Appendix B

Probability Theory, Random Variables, and Random Processes

B.1 Probability Theory

Probability theory provides a mathematical characterization for random events. Such events are defined on an underlying probability space $(\Omega, \mathcal{E}, p(\cdot))$. The probability space consists of a sample space Ω of possible outcomes for random events, a set of random events $\mathcal{E}$ where each $A \in \mathcal{E}$ is a subset of Ω, and a probability measure $p(\cdot)$ defined on these subsets. Thus, $\mathcal{E}$ is a set of sets, and the probability measure $p(A)$ is defined for every set $A \in \mathcal{E}$. A probability space requires that the set $\mathcal{E}$ is a σ-field. Intuitively, a set of sets $\mathcal{E}$ is a σ-field if it contains all intersections, unions, and complements of its elements.[1] More precisely, $\mathcal{E}$ is a σ-field if the set of all possible outcomes Ω is one of the sets in $\mathcal{E}$, a set $A \in \mathcal{E}$ implies that $A^c \in \mathcal{E}$, and for any sets $A_1, A_2, \ldots$ with $A_i \in \mathcal{E}$, we have $\cup_{i=1}^{\infty} A_i \in \mathcal{E}$. The set $\mathcal{E}$ must be a σ-field for the probability of intersections and unions of random events to be defined. We also require that the probability measure associated with a probability space have the following three fundamental properties:

1. $p(\Omega) = 1$.
2. $0 \leq p(A) \leq 1$ for any event $A \in \mathcal{E}$.
3. If A and B are mutually exclusive, i.e. their intersection is zero, then $p(A \cup B) = p(A) + p(B)$.

Throughout this section, we only consider sets in $\mathcal{E}$, since the probability measure is only defined on these sets.

Several important characteristics of the probability measure $p(\cdot)$ can be derived from its fundamental properties. In particular, $p(A^c) = 1 - p(A)$. Moreover, consider sets $A_1, \ldots, A_n$, where A_i and A_j, $i \neq j$, are disjoint ($A_i \cap A_j = \emptyset$). Then if $A_1 \cup A_2 \cup \ldots \cup A_n = \Omega$, we have that $\sum_{i=1}^{n} p(A_i) = 1$. We call the set $\{A_1, \ldots, A_n\}$ with these properties a *partition* of Ω. For two sets A_i and A_j that are not disjoint, $p(A_i \cup A_j) = p(A_i) + p(A_j) - p(A_i \cap A_j)$. This leads to the *union bound*, which says that for any sets $A_1, \ldots, A_n$,

$$p(A_1 \cup A_2 \cup \ldots \cup A_n) \leq \sum_{i=1}^{n} p(A_i). \tag{B.1}$$

[1] We use the notation $A \cap B$ to denote the intersection of A and B, i.e. all elements in both A and B. The union of A and B, denoted $A \cup B$ is the set of all elements in A or B. The complement of a set $A \subset \Omega$, denoted by A^c, is defined as all elements in Ω that are not in the set A.

The occurence of one random event can affect the probability of another random event, since observing one random event indicates which subsets in $\mathcal{E}$ could have contributed to the observed outcome. To capture this effect, we define the probability of event B conditioned on the occurence of event A as $p(B|A) = p(A \cap B)/p(A)$, assuming $p(A) \neq 0$. This implies that

$$p(A \cap B) = p(A|B)p(B) = p(B|A)p(A). \tag{B.2}$$

The conditional probability $p(B|A) = p(A \cap B)/p(A)$ essentially normalizes the probability of B with respect to the outcomes associated with A, since it is known that A has occured. We obtain *Bayes Rule* from (B.2) as

$$p(B|A) = \frac{p(A|B)p(B)}{p(A)}. \tag{B.3}$$

Independence of events is a function of the probability measure $p(\cdot)$. In particular, events A and B are independent if $p(A \cap B) = p(A)p(B)$. This implies that $p(B|A) = p(B)$ and $p(A|B) = p(A)$.

