# Chapter 7

# Random Processes

## 7.1 Correlation in Random Variables

A random variable $X$ takes on numerical values as the result of an experiment. Suppose that the experiment also produces another random variable, $Y$. What can we say about the relationship between $X$ and $Y$?.

One of the best ways to visualize the possible relationship is to plot the $(X, Y)$ pair that is produced by several trials of the experiment. An example of correlated samples is shown in Figure 7.1. The points fall within a somewhat elliptical contour, slanting downward, and centered at approximately (4,0). The points were created with a random number generator using a correlation coefficient of $\rho = -0.5$, $E[X] = 4$, $E[Y] = 0$. The mean values are the coordinates of the cluster center. The negative correlation coefficient indicates that an increase in $X$ above its mean value generally corresponds to a decrease in $Y$ below its mean value. This tendency makes it possible to make predictions about the value that one variable will take given the value of the other, something which can be useful in many settings.

The joint behavior of $X$ and $Y$ is fully captured in the joint probability distribution. If the random variables are continuous then it is appropriate to use a probability density function, $f_{XY}(x, y)$. We will presume that the pdf is known or can be estimated. Computation of the usual expected values is then straightforward.

$$E[X^m Y^n] = \iint_{-\infty}^{\infty} x^m y^n f_{XY}(x, y) dx dy \tag{7.1}$$

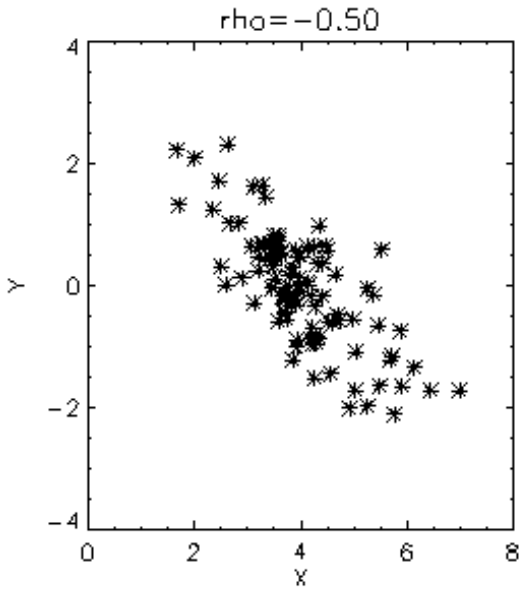


Figure 7.1: Scatter plot of random variables $X$ and $Y$. These random variables have a correlation of $\rho = -0.5$.

### 7.1.1 Covariance Function

The covariance function is a number that measures the common variation of $X$ and $Y$. It is defined as

$$\begin{aligned} cov(X,Y) &= E[(X - E[X])(Y - E[Y])] && (7.2) \\ &= E[XY] - E[X]E[Y] && (7.3) \end{aligned}$$

The covariance is determined by the difference in $E[XY]$ and $E[X]E[Y]$. If $X$ and $Y$ were statistically independent then $E[XY]$ would equal $E[X]E[Y]$ and the covariance would be zero. Hence, the covariance, as its name implies, measures the common variation. The covariance can be normalized to produce what is known as the correlation coefficient, $\rho$.

$$\rho = \frac{\mathrm{cov(X,Y)}}{\sqrt{\mathrm{var(X)var(Y)}}} \tag{7.4}$$

The correlation coefficient is bounded by $-1 \leq \rho \leq 1$. It will have value $\rho = 0$ when the covariance is zero and value $\rho = \pm 1$ when $X$ and $Y$ are perfectly

correlated or anti-correlated.

### 7.1.2 Autocorrelation Function

The autocorrelation[1] function is very similar to the covariance function. It is defined as

$$R(X,Y) = E[XY] = cov(X,Y) + E[X]E[Y] \tag{7.5}$$

It retains the mean values in the calculation of the value. The random variables are *orthogonal* if $R(X,Y) = 0$.

### 7.1.3 Joint Normal Distribution

If $X$ and $Y$ have a joint normal distribution then the probability density function is

$$f_{XY}(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left( -\frac{\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right) + \left(\frac{y-\mu_y}{\sigma_y}\right)^2}{2(1-\rho^2)} \right) \tag{7.6}$$

The contours of equal probability are ellipses, as shown in Figure 7.2. The probability changes much more rapidly along the minor axis of the ellipses than along the major axes. The orientation of the elliptical contours is along the line $y = x$ if $\rho > 0$ and along the line $y = -x$ if $\rho < 0$. The contours are a circle, and the variables are uncorrelated, if $\rho = 0$. The center of the ellipse is $(\mu_x, \mu_y)$.

## 7.2 Linear Estimation

It is often the case that one would like to estimate or predict the value of one random variable based on an observation of the other. If the random variables are correlated then this should yield a better result, on the average, than just guessing. We will see this is indeed the case.

[1] Be careful to not confuse the term "autocorrelation function" with "correlation coefficient".

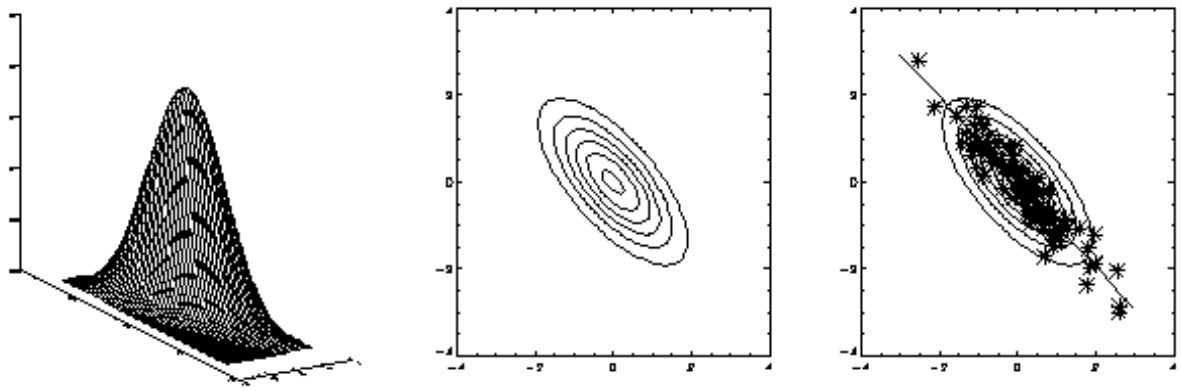

Figure 7.2: The normal probability distribution shown as a surface plot on the left and a contour plot in the center. A number of sample points are shown overlaid on the contour plot in the right frame. The linear predictor line is drawn in the right frame. $\rho = -0.7$, $\sigma_x = \sigma_y = 1$, $\mu_x = \mu_y = 0$.

