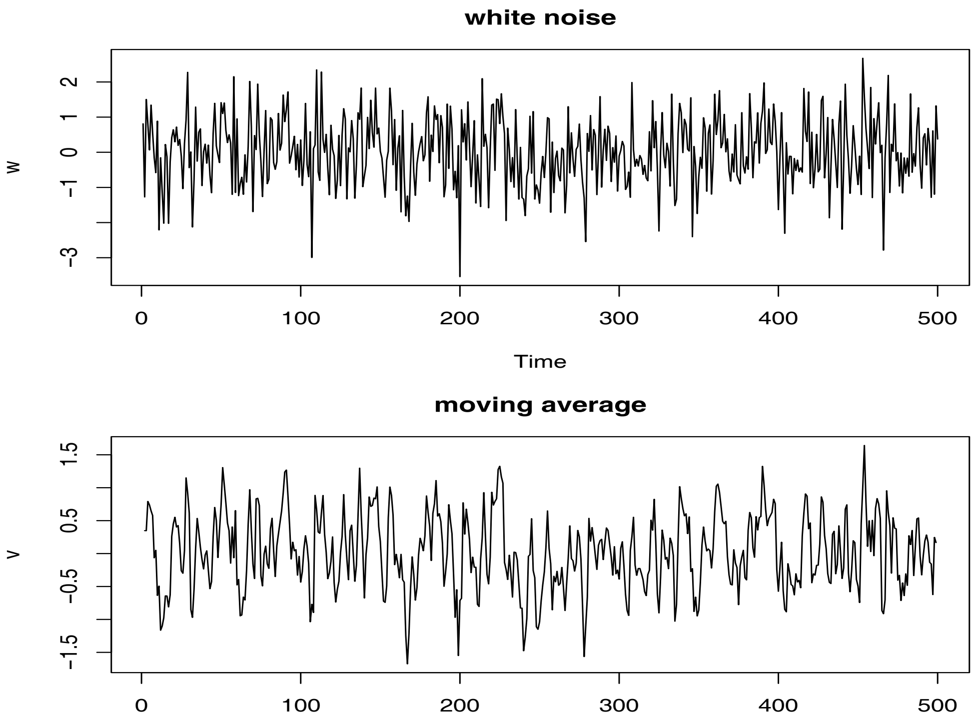

Fig. 1.8. Gaussian white noise series (top) and three-point moving average of the Gaussian white noise series (bottom).

```
w = rnorm(500,0,1)                       # 500 N(0,1) variates
v = filter(w, sides=2, rep(1/3,3))  # moving average
par(mfrow=c(2,1))
plot.ts(w, main="white noise")
plot.ts(v, main="moving average")
```

The speech series in Figure 1.3 and the Recruitment series in Figure 1.5, as well as some of the MRI series in Figure 1.6, differ from the moving average series because one particular kind of oscillatory behavior seems to predominate, producing a sinusoidal type of behavior. A number of methods exist for generating series with this quasi-periodic behavior; we illustrate a popular one based on the autoregressive model considered in Chapter 3.

Example 1.10 Autoregressions

Suppose we consider the white noise series w_t of Example 1.8 as input and calculate the output using the second-order equation

$$x_t = x_{t-1} - .9x_{t-2} + w_t \tag{1.2}$$

successively for $t = 1, 2, \ldots, 500$. Equation (1.2) represents a regression or prediction of the current value x_t of a time series as a function of the past two values of the series, and, hence, the term autoregression is suggested

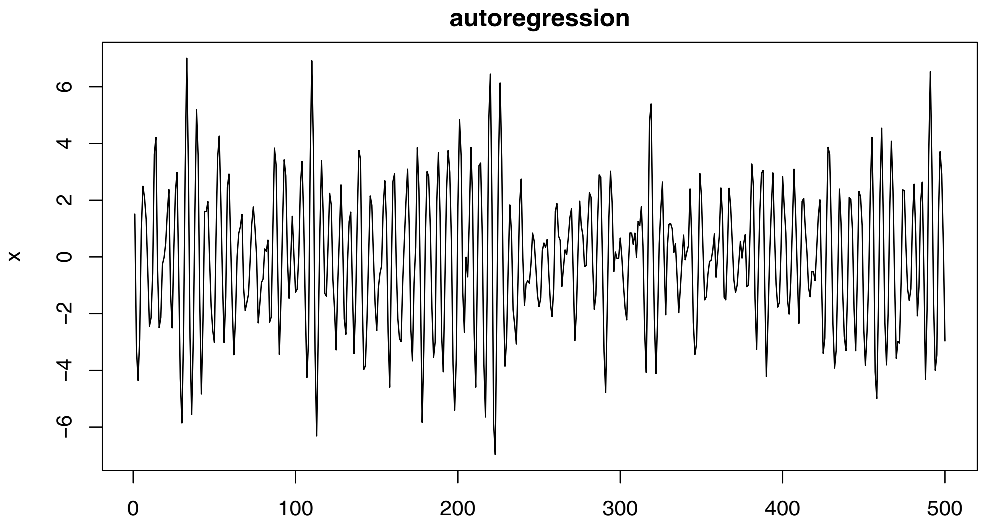

Fig. 1.9. Autoregressive series generated from model (1.2).

for this model. A problem with startup values exists here because (1.2) also depends on the initial conditions x_0 and x_{-1}, but, for now, we assume that we are given these values and generate the succeeding values by substituting into (1.2). The resulting output series is shown in Figure 1.9, and we note the periodic behavior of the series, which is similar to that displayed by the speech series in Figure 1.3. The autoregressive model above and its generalizations can be used as an underlying model for many observed series and will be studied in detail in Chapter 3.

