Psi_{t'}E[X_0(t)X_0(t')] \\
&= \Psi_t\Psi_{t'}K_X(t, t').
\end{aligned} \tag{1.79}
$$

Since the roles of Ψ_t and $\Psi_{t'}$ can be interchanged, we obtain the next theorem, completing the commuting diagram of Eq. 1.66. The same technique applies to the autocorrelation.

Theorem 1.3. If $X(t)$ is a random function for which $\Psi E X = E\Psi X$, then

$$
\text{(i) } \mu_{\Psi X}(s) = \Psi(\mu_X(t)), \tag{1.80}
$$

$$
\text{(ii) } K_{\Psi X}(s, s') = \Psi_t\Psi_{t'}K_X(t, t') = \Psi_{t'}\Psi_t K_X(t, t'). \tag{1.81}
$$

∎

If $g(t, u)$ is the impulse response function for Ψ, and we let $Y = \Psi X$, then

$$
Y(t) = \int_{-\infty}^{\infty} g(t, u)X(u)du, \tag{1.82}
$$

and the conclusions of Theorem 1.3 can be rewritten as

$$
\text{(i') } \mu_Y(t) = \int_{-\infty}^{\infty} g(t, u)\mu_X(u)du, \tag{1.83}
$$

$$
\text{(ii') } K_Y(t, t') = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(t, u)g(t', u')K_X(u, u')dudu'. \tag{1.84}
$$

Letting $t = t'$ yields the output variance:

$$
\text{Var}[Y(t)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(t, u)g(t, u')K_X(u, u')dudu'. \tag{1.85}
$$

By letting Ψ be the differential operator and recognizing that Ψ_t and $\Psi_{t'}$ are partial derivatives with respect to t and t' for the differential operator, if $s = t$, then it follows from Eq. 1.81 that

$$K_{\Psi X}(t, t') = \frac{\partial \partial}{\partial t' \partial t} K_X(t, t').$$ (1.86)

This relation holds for generalized derivatives involving delta functions.