The task is to construct a rule for the prediction of $Y$ based on an observation of $X$. We will call the prediction $\hat{Y}$, and compute its value with the simple linear equation

$$\hat{Y} = aX + b \tag{7.7}$$

where $a$ and $b$ are parameters to be chosen to provide the best results. We are encouraged to select this linear rule when we note that the sample points tend to fall about a sloping line. We would expect $a$ to correspond to the slope and $b$ to the intercept.

To find a means of calculating the coefficients from a set of sample points, construct the predictor error

$$\varepsilon = E[(Y - \hat{Y})^2] \tag{7.8}$$

We want to choose $a$ and $b$ to minimize $\varepsilon$. Therefore, compute the appropriate derivatives and set them to zero.

$$\frac{\partial \varepsilon}{\partial a} = -2E[(Y - \hat{Y})\frac{\partial \hat{Y}}{\partial a}] = 0 \tag{7.9}$$

$$\frac{\partial \varepsilon}{\partial b} = -2E[(Y - \hat{Y})\frac{\partial \hat{Y}}{\partial b}] = 0 \tag{7.10}$$

upon substitution of $\hat{Y} = aX + b$ and rearrangement we get the pair of equations

$$E[XY] = aE[X^2] + bE[X] \tag{7.11}$$
$$E[Y] = aE[X] + b \tag{7.12}$$

These can be solved for $a$ and $b$ in terms of the expected values. The expected values can be themselves estimated from the sample set.

$$a = \frac{\mathrm{cov(X, Y)}}{\mathrm{var(X)}} \tag{7.13}$$
$$b = E[Y] - \frac{\mathrm{cov(X, Y)}}{\mathrm{var(X)}} E[X] \tag{7.14}$$

The prediction error with these parameter values is

$$\varepsilon = (1 - \rho^2)\mathrm{var(Y)} \tag{7.15}$$

When the correlation coefficient $\rho = \pm 1$ the error is zero, meaning that perfect prediction can be made. When $\rho = 0$ the variance in the prediction is as large as the variation in $Y$, and the predictor is of no help at all. For intermediate values of $\rho$, whether positive or negative, the predictor reduces the error.

## 7.3 Random Processes

We have seen that a random variable $X$ is a rule which assigns a number to every outcome $e$ of an experiment. The random variable is a function $X(e)$ that maps the set of experiment outcomes to the set of numbers. A random process is a rule that maps every outcome $e$ of an experiment to a *function* $X(t, e)$. A random process is usually conceived of as a function of time, but there is no reason to not consider random processes that are functions of other independent variables, such as spatial coordinates. The function $X(u, v, e)$ would be a function whose value depended on the location $(u, v)$ and the outcome $e$, and could be used in representing random variations in an image.

In the following we will deal with one-dimensional random processes to develop a number of basic concepts. Having them in hand, we can then go on to multiple dimensions.

The domain of $e$ is the set of outcomes of the experiment. We assume that a probability distribution is known for this set. The domain of $t$ is a set, $\mathcal{T}$, of real numbers. If $\mathcal{T}$ is the real axis then $X(t, e)$ is a *continuous-time* random process, and if $\mathcal{T}$ is the set of integers then $X(t, e)$ is a *discrete-time* random process[2].

We can make the following statements about the random process:

1. It is a family of functions, $X(t, e)$. Imagine a giant strip chart recording in which each pen is identified with a different $e$. This family of functions is traditionally called an *ensemble*.

2. A single function $X(t, e_k)$ is selected by the outcome $e_k$. This is just a time function that we could call $X_k(t)$. Different outcomes give us different time functions.

3. If $t$ is fixed, say $t = t_1$, then $X(t_1, e)$ is a random variable. Its value depends on the outcome $e$.

4. If both $t$ and $e$ are given then $X(t, e)$ is just a number.

The particular interpretation must be understood from the context of the problem.

### 7.3.1 Averages

Because $X(t_1, e)$ is a random variable that represents the set of samples across the ensemble at time $t_1$, we can make use of all of the concepts that have been developed for random variables.

#### Moments

For example, if it has a probability density function[3] $f_X(x; t_1)$ then the moments are

$$m_n(t_1) = E[X^n(t_1)] = \int_{-\infty}^{\infty} x^n f_X(x; t_1)\, dx \tag{7.16}$$

[2] We will often suppress the display of the variable $e$ and write $X(t)$ for a continuous-time RP and $X[n]$ or $X_n$ for a discrete-time RP. Always go back to the basic definition when you need to be sure that an idea is clear.

[3] We will do the development in terms of random variables in which $X$ has a continuous range of values. If $X$ has a discrete set of possible values, then the integrals are replaced by sums, etc.

We need the notation $f_X(x;t_1)$ because it is very possible that the probability density will depend upon the time the samples are taken. The mean value is $\mu_X = m_1$, which can be a function of time. The central moments are

$$E[(X(t_1) - \mu_X(t_1))^n] = \int_{-\infty}^{\infty} (x - \mu_X(t_1))^n \, f_X\,(x;t_1)\,dx \tag{7.17}$$

**Correlation**

The numbers $X(t_1, e)$ and $X(t_2, e)$ are samples from the same time function at different times. This is a pair of random variables which we could write conveniently in terms of a doublet $(X_1, X_2)$. It is described by a joint probability density function[4] $f(x_1, x_2; t_1, t_2)$. The notation includes the times because the result surely can depend on when the samples are taken.