B.2 Random Variables

Random variables are defined on an underlying probability space $(\Omega, \mathcal{E}, p(\cdot))$. In particular, a random variable X is a function mapping from the sample space Ω to a subset of the real line. If X takes discrete values on the real line it is called a discrete random variable, and if it takes continuous values it is called a continuous random variable. The *cumulative distribution function* (CDF) of a random variable X is defined as $P_X(x) \stackrel{\triangle}{=} p(X \leq x)$ for some $x \in \mathcal{R}$. The CDF is derived from the underlying probability space as $p(X \leq x) = p(X^{-1}(-\infty, x))$, where $X^{-1}(\cdot)$ is the inverse mapping from the real line to a subset of Ω: $X^{-1}(-\infty, x) = \{\omega \in \Omega : X(\omega) \leq x\}$. Properties of the CDF are based on properties of the underlying probability measure. In particular, the CDF satisfies $0 \leq P_X(x) = p(X^{-1}(-\infty, x)) \leq 1$. In addition, the CDF is nondecreasing: $P_X(x_1) \leq P_X(x_2)$ for $x_1 \leq x_2$. That is because $P_X(x_2) = p(X^{-1}(-\infty, x_2)) = p(X^{-1}(-\infty, x_1)) + p(X^{-1}(x_1, x_2)) \geq p(X^{-1}(-\infty, x_1)) = P_X(x_1)$.

The *probability density function* (pdf) of a random variable X is defined as the derivative of its CDF, $p_X(x) \stackrel{\triangle}{=} \frac{d}{dx}P_X(x)$. For X a continuous random variable $p_X(x)$ is a function over the entire real line. For X a discrete random variable $p_X(x)$ is a set of delta functions at the possible values of X. The pdf, also refered to as the *probability distribution* or *distribution* of X, defines the probability that X lies in a given range of values:

$$p(x_1 < X \leq x_2) = p(X \leq x_2) - p(X \leq x_1) = P_X(x_2) - P_x(x_1) = \int_{x_1}^{x_2} p_X(x)dx. \tag{B.4}$$

Since $P_X(\infty) = 1$ and $P_X(-\infty) = 0$, the pdf integrates to 1,

$$\int_{-\infty}^{\infty} p_X(x)dx = 1. \tag{B.5}$$

Note that the subscript X is often omitted from the pdf and CDF when it is clear from the context that these functions characterize the distribution of X. In this case the pdf is written as $p(x)$ and the CDF as $P(x)$.

The *mean* or *expected value* of a random variable X is its probabalistic average, defined as

$$\mu_X = \mathbf{E}[X] \stackrel{\triangle}{=} \int_{-\infty}^{\infty} xp_X(x)dx. \tag{B.6}$$

The expectation operator $\mathbf{E}[\cdot]$ is linear and can also be applied to functions of random variables. In particular, the mean of a function of X is given by

$$\mathbf{E}[g(X)] = \int_{-\infty}^{\infty} g(x)p_X(x)dx. \tag{B.7}$$

A function of particular interest is the *nth moment* of X,

$$\mathbf{E}[X^n] = \int_{-\infty}^{\infty} x^n p_X(x) dx. \tag{B.8}$$

The variance of X is defined in terms of its mean and second moment as

$$\text{Var}[X] = \sigma_X^2 \stackrel{\triangle}{=} \mathbf{E}[(X - \mu_X)^2] = \mathbf{E}[X^2] - \mu_X^2. \tag{B.9}$$

The variance characterizes the average squared difference between X and its mean μ_X. The standard deviation of X, σ_X, is the square root of its variance. From the linearity of the expectation operator, it is easily shown that for any constant c, $\mathbf{E}[cX] = c\mathbf{E}[X]$, $\text{Var}[cX] = c^2\text{Var}[X]$, $\mathbf{E}[X + c] = \mathbf{E}[X] + c$, and $\text{Var}[X + c] = \text{Var}[X]$. Thus, scaling a random variable by a constant scales its mean by the same constant and its variance by the constant squared. Adding a constant to a random variable shifts the mean by the same constant but doesn't affect the variance.