From the joint density function one can compute the marginal densities, conditional probabilities and other quantities that may be of interest. A measure of particular interest is the correlation and covariance. The covariance is[5]

$$C(t_1, t_2) = E[(X_1 - \mu_1)(X_2 - \mu_2)] = \iint_{-\infty}^{\infty} (x_1 - \mu_1)(x_2 - \mu_2) f(x_1, x_2; t_1, t_2) dx_1 dx_2 \tag{7.18}$$

The correlation function[6] is

$$R(t_1, t_2) = E[X_1 X_2] = \iint_{-\infty}^{\infty} x_1 x_2 f(x_1, x_2; t_1, t_2) dx_1 dx_2 \tag{7.19}$$

$$C(t_1, t_2) = R(t_1, t_2) - \mu_1 \mu_2 \tag{7.20}$$

Note that both the covariance and correlation functions are symmetric in $t_1$ and $t_2$. $C(t_1, t_2) = C(t_2, t_1)$ and $R(t_1, t_2) = R(t_2, t_1)$

The average power in the process at time $t$ is represented by

$$R(t, t) = E[X^2(t)] \tag{7.21}$$

and $C(t, t)$ represents the power in the fluctuation about the mean value.

[4] Here we will not use subscripts on the function, as in $f_{X_1 X_2}(x_1, x_2; t_1, t_2)$, to avoid becoming overly baroque. The subscripts are implied by the argument. Here we see a reason why this is such a common practice in probability publications.

[5] Although not explicitly shown, the mean values can be functions of $t_1$ and $t_2$.

[6] This correlation function is called *autocorrelation* when it is necessary to clearly indicate that the variable is being correlated with itself rather than another variable, which is called *cross-correlation.*

**Example 7.3.1** ***Poisson Process*** *Let $N(t_1, t_2)$ be the number of events produced by a Poisson process in the interval $(t_1, t_2)$ when the average rate is $\lambda$ events per second. The probability that $N = n$ is*

$$P[N = n] = \frac{(\lambda\tau)^n e^{-\lambda\tau}}{n!} \tag{7.22}$$

*where $\tau = t_2 - t_1$. Then $E[N(t_1, t_2)] = \lambda\tau$. A random process can be defined as the number of events in the interval $(0, t)$. Thus, $X(t) = N(0, t)$. The expected number of events in $t$ is $E[X(t)] = \lambda t$. For a Poisson distribution we know that the variance is*

$$E[(X(t) - \lambda t)^2] = E[X^2(t)] - (\lambda t)^2 = \lambda t \tag{7.23}$$

*from which we readily find that the "average power" in the function $X(t)$ is*

$$E[X^2(t)] = \lambda t + \lambda^2 t^2 \tag{7.24}$$

*A graph of $X(t)$ would show a function fluctuating about an average trend line with a slope $\lambda$. An example is shown in Figure 7.3.*
*Finding the correlation $R(t_1, t_2) = E[X(t_1)X(t_2)]$ has to solve the problem that $X(t_1)$ and $X(t_2)$ have overlapping ranges. If $t_2 > t_1$ then $X(t_1)$ and $X(t_2) - X(t_1)$ are statistically independent because the ranges do not overlap. Then expand the identity*

$$\begin{aligned} E[X(t_1)X(t_2)] &= E[X(t_1)(X(t_1) + X(t_2) - X(t_1))] \\ &= E[X^2(t_1)] + E[(X(t_1)(X(t_2) - X(t_1))] \end{aligned} \tag{7.25}$$

$$E[X(t_1)(X(t_2) - X(t_1))] = E[X(t_1)]E[X(t_2) - X(t_1)] = \lambda t_1 \lambda(t_2 - t_1) \tag{7.26}$$

*Combining Equations 7.24, 7.25 and 7.26 we now have*

$$R(t_1,t_2) = \lambda t_1 + \lambda^2 t_1^2 + \lambda t_1 \lambda(t_2 - t_1) = \lambda t_1 + \lambda^2 t_1 t_2 \quad \text{for } t_2 \geq t_1 \tag{7.27}$$

*Because $R(t_1,t_2) = R(t_2,t_1)$ we can get the result for the case $t_1 \geq t_2$ by interchanging variables. The final result is*

$$R(t_1,t_2) = \begin{cases} \lambda t_2 + \lambda^2 t_1 t_2, & t_1 \geq t_2 \\ \lambda t_1 + \lambda^2 t_1 t_2, & t_2 \geq t_1 \end{cases} \tag{7.28}$$

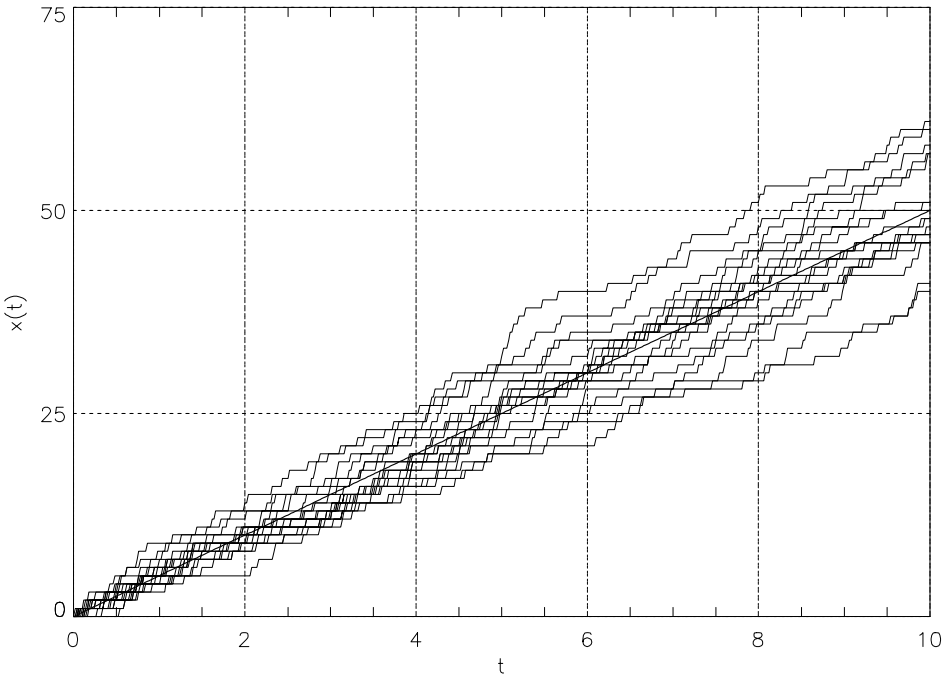


Figure 7.3: An ensemble of 20 Poisson random processes. The rate is $\lambda = 5$ so that in $t = 10$ seconds one would expect $\lambda t = 50$ events. Note that the variance about the trend line increases with time.

The Poisson process provides an illustration of the content of a photon detector over time. If one had an array of twenty photon detectors, then the photon count in the individual detectors may be like the individual tracks in Figure 7.3. For this process the correlation function is clearly a function of both $t_1$ and $t_2$. In many instances the result is a function only of $|t_2 - t_1|$. We will see an example of this below.

**Example 7.3.2** ***Telegraph Signal*** *Consider a random process that has the following properties: (1) $X(t) = \pm 1$, (2) the number of zero crossings in the interval $(0, t)$ is described by a Poisson process, and (3) $X(0) = 1$. We will remove the third condition later in the example, but it is helpful in setting up the result. We first find the expected value at time $t$. Let $N(t)$ equal the*

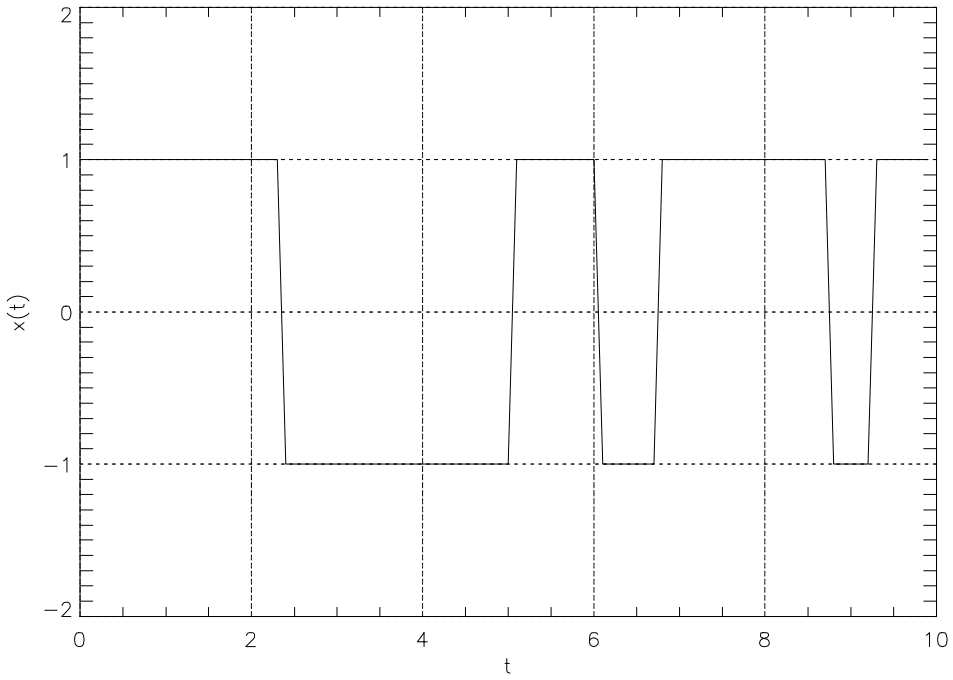


Figure 7.4: An example of a telegraph signal. The zero cossings are described by a Poisson process, with an average rate of $\lambda = 1$ per second.

*number of zero crossings in the interval* $(0, t)$. *with* $t \geq 0$. *Then*

$$P(N = n) = \frac{(\lambda t)^n \, e^{-\lambda t}}{n!} \tag{7.29}$$

$$\begin{aligned} P[X(t) = 1] &= P\left[N = \textit{even number}\right] \\ &= e^{-\lambda t}\left[1 + \frac{(\lambda t)^2}{2!} + \frac{(\lambda t)^4}{4!} + \cdots\right] \\ &= e^{-\lambda t}\cosh \lambda t \end{aligned} \tag{7.30}$$

$$\begin{aligned} P[X(t) = -1] &= P\left[N = \textit{odd number}\right] \\ &= e^{-\lambda t}\left[\lambda t + \frac{(\lambda t)^3}{3!} + \frac{(\lambda t)^5}{5!} + \cdots\right] \\ &= e^{-\lambda t}\sinh \lambda t \end{aligned} \tag{7.31}$$

*The expected value is*

$$E[X(t)] = e^{-\lambda t}\cosh \lambda t - e^{-\lambda t}\sinh \lambda t = e^{-2\lambda t} \tag{7.32}$$

*Note that the expected value decays toward $x = 0$ for large $t$. That is because the influence of knowing the value at $t = 0$ decays exponentially. The autocorrelation function is computed by finding $R(t_1, t_2) = E\left[X\left(t_1\right) X\left(t_2\right)\right]$. Let $x_0 = -1$ and $x_1 = 1$ denote the two values that $X$ can attain. For the moment assume that $t_2 \geq t_1$. Then*

$$R(t_1, t_2) = \sum_{j=0}^{1}\sum_{k=0}^{1} x_j x_k P[X(t_1) = x_k] P[X(t_2) = x_j | X(t_1) = x_k] \tag{7.33}$$

*The first term in each product is given above. To find the conditional probabilities we take note of the fact that the number of sign changes in $t_2 - t_1$ is a Poisson process. Hence, in a manner that is similar to the analysis above,*

$$P[X(t_2) = 1 | X(t_1) = 1] = P[X(t_2) = -1 | X(t_1) = -1] = e^{-\lambda(t_2 - t_1)}\cosh \lambda(t_2 - t_1) \tag{7.34}$$

$$P[X(t_2) = -1 | X(t_1) = 1] = P[X(t_2) = 1 | X(t_1) = -1] = e^{-\lambda(t_2 - t_1)}\sinh \lambda(t_2 - t_1) \tag{7.35}$$