The distribution of a random variable X can be determined from its *characteristic function*, defined as

$$\phi_X(\nu) \stackrel{\triangle}{=} \mathbf{E}[e^{j\nu X}] = \int_{-\infty}^{\infty} p_X(x) e^{j\nu x} dx. \tag{B.10}$$

We see from (B.10) that the characteristic function $\phi_X(\nu)$ of $X(t)$ is the inverse Fourier transform of the distribution $p_X(x)$ evaluated at $f = \nu/(2\pi)$. Thus we can obtain $p_X(x)$ from $\phi_X(\nu)$ as

$$p_X(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \phi_X(\nu) e^{-j\nu x} dx. \tag{B.11}$$

This will become significant in finding the distribution for sums of random variables.

Let X be a random variable and $g(x)$ be a function on the real line. Let $Y = g(X)$ define another random variable. Then $P_Y(y) = \int_{x:g(x)\leq y} p_X(x)dx$. For g monotonically increasing and one-to-one this becomes $P_Y(y) = \int_{-\infty}^{g^{-1}(y)} p_X(x)dx$. For g monotonically decreasing and one-to-one this becomes $P_Y(y) = \int_{g^{-1}(y)}^{\infty} p_X(x)dx$.

We now consider joint random variables. Two random variables must share the same underlying probability space for their joint distribution to be defined. Let X and Y be two random variables defined on the same probability space $(\Omega, \mathcal{E}, p(\cdot))$. Their joint CDF is defined as $P_{XY}(x, y) \stackrel{\triangle}{=} p(X \leq x, Y \leq y)$. Their joint pdf (distribution) is defined as the derivative of the joint CDF:

$$p_{XY}(x, y) \stackrel{\triangle}{=} \frac{\partial^2 P_{XY}(x, y)}{\partial x \partial y}. \tag{B.12}$$

Thus,

$$P_{XY}(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} p_{XY}(v, w) dv dw. \tag{B.13}$$

For joint random variables X and Y, we can obtain the distribution of X by integrating the joint distribution with respect to Y:

$$p_X(x) = \int_{-\infty}^{\infty} p_{XY}(x, y) dy. \tag{B.14}$$

Similarly,

$$p_Y(y) = \int_{-\infty}^{\infty} p_{XY}(x, y) dx. \tag{B.15}$$

The distributions $p_X(x)$ and $p_Y(y)$ obtained in this manner are sometimes refered to as the *marginal* distributions relative to the joint distribution $p_{XY}(x,y)$. Note that the joint distribution must integrate to one:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p_{XY}(x,y)dxdy = 1. \tag{B.16}$$

The definitions for joint CDF and joint pdf of two random variables extend in a straightforward manner to any finite number of random variables.

As with random events, observing the value for one random variable can affect the probability of another random variable. We define the conditional distribution of the random variable Y given a realization $X = x$ of random variable X as $p_Y(y|X = x) = p_{XY}(x,y)/p_X(x)$. This implies that $p_{XY}(x,y) = p_Y(y|X = x)p_X(x)$. Independence between two random variables X and Y is a function of their joint distribution. Specifically, X and Y are independent random variables if their joint distribution $p_{XY}(x,y)$ factors into separate distributions for X and Y: $p_{XY}(x,y) = p_X(x)p_Y(y)$. For independent random variables, it is easily shown that for any functions $f(x)$ and $g(y)$, $\mathbf{E}[f(X)g(Y)] = \mathbf{E}[f(X)]\mathbf{E}[g(Y)]$.

For X and Y joint random variables with joint pdf $p_{XY}(x,y)$, we define their ijth *joint moment* as

$$\mathbf{E}[X^iY^j] \stackrel{\triangle}{=} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^iy^jp_{XY}(x,y)dxdy. \tag{B.17}$$

The *correlation* of X and Y is defined as $\mathbf{E}[XY]$. The *covariance* of X and Y is defined as $\text{Cov}[XY] \stackrel{\triangle}{=} \mathbf{E}[(X - \mu_X)(Y - \mu_Y)] = \mathbf{E}[XY] - \mu_X\mu_Y$. Note that the covariance and correlation of X and Y are equal if either X or Y has mean zero. The *correlation coefficient* of X and Y is defined in terms of their covariance and standard deviations as $\rho \stackrel{\triangle}{=} \text{Cov}[XY]/(\sigma_X\sigma_Y)$. We say that X and Y are *uncorrelated* if their covariance is zero or, equivalently, their correlation coefficient is zero. Note that uncorrelated random variables (i.e. X and Y with $\text{Cov}[XY] = \mathbf{E}[XY] - \mu_X\mu_Y = 0$) will have a nonzero correlation ($\mathbf{E}[XY] \neq 0$) if their means are not zero. For random variables $X_1, \ldots, X_n$, we define their *covariance matrix* $\mathbf{\Sigma}$ as an $n \times n$ matrix with ijth element $\mathbf{\Sigma}_{ij} = \text{Cov}[X_iX_j]$. In particular, the ith diagonal element of $\mathbf{\Sigma}$ is the variance of X_i: $\mathbf{\Sigma}_{ii} = \text{Var}[X_i]$.