*Hence*

$$\begin{aligned} R(t_1, t_2) &= e^{-\lambda t_1}\cosh \lambda t_1 \left[e^{-\lambda(t_2 - t_1)}\cosh \lambda(t_2 - t_1) - e^{-\lambda(t_2 - t_1)}\sinh \lambda(t_2 - t_1)\right] \\ &\quad -e^{-\lambda t_1}\sinh \lambda t_1 \left[e^{-\lambda(t_2 - t_1)}\cosh \lambda(t_2 - t_1) - e^{-\lambda(t_2 - t_1)}\sinh \lambda(t_2 - t_1)\right] \end{aligned}$$

*After some algebra this reduces to*

$$R(t_1, t_2) = e^{-\lambda(t_2 - t_1)} \quad \textit{for } t_2 \geq t_1 \tag{7.36}$$

*A parallel analysis applies to the case $t_2 \leq t_1$, so that*

$$R(t_1, t_2) = e^{-\lambda|t_2 - t_1|} \tag{7.37}$$

*The autocorrelation for the telegraph signal depends only upon the time difference, not the location of the time interval. We will see soon that this is a very important characteristic of stationary random processes. We can now remove condition (3) on the telegraph process. Let* $Y(t) = AX(t)$ *where* $A$ *is a random variable independent of* $X$ *that takes on the values* $\pm 1$ *with equal probability. Then* $Y(0)$ *will equal* $\pm 1$ *with equal probability, and the telegraph process will no longer have the restriction of being positive at* $t = 0$*. Since* $A$ *and* $X$ *are independent, the autocorrelation for* $Y(t)$ *is given by*

$$E\left[Y(t_1)Y\left(t_2\right)\right] = E[A^2]E\left[X(t_1)X\left(t_2\right)\right] = e^{-\lambda|t_2-t_1|} \tag{7.38}$$

*since* $E[A^2] = 1$.

### 7.3.2 Stationary Random Processes

The random telegraph is one example of a process that has at least some statistics that are independent of time. Random processes whose statistics do not depend on time are called *stationary*. In general, random processes can have joint statistics of any order. If the process is stationary, they are independent of time shift. We will explain this by building up the description.

The first order statistics are described by the cumulative distribution function $F(x;t)$. If the process is stationary then the distribution function at times $t = t_1$ and $t = t_2$ will be identical. An example of a random process is shown in Figure 7.5. At any particular time, $F[x;t] = P[X \leq x;t]$ is just the cumulative distribution function over the random variable $X(t)$. The distribution function is independent of time for a stationary process. Given the distribution function, we can compute other first-order statistics such as the probability density.

The second-order statistics are described by the joint statistics of random variables $X(t_1)$ and $X\left(t_2\right)$. The second-order cumulative distribution function is

$$F(x_1, x_2; t_1, t_2) = P[X(t_1) \leq x_1, X(t_2) \leq x_2] \tag{7.39}$$

from which one may compute the density

$$f(x_1, x_2; t_1, t_2) = \frac{\partial^2 F(x_1, x_2; t_1, t_2)}{\partial x_1 \partial x_2} \tag{7.40}$$

When the process is stationary these functions depend on the time separation $(t_2 - t_1)$ but not on the location of the interval.

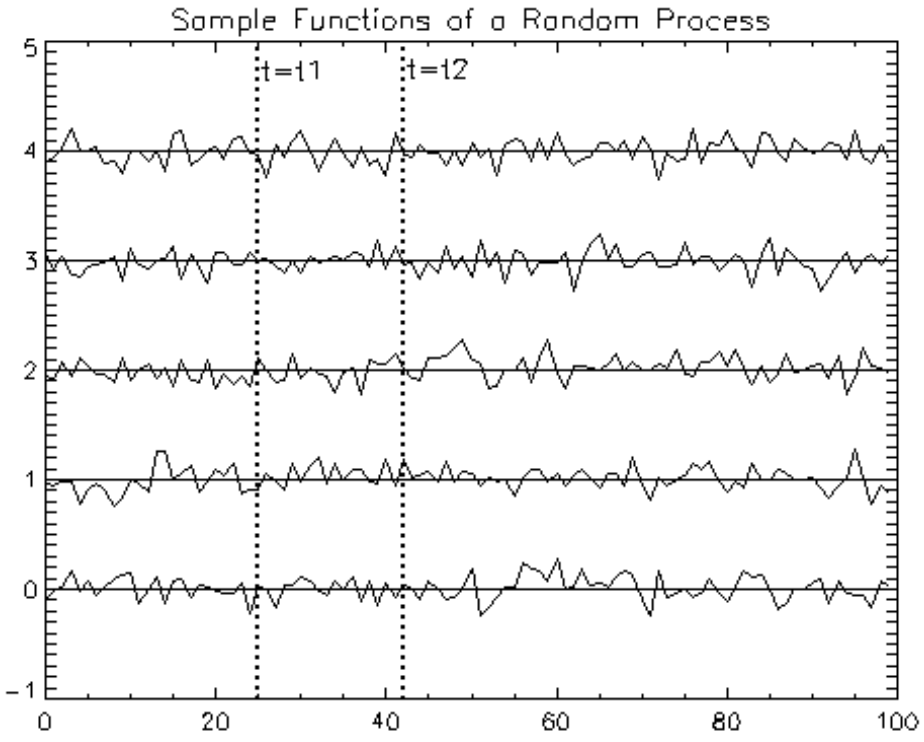


Figure 7.5: Several snapshots of sample functions of a random process are shown. Particular times $t = t_1$ and $t = t_2$ are labeled.