Consider two independent random variables X and Y. Let $Z = X + Y$ define a new random variable on the probability space $(\Omega, \mathcal{E}, p(\cdot))$. We can show directly or by using characteristic functions that the distribution of Z is the convolution of the distributions of X and Y: $p_Z(z) = p_X(x) * p_Y(y)$. Equivalently, $\phi_Z(\nu) = \phi_X(\nu)\phi_Y(\nu)$. With this distribution it can be shown that $\mathbf{E}[Z] = \mathbf{E}[X] + \mathbf{E}[Y]$, and $\text{Var}[Z] = \text{Var}[X] + \text{Var}[Y]$. So for sums of independent random variables, the mean of the sum is the sum of the means and the variance of the sum is the sum of the variances.

A distribution that arises frequently in the study of communication systems is the Gaussian distribution. The Gaussian distribution for a random variable X is defined in terms of its mean μ_X and variance σ_X^2 as

$$p_X(x) = \frac{1}{\sqrt{2\pi}\sigma_X}e^{-[(x-\mu_X)^2/(2\sigma_X^2)]}. \tag{B.18}$$

The Gaussian distribution, also called the normal distribution, is denoted as $\mathcal{N}(\mu_X, \sigma_X^2)$. Note that the tail of the distribution, i.e. the value of $p_X(x)$ as x moves away from μ_X, decreases exponentially. The CDF $P_X(x) = p(X \leq x)$ for this distribution does not exist in closed form. It is defined in terms of the Gaussian Q function as

$$P_X(x) = p(X \leq x) = 1 - Q\left(\frac{x - \mu_X}{\sigma_X}\right), \tag{B.19}$$

where the Gaussian Q function, defined by

$$Q(x) \stackrel{\triangle}{=} \int_x^{\infty} \frac{1}{\sqrt{2\pi}}e^{-y^2/2}dy, \tag{B.20}$$

is the probability that a Gaussian random variable X with mean zero and variance one is bigger than x: $Q(x) = p(X \geq x)$ for $X \sim \mathcal{N}(0,1)$. The Gaussian Q function is related to the complementary error function as $Q(x) = .5\text{erfc}(x/\sqrt{2})$. These functions are typically calculated using standard computer math packages.

Let $\mathbf{X} = (X_1, \ldots, X_n)$ denote a vector of jointly Gaussian random variables. Their joint distribution is given by

$$p_{X_1 \ldots X_n}(x_1, \ldots, x_n) = \frac{1}{\sqrt{(2\pi)^n \det[\mathbf{\Sigma}]}} \exp\left[-.5(\mathbf{x} - \mu_{\mathbf{X}})^T \mathbf{\Sigma}^{-1} (\mathbf{x} - \mu_{\mathbf{X}})\right], \tag{B.21}$$

where $\mu_{\mathbf{X}} = \mathbf{E}[\mathbf{X}]^T = (\mathbf{E}[X_1], \ldots, \mathbf{E}[X_n])^T$ is the mean of $\mathbf{X}$ and $\mathbf{\Sigma}$ is the $n \times n$ covariance matrix of $\mathbf{X}$, i.e. $\mathbf{\Sigma}_{ij} = \text{Cov}[X_i X_j]$. It can be shown from (B.21) that for jointly Gaussian random variables X and Y, if $\text{Cov}[XY] = 0$ then $p_{XY}(x,y) = p_X(x) p_Y(y)$. In other words, Gaussian random variables that are uncorrelated are independent.