The statistics of a random process can be defined for any order $n$ and any set of sample times $\{t_1, \ldots, t_n\}$. For a stationary process, $F[x_1, \ldots, x_n; x_t, \ldots, x_t]$ is independent of translatiion through time so long as the relative locations of the sample times are maintained. If the process is stationary for all values of $n$ then it is said to be *strict-sense stationary.*

### Wide-sense Stationary Processes

In practice it is not possible to determine the statistics at all orders of $n$, although they may be postulated in a model. But we often are particularly interested in processes that are stationary up to at least order $n = 2$. Such processes are called *wide-sense stationary* (wss). If a process is wss then its mean, variance, autocorrelation function and other first and second order

statistical measures are independent of time. We have seen that a Poisson random process has mean $\mu(t) = \lambda t$, so it is not stationary in any sense. The telegraph signal has mean $\mu = 0$, variance $\sigma^2 = 1$ and autocorrelation function $R(t_1, t_2) = R(\tau) = e^{-\lambda\tau}$ where $\tau = |t_2 - t_1|$. It is a wss process. When a process is wss it is common practice to not show the variable $t$ in first-order parameters, to replace $t_2 - t_1$ by $\tau$, and to use the notation $R(\tau)$ instead of $R(t_1, t_2)$. The following is true for the autocorrelation function of a wss process

$$R(\tau) = E[X(t)X(t+\tau)] \tag{7.41}$$

for any value of $t$. Then

$$R(0) = E[X^2] \tag{7.42}$$

The covariance function is

$$C(\tau) = E[(X(t) - \mu)(X(t+\tau) - \mu)] = R(\tau) - \mu^2 \tag{7.43}$$

$$C(0) = E[(X(t) - \mu)^2] = \sigma^2 \tag{7.44}$$

Two random processes $X(t)$ and $Y(t)$ are called *jointly wide-sense stationary* if each is wss and their cross correlation depends only on $\tau = t_2 - t_1$. Then

$$R_{xy}(\tau) = E[X(t)Y(t+\tau)] \tag{7.45}$$

is called the cross-correlation function and

$$C_{xy}(\tau) = R_{xy}(\tau) - \mu_x\mu_y \tag{7.46}$$

is called the cross-covariance function.

### 7.3.3 Filtered Random Processes

It is often the case that we need to describe the output of a system when the input is a random process. This is generally the case with all communication systems and sensor systems. This is clearly a very large class of systems. Here we will deal only with linear systems because they lend themselves to general results. Nonlinear systems must be treated with special analytical tools on a case-by-case basis.

Let $X(t, e)$ be a random process. For the moment we show the outcome $e$ of the underlying random experiment that selects a particular function of time. An operation can be done on the time function to produce an output

$Y(t,e) = \mathcal{L}[X(t,e)]$. Clearly, $Y(t,e)$ is an ensemble of functions selected by $e$, and is a random process. We will now discontinue display of the outcome $e$.

Let $X_1(t)$ and $X_2(t)$ be any two random processes that are permitted as system inputs and let $a_1$ and $a_2$ be any scalar multipliers. The system represented by the operation $\mathcal{L}$ is linear if

$$\mathcal{L}[a_1X_1(t) + a_2X_2(t)] = a_1\mathcal{L}[X_1(t)] + a_2\mathcal{L}[X_2(t)] \tag{7.47}$$

The coefficients may themselves be random variables without affecting the definition of linearity. The system is time-invariant if the response to a time-shifted input is just the time-shifted output.

$$Y(t+\tau) = \mathcal{L}[X(t+\tau)] \tag{7.48}$$

A time-invariant linear system can be represented by its impulse response, $h(t)$, so that the output and input are related through the convolution:

$$Y(t) = \int_{-\infty}^{\infty} X(t-s)h(s)ds \tag{7.49}$$

### Mean Value

The following result holds for any linear system, whether or not it is time invariant and even stationary.

$$E[\mathcal{L}X(t)] = \mathcal{L}E[X(t)] = \mathcal{L}[\mu(t)] \tag{7.50}$$

When the process is stationary we find $\mu_y = \mathcal{L}[\mu_x]$, which is just the response to a constant of value $\mu_x$.

### Output Autocorrelation

We would like to know the autocorrelation function $R_{yy}(\tau)$ of the output of a linear system when its input is a wss random process. When the input is wss and the system is time invariant the output is also wss.

Let us first find the cross-correlation function

$$\begin{aligned} R_{xy}(\tau) &= E[X(t)Y(t+\tau)] \\ &= \int_{-\infty}^{\infty} E[X(t)X(t+\tau-s)]\,h(s)ds \\ &= \int_{-\infty}^{\infty} R_{xx}(\tau-s)h(s)ds \end{aligned} \tag{7.51}$$

The cross-correlation $R_{xy}(\tau)$ is the convolution of the autocorrelation function $R_{xx}(\tau)$ with the impulse response of the system, a very important result in its own right.

Now, multiply through Equation 7.48 by $Y(t)$ and take averages.

$$R_{yy}(\tau) = \mathcal{L}E[X(t+\tau)Y(t)] = \mathcal{L}R_{xy}(\tau) \tag{7.52}$$

Applying the shift-invariant linear filter as the operator,

$$R_{yy}(\tau) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} R_{xx}(\tau - s_1 + s_2)h(s_1)h(s_2)ds_1 ds_2 \tag{7.53}$$

**Example 7.3.3** ***Filtering a white noise process*** *White noise is defined as a stationary process $W(t)$ with the property that the $R_w(\tau) = \sigma_w^2\delta(\tau)$. The process is completely uncorrelated. If the input to a linear tim-invariant system is $X(t) = W(t)$ then*

$$\begin{aligned} R_{xy}(\tau) &= \int_{-\infty}^{\infty} R_{xx}(\tau - s)h(s)ds \\ &= \int_{-\infty}^{\infty} \sigma_w^2\delta(\tau - s)h(s)ds \\ &= \sigma_w^2 h(\tau) \end{aligned} \tag{7.54}$$

*We can measure the impulse response by finding the cross-correlation between the input and output when the input is white noise. The autocorrelation function of the output is*

$$R_{yy}(\tau) = \int_{-\infty}^{\infty} \sigma_w^2 h(\tau + s)h(s)ds = \sigma_w^2 R_{hh}(\tau) \tag{7.55}$$

*The mean-squared value of the output is*

$$R_{yy}(0) = \sigma_w^2 \int_{-\infty}^{\infty} h^2(s)ds = \sigma_w^2 E_h \tag{7.56}$$