The underlying reason why the Gaussian distribution commonly occurs in communication system models is the Central Limit Theorem (CLT), which defines the limiting distribution for the sum of a large number of independent random variables with the same distribution. Specifically, let X_i be independent and identically distributed (i.i.d.) joint random variables. Let $Y_n = \sum_{i=1}^n X_i$ and $Z_n = (Y_n - \mu_{Y_n})/\sigma_{Y_n}$. The CLT states that the distribution of Z_n as n goes to infinity converges to a Gaussian distribution with mean zero and variance one: $\lim_{n\to\infty} p_{Z_n}(x) = \mathcal{N}(0,1)$. Thus, any random variable equal to the sum of a large number of i.i.d. random components has a distribution that is approximately Gaussian. For example, noise in a radio receiver typically consists of spurious signals generated by the various hardware components, and with a large number of i.i.d. components this noise is accurately modeled as Gauss-distributed.

Two other common distributions that arise in communication systems are the uniform distribution and the binomial distribution. A random variable X that is uniformly distributed has pdf $p_X(x) = 1/(b-a)$ for x in the interval $[a,b]$ and zero otherwise. A random phase θ is commonly modeled as uniformly-distributed on the interval $[0, 2\pi]$, which we denote as $\theta \sim \mathcal{U}[0, 2\pi]$. The binomial distribution often arises in coding analysis. Let X_i, $i = 1, \ldots, n$, be discrete random variables that take two possible values, 0 and 1. Suppose the X_i are i.i.d. with $p(X_i = 1) = p$ and $p(X_i = 0) = 1 - p$. Let $Y = \sum_{i=1}^n X_i$. Then Y is a discrete random variable that takes integer values $k = 0, 1, 2, \ldots$. The distribution of Y is the binomial distribution, given by

$$p(Y = k) = \left(\begin{array}{c} n \\ k \end{array} \right) p^k (1-p)^{n-k}, \tag{B.22}$$

where

$$\left(\begin{array}{c} n \\ k \end{array} \right) \triangleq \frac{n!}{k!(n-k)!}. \tag{B.23}$$

B.3 Random Processes

A random process $X(t)$ is defined on an underlying probability space $(\Omega, \mathcal{E}, p(\cdot))$. In particular, it is a function mapping from the sample space Ω to a set of real functions $\{x_1(t), x_2(t), \ldots\}$, where each $x_i(t)$ is a possible realization of $X(t)$. Samples of $X(t)$ at times $t_0, t_1, \ldots, t_n$ are joint random variables defined on the underlying probability space. Thus, the joint CDF of samples at times $t_0, t_1, \ldots, t_n$ is given by $P_{X(t_0)X(t_1)\ldots X(t_n)}(x_0, \ldots, x_n) = p(X(t_0) \leq x_0, X(t_1) \leq x_1, \ldots, X(t_n) \leq x_n)$. The random process $X(t)$ is fully characterized by its joint CDF $P_{X(t_0)X(t_1)\ldots X(t_n)}(x_0, \ldots, x_n)$ for all possible sets of sample times $\{t_0, t_1, \ldots, t_n\}$.

A random process $X(t)$ is stationary if for all T and all sets of sample times $\{t_0, \ldots, t_n\}$, we have that $p(X(t_0) \leq x_0, X(t_1) \leq x_1, \ldots, X(t_n) \leq x_n) = p(X(t_0 + T) \leq x_0, X(t_1 + T) \leq x_1, \ldots, X(t_n + T) \leq x_n)$. Intuitively, a random process is stationary if time shifts do not affect its probability. Stationarity of a process is

often difficult to prove since it requires checking the joint CDF of all possible sets of samples for all possible time shifts. Stationarity of a random process is often infered from the stationarity of the source generating the process.

The *mean* of a random process is defined as $E[X(t)]$. Since the mean of a stationary random process is independent of time shifts, it must be constant: $\mathbf{E}[X(t)] = \mathbf{E}[X(t-t)] = \mathbf{E}[X(0)] = \mu_X$. The autocorrelation of a random process is defined as $A_X(t, t+\tau) \stackrel{\triangle}{=} \mathbf{E}[X(t)X(t+\tau)]$. The autocorrelation of $X(t)$ is also called its second moment. Since the autocorrelation of a stationary process is independent of time shifts, $A_X(t, t+\tau) = \mathbf{E}[X(t-t)X(t+\tau-t)] = \mathbf{E}[X(0)X(\tau)] \stackrel{\triangle}{=} A_X(\tau)$. So for stationary processes, the autocorrelation depends only on the time difference τ between the samples $X(t)$ and $X(t+\tau)$ and not on the absolute time t. The autocorrelation of a process measures the correlation between samples of the process taken at different times.

Two random processes $X(t)$ and $Y(t)$ defined on the same underlying probability space have a joint CDF characterized by

$$\begin{aligned} & P_{X(t_0)X(t_1)\ldots X(t_n)Y(t'_0)\ldots Y(t'_m)}(x_0, \ldots, x_n, y_0, \ldots, y_m) \\ &= p(X(t_0) \leq x_0, \ldots, X(t_n) \leq x_n, Y(t'_0) \leq y_0, \ldots, Y(t'_m) \leq y_m) \end{aligned} \tag{B.24}$$

for all possible sets of sample times $\{t_0, t_1, \ldots, t_n\}$ and $\{t'_0, t_1, \ldots, t'_m\}$. Two random processes $X(t)$ and $Y(t)$ are independent if for all such sets we have that

$$\begin{aligned} & p_{X(t_0)X(t_1)\ldots X(t_n)Y(t'_0)\ldots Y(t'_m)}(X(t_0) \leq x_0, \ldots, X(t_n) \leq x_n, Y(t'_0) \leq y_0, \ldots, Y(t'_m) \leq y_m) \\ &= p_{X(t_0)X(t_1)\ldots X(t_n)}(X(t_0) \leq x_0, \ldots, X(t_n) \leq x_n) p_{Y(t'_0)\ldots Y(t'_m)}(Y(t'_0) \leq y_0, \ldots, Y(t'_m) \leq y_m) \end{aligned} \tag{B.25}$$

The cross-correlation between two random processes $X(t)$ and $Y(t)$ is defined as $A_{XY}(t, t+\tau) \stackrel{\triangle}{=} \mathbf{E}[X(t)Y(t+\tau)]$. The two processes are uncorrelated if $\mathbf{E}[X(t)Y(t+\tau)] = \mathbf{E}[X(t)]\mathbf{E}[Y(t+\tau)]$ for all t and τ. As with the autocorrelation, if both $X(t)$ and $Y(t)$ are stationary, the cross-correlation is only a function of τ: $A_{XY}(t, t+\tau) = \mathbf{E}[X(t-t)Y(t+\tau-t)] = \mathbf{E}[X(0)Y(\tau)] \stackrel{\triangle}{=} A_{XY}(\tau)$.

In most analysis of random processes we focus only on the first and second moments. *Wide-sense stationarity* is a notion of stationarity that only depends on the first two moments of a process, and it can also be easily verified. Specifically, a process is wide-sense stationary (WSS) if its mean is constant, $\mathbf{E}[X(t)] = \mu_X$, and its autocorrelation depends only on the time difference of the samples, $A_X(t, t+\tau) = \mathbf{E}[X(t)X(t+\tau)] = A_X(\tau)$. Stationary processes are WSS but in general WSS processes are not necessarily stationary. For WSS processes, the autocorrelation is a symmetric function of τ, since $A_X(\tau) = \mathbf{E}[X(t)X(t+\tau)] = \mathbf{E}[X(t+\tau)X(t)] = A_X(-\tau)$. Moreover, it can be shown that $A_X(\tau)$ takes its maximum value at $\tau = 0$, i.e. $|A_X(\tau)| \leq A_X(0) = \mathbf{E}[X^2(t)]$. As with stationary processes, if two processes $X(t)$ and $Y(t)$ are both WSS then their cross-correlation is independent of time shifts, and thus depends only on the time difference of the processes: $A_{XY}(t, t+\tau) = \mathbf{E}[X(0)Y(\tau)] = A_{XY}(\tau)$.