*where $E_h$ is the "energy" in the impulse response function. These results indicate how one might construct a random process that has a particular autocorrelation function of interest. Simply select an impulse response that has the right properties and use it to filter white noise.*

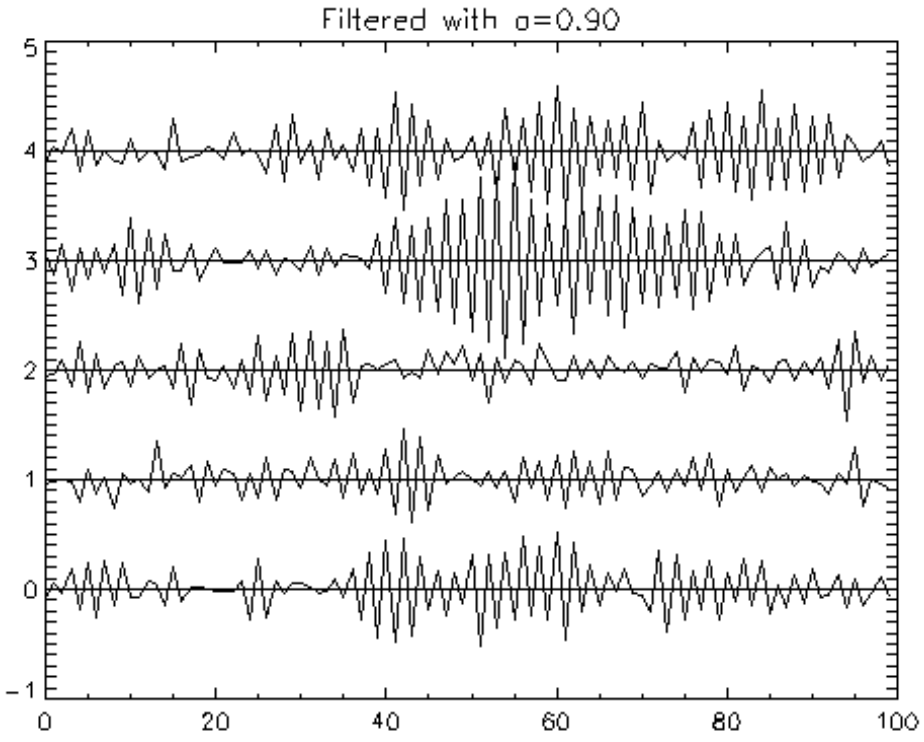


Figure 7.6: Ensemble of outputs of digital filter with white noise input and filter equation $y[n] = x[n] + 0.9y[n-1]$, which represents a low-pass digital filter.

**Example 7.3.4** ***Discrete Low-pass Filter*** *Let $X[n]$ be a discrete random process with autocorrelation function $R_{xx}[m] = E[X[n]X[n+m]] = \sigma_x^2\delta(n-m)$. Consider the random process $Y$ that is represented by the difference equation*

$$Y[n] = X[n] + aY[n-1] \tag{7.57}$$

*where $a$ is a constant. This difference equation represents a digital filter with impulse response*

$$h[n] = a^n step(n) \tag{7.58}$$

*Then the cross-correlation function is given by the difference equation*

$$R_{xy}[m] = E[Y[n]X[n-m]] = \sigma_x^2\delta(m) + aR_{xy}(m-1) \tag{7.59}$$

*This can be solved recursively to find* $R_{xy}[m] = a^m R_{xy}[0]$. *This function is bounded if* $|a| < 1$. *The impulse response of this filter is* $h[m] = a^m step[m]$. *The autocorrelation function of the output is*

$$R_{yy}[m] = \sigma_x^2 \sum_{n=0}^{\infty} a^n a^{n+m} = \frac{\sigma_x^2 a^{|m|}}{1-a^2} \tag{7.60}$$

*where we use* $|m|$ *because* $R_{yy}(-m) = R_{yy}(m)$.

### 7.3.4 Ergodic Random Process

A practical problem arises when we want to calculate parameters such as mean or variance of a random process. The definition would require that we have a large number of examples of the random process and that we calculate the parameters for different values of $t$ by averaging across the ensemble. Often we are faced with the situation of having only one member of the ensemble–that is, one of the time functions. Under what circumstances is it appropriate to use it to draw conclusions about the whole ensemble? It is appropriate to use this approach if and only if the process is ergodic.

A random process is *ergodic* if every member of the process carries with it the complete statistics of the whole process. Then its ensemble averages will equal appropriate time averages. Of necessity, an ergodic process must be stationary, but not all stationary processes are ergodic.

The definition of an ergodic process does not help us much when we need to determine whether or not a particular process is ergodic. We will proceed with some examples.

#### Mean-ergodic Processes

Let $X(t, e)$ be a random process. We know that the mean value, calculated over the ensemble, is $\mu_x(t) = E[X(t, e)]$. If the process is stationary then $\mu_x$ is independent of $t$, a necessary condition for ergodicity. Suppose that we have a particular sample function, say $X(t, e_0)$. We can calculate its time average with an operation such as

$$\tilde{X}_T(e_0) = \frac{1}{2T} \int_{-T}^{T} X(t, e_0) dt \tag{7.61}$$

The answer is a random variable for any value of $T$. The mean value of such measures (computed across the ensemble!) is

$$E[\tilde{X}_T(e)] = \frac{1}{2T}\int_{-T}^{T} E[X(t,e)]dt = \frac{1}{2T}\int_{-T}^{T} \mu_x dt = \mu_x \tag{7.62}$$

We expect (hope) that the variation will decrease with increasing $T$, and the law of large numbers leads us to expect a decrease in the variance in proportion to $1/T$. Do we also expect that we would get the same result with a different sample function, say $X(t, e_1)$? If the function is ergodic then we will measure the same mean value using the time average approach and the measured mean will converge on $\mu_x$. A process with this property is called *mean-ergodic*.