The power spectral density (PSD) of a WSS process is defined as the Fourier transform of its autocorrelation function with respect to τ:

$$S_X(f) = \int_{-\infty}^{\infty} A_X(\tau) e^{-j2\pi f\tau} d\tau. \tag{B.26}$$

The autocorrelation can be obtained from the PSD through the inverse transform:

$$A_X(\tau) = \int_{-\infty}^{\infty} S_X(f) e^{j2\pi f\tau} df. \tag{B.27}$$

The PSD takes its name from the fact that the expected power of a random process $X(t)$ is the integral of its PSD:

$$\mathbf{E}[X^2(t)] = A_X(0) = \int_{-\infty}^{\infty} S_X(f) df, \tag{B.28}$$

which follows from (B.27). Similarly, from (B.26) we get that $S_X(0) = \int_{-\infty}^{\infty} A_X(\tau)d\tau$. The symmetry of $A_X(\tau)$ can be used with (B.26) to show that $S_X(f)$ is also symmmetric, i.e. $S_X(f) = S_X(-f)$. *White noise* is defined as a zero mean WSS random process with a PSD that is constant over all frequencies. Thus, a white noise process $X(t)$ has $\mathbf{E}[X(t)] = 0$ and $S_X(f) = N_0/2$ for some constant N_0 which is typically refered to as the (one-sided) white noise PSD. By the inverse Fourier transform, the autocorrelation of white noise is given by $A_X(\tau) = (N_0/2)\delta(\tau)$. In some sense, white noise is the most random of all possible noise processes, since it decorrelates instantaneously.

Random processes are often filtered or modulated, and when the process is WSS the impact of these operations can be characterized in a simple way. In particular, if a WSS process with PSD $S_X(f)$ is passed through a linear time-invariant filter with frequency response $H(f)$, then the filter output is also a WSS process with power spectral density $|H(f)|^2 S_X(f)$. If a WSS process $X(t)$ with PSD $S_X(f)$ is multiplied by a carrier $\cos(2\pi f_c t + \theta)$ with $\theta \sim \mathcal{U}[0, 2\pi]$, the multiplication results in a WSS process $X(t)\cos(2\pi f_c t + \theta)$ with PSD $.25[S_X(f - f_c) + S_X(f + f_c)]$.

Stationarity and WSS are properties of the underlying probability space associated with a random process. We are also often interested in time-averages associated with random processes, which can be characterized by different notions of *ergodicity*. A random process $X(t)$ is *ergodic in the mean* if its time-averaged mean, defined as

$$\mu_X^{ta} = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} X(t)dt, \tag{B.29}$$

is constant for all possible realizations of $X(t)$. In other words, $X(t)$ is ergodic in the mean if $\lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} x_i(t)dt$ equals the same constant μ_X^{ta} for all possible realizations $x_i(t)$ of $X(t)$. Similarly, a random process $X(t)$ is *ergodic in the nth moment* if its time-averaged nth moment

$$\mu_{X^n}^{ta} = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} X^n(t)dt \tag{B.30}$$

is constant for all possible realizations of $X(t)$. We can also define ergodicity of $X(t)$ relative to its time-averaged autocorrelation

$$A_X^{ta}(\tau) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} X(t)X(t+\tau)dt. \tag{B.31}$$

Specifically, $X(t)$ is *ergodic in autocorrelation* if $\lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} x_i(t)x_i(t+\tau)dt$ equals the same value $A_X^{ta}(\tau)$ for all possible realizations $x_i(t)$ of $X(t)$. Ergodicity of the autocorrelation in higher order moments requires that the nmth order time-averaged autocorrelation

$$A_X^{ta}(n, m, \tau) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} X^n(t)X^m(t+\tau)dt \tag{B.32}$$

is constant for all realizations of $X(t)$. A process that is ergodic in all order moments and autocorrelations is called *ergodic*. Ergodicity of a process requires that its time-averaged nth moment and ijth autocorrelation, averaged over all time, be constant for all n, i, and j. This implies that the probability associated with an ergodic process is independent of time shifts, and thus the process is stationary. In other words, an ergodic process must be stationary. However, a stationary process can be either ergodic or nonergodic. Since an ergodic process is stationary,

$$\begin{aligned}
\mu_X^{ta} &= \mathbf{E}[\mu_X^{ta}] \\
&= \mathbf{E}\left[\lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} X(t)dt\right] \\
&= \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} \mathbf{E}[X(t)]dt \\
&= \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} \mu_X dt = \mu_X.
\end{aligned} \tag{B.33}$$

Thus, the time-averaged mean of $X(t)$ equals its probabilistic mean. Similarly,

$$\begin{aligned} A_X^{ta}(\tau) &= \mathbf{E}[A_X^{ta}(\tau)] \\ &= \mathbf{E}\left[\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T} X(t)X(t+\tau)dt\right] \\ &= \lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}\mathbf{E}[X(t)(t+\tau)]dt \\ &= \lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T} A_X(\tau)dt = A_X(\tau), \end{aligned} \tag{B.34}$$

so the time-averaged autocorrelation of $X(t)$ equals its probabilistic autocorrelation.

B.4 Gaussian Processes

Noise processes in communication systems are commonly modeled as a Gaussian process. A random process $X(t)$ is a Gaussian process if for all values of T and all functions $g(t)$ the random variable

$$X_g = \int_0^T g(t)X(t)dt \tag{B.35}$$

has a Gaussian distribution. Since a communication receiver typically uses an integrator in signal detection, this definition implies that if the channel introduces a Gaussian noise process at the receiver input, the distribution of the random variable associated with the noise at the output of the integrator will have a Gaussian distribution. The mean of X_g is

$$\mathbf{E}[X_g] = \int_0^T g(t)\mathbf{E}[X(t)]dt \tag{B.36}$$

and the variance is

$$\text{Var}[X_g] = \int_0^T\int_0^T g(t)g(s)\mathbf{E}[X(t)X(s)]dtds - (\mathbf{E}[X_g])^2 \tag{B.37}$$

If $X(t)$ is WSS these simplify to

$$\mathbf{E}[X_g] = \int_0^T g(t)\mu_X dt \tag{B.38}$$

and

$$\text{Var}[X_g] = \int_0^T\int_0^T g(t)g(s)R_X(s-t)dtds - (\mathbf{E}[X_g])^2. \tag{B.39}$$

Several important properties of Gaussian random processes can be obtained from the definition. In particular, if a Gaussian random process is input to a linear time-invariant filter, the filter output is also a Gaussian random process. Moreover, we expect samples $X(t_i), i = 0, 1, \ldots$ of a Gaussian random process to be jointly Gaussian random variables, and indeed that follows from the definition by setting $g(t) = \delta(t - t_i)$ in (B.35). Since these samples are Gaussian random variables, if the samples are uncorrelated, they are also independent. In addition, for a WSS Gaussian processes, the distribution of X_g in (B.35) only depends on the mean and autocorrelation of the process $X(t)$. Finally, note that a random process is completely defined by the joint probability of its samples over all sets of sample times. For a Gaussian process, these samples are jointly Gaussian with their joint distribution determined by the mean and autocorrelation of the process. Thus, since the underlying probability of a Gaussian process is completely determined by its mean and autocorrelation, there are no higher moments for the process, so a WSS Gaussian process is also stationary. Similarly, a Gaussian process that is ergodic in the mean and autocorrelation is an ergodic process.

Bibliography

[1] A. Papoulis and S.U. Pillai, *Probability, Random Variables and Stochastic Processes*, McGraw-Hill, 2002.

[2] A. Leon-Garcia, *Probability and Random Processes for Electrical Engineering*, 2nd Ed., Addison-Wesley, 1994.

[3] R.M. Gray and L.D. Davisson, *Random Processes: A Mathematical Approach for Engineers*, Prentice-Hall, 1986.

[4] W. B. Davenport, Jr. and W. L. Root, *An Introduction to the Theory of Random Signals and Noise*, McGraw Hill, 1987.

[5] H. Stark and J. W. Woods, *Probability and Random Processes with Applications to Signal Processing*, 3rd Ed., Prentice Hall, 2001.

[6] R. G. Gallager, *Discrete Stochastic Processes*. Kluwer, 1996.

[7] W. Feller, *An Introduction to Probability Theory and Its Applications*, Vol. I and Vol II, Wiley, 1968/1971.

[8] P. Billingsley, *Probability and Measure*, 3rd. Ed., Wiley, 1995.