The covariance of $X(t,e)$ is $C(t_1,t_2) = E[X(t_1,e)X(t_2,e)] - \mu_x^2$. The random process is mean-ergodic if and only if[7]

$$\lim_{t\to\infty} \frac{1}{4T^2}\iint_{-T}^{T} C(t_1, C_2)dt_1 dt_2 = 0 \tag{7.63}$$

As a corrolary, a wss process is ergodic if and only if the autocovariance $C(\tau) = R(\tau) - \mu^2$ is such that

$$\lim_{t\to\infty} \frac{1}{2T}\int_{-T}^{T} C(\tau)\left(1 - \frac{|\tau|}{2T}\right) d\tau = 0 \tag{7.64}$$

A sufficient condition for a process to be mean-ergodic is that it be wss and

$$\int_{-\infty}^{\infty} |C(\tau)|\, d\tau < \infty \tag{7.65}$$

A wss random process is mean-ergodic if the random variables $X(t)$ and $X(t+\tau)$ are uncorrelated for large $\tau$. This is a condition that does hold for most physical processes.

**Example 7.3.5** ***Filtered White Noise*** *We have seen in Equation 7.55 that filtered white noise has the autocorrelation function $R_{yy}(\tau) = \sigma_w^2 R_{hh}(\tau)$. Since the mean value is zero, this is also the autocovariance function. If the filter is such that $\int_{-\infty}^{\infty} |R_{hh}(\tau)|\, d\tau < \infty$ then the filter output is mean-ergodic. This is a condition that holds for most filters because the impulse response declines toward zero for a stable system.*

[7] A. Popoulis, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, 1984, p.247.

### Distribution-Ergodic Processes

If a process is mean-ergodic then the mean value can be found as the time average over a single sample function from the random process. But what if we want to measure the probability distribution. How and when might we do that? It turns out that the cumulative distribution function can be found from a sample function by a simple circuit.

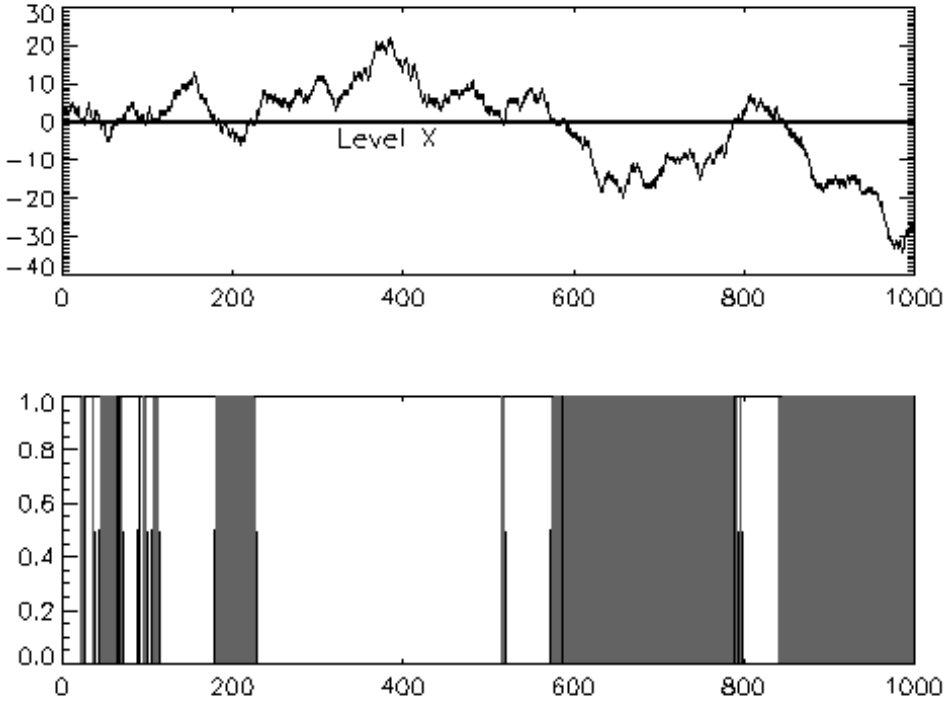


Figure 7.7: The fraction of the time that a random process is lower than a threshold level $x$ is equal to the cumulative distribution function $F(x) = P[X \leq x]$. The fraction of the area that is shaded in the lower figure equals $F(x)$.

Recall that the cumulative distribution function is

$$F(x) = P[X(t) \leq x] \tag{7.66}$$

If the processis stationary, which is a necessary condition for ergodicity, then $F[x]$ is independent of $t$. The above probability is measured across the *ensemble* at any $t$.

Given any sample function $X(t, e_0)$, we can measure the above probability by finding the fraction of the *time* that $X(t)$ is less than $x$. Consider passing $X(t)$ through a system that has the input-output relationship

$$Y_x(t) = \begin{cases} 1, & X(t) \leq x \\ 0, & X(t) > x \end{cases} \tag{7.67}$$

This is a random process in its own right. It has the property that, for any $t$,

$$E[Y_x(t)] = F(x) \tag{7.68}$$

If $Y_x(t)$ is mean-ergodic then we can measure its expected value by computing the time average $\tilde{Y}_x$. The fraction of the area under a sample $Y_x(t)$ waveform will equal the value of $F(x)$. By changing $x$ and measuring the area, we find $F(x)$. This is illustrated in Figure 7.7. A process $X(t)$ is distribution-ergodic if the related process $Y_x(t)$ is mean-ergodic for every value of $x$.

### Correlation-Ergodic Processes

We would often like to find the autocorrelation function of a random process. If $X(t)$ is wss then $R_{xx}(\tau) = E[X(t)X(t+\tau)]$. To compute this average over time, construct the function

$$Z_\tau(t) = X(t)X(t+\tau) \tag{7.69}$$

This is a random process with the property that $R_{xx}(\tau) = E[Z_\tau(t)]$. If $Z_\tau(t)$ is mean-ergodic then we can compute the average over time for any sample function of the random process. Let us denote this average by a script $\mathcal{R}$ to differentiate it from the ensemble average $R$.

$$\mathcal{R}_{xx}(\tau) = \tilde{Z}_\tau = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} X(t)X(t+\tau)dt \tag{7.70}$$

When we say that a stationary random process is ergodic we mean that any average that we may want to find can be computed from any sample function of the process by an appropriate average over time. We will make use of the ergodic assumption when we examine the spectrum of random processes.