One way to simulate and plot data from the model (1.2) in R is to use the following commands (another way is to use `arima.sim`).

```
w = rnorm(550,0,1)  # 50 extra to avoid startup problems
x = filter(w, filter=c(1,-.9), method="recursive")[-(1:50)]
plot.ts(x, main="autoregression")
```

Example 1.11 Random Walk with Drift

A model for analyzing trend such as seen in the global temperature data in Figure 1.2, is the random walk with drift model given by

$$x_t = \delta + x_{t-1} + w_t \tag{1.3}$$

for $t = 1, 2, \ldots$, with initial condition $x_0 = 0$, and where w_t is white noise. The constant δ is called the drift, and when $\delta = 0$, (1.3) is called simply a random walk. The term random walk comes from the fact that, when $\delta = 0$, the value of the time series at time t is the value of the series at time $t-1$ plus a completely random movement determined by w_t. Note that we may rewrite (1.3) as a cumulative sum of white noise variates. That is,

$$x_t = \delta\, t + \sum_{j=1}^{t} w_j \tag{1.4}$$

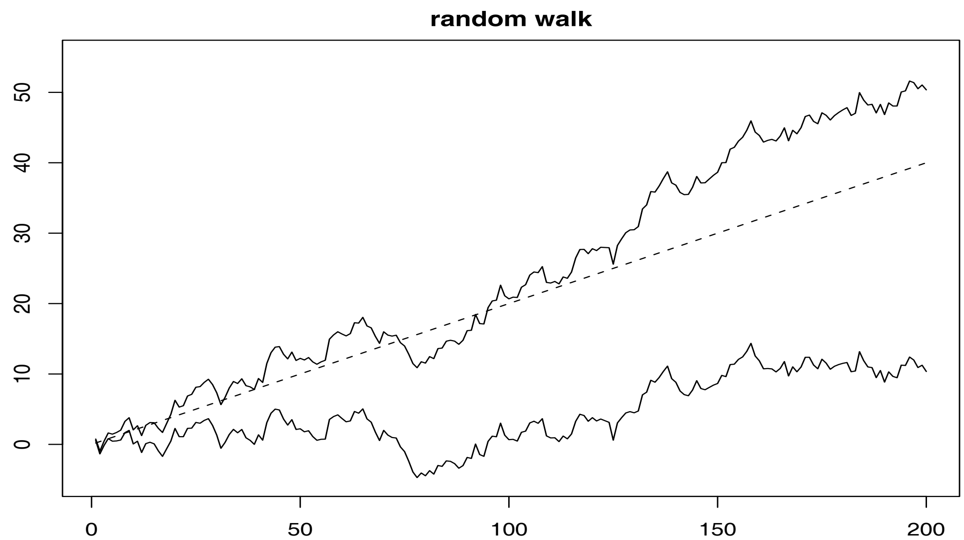

Fig. 1.10. Random walk, $\sigma_w = 1$, with drift $\delta = .2$ (upper jagged line), without drift, $\delta = 0$ (lower jagged line), and a straight line with slope .2 (dashed line).

for $t = 1, 2, \ldots$; either use induction, or plug (1.4) into (1.3) to verify this statement. Figure 1.10 shows 200 observations generated from the model with $\delta = 0$ and .2, and with $\sigma_w = 1$. For comparison, we also superimposed the straight line $.2t$ on the graph.

To reproduce Figure 1.10 in R use the following code (notice the use of multiple commands per line using a semicolon).

```
set.seed(154)                  # so you can reproduce the results
w = rnorm(200,0,1);  x = cumsum(w)    # two commands in one line
wd = w +.2;    xd = cumsum(wd)
plot.ts(xd, ylim=c(-5,55), main="random walk")
lines(x); lines(.2*(1:200), lty="dashed")
```

Example 1.12 Signal in Noise

Many realistic models for generating time series assume an underlying signal with some consistent periodic variation, contaminated by adding a random noise. For example, it is easy to detect the regular cycle fMRI series displayed on the top of Figure 1.6. Consider the model

$$x_t = 2\cos(2\pi t/50 + .6\pi) + w_t \tag{1.5}$$

for $t = 1, 2, \ldots, 500$, where the first term is regarded as the signal, shown in the upper panel of Figure 1.11. We note that a sinusoidal waveform can be written as

$$A\cos(2\pi\omega t + \phi), \tag{1.6}$$

where A is the amplitude, ω is the frequency of oscillation, and ϕ is a phase shift. In (1.5), $A = 2$, $\omega = 1/50$ (one cycle every 50 time points), and $\phi = .6\pi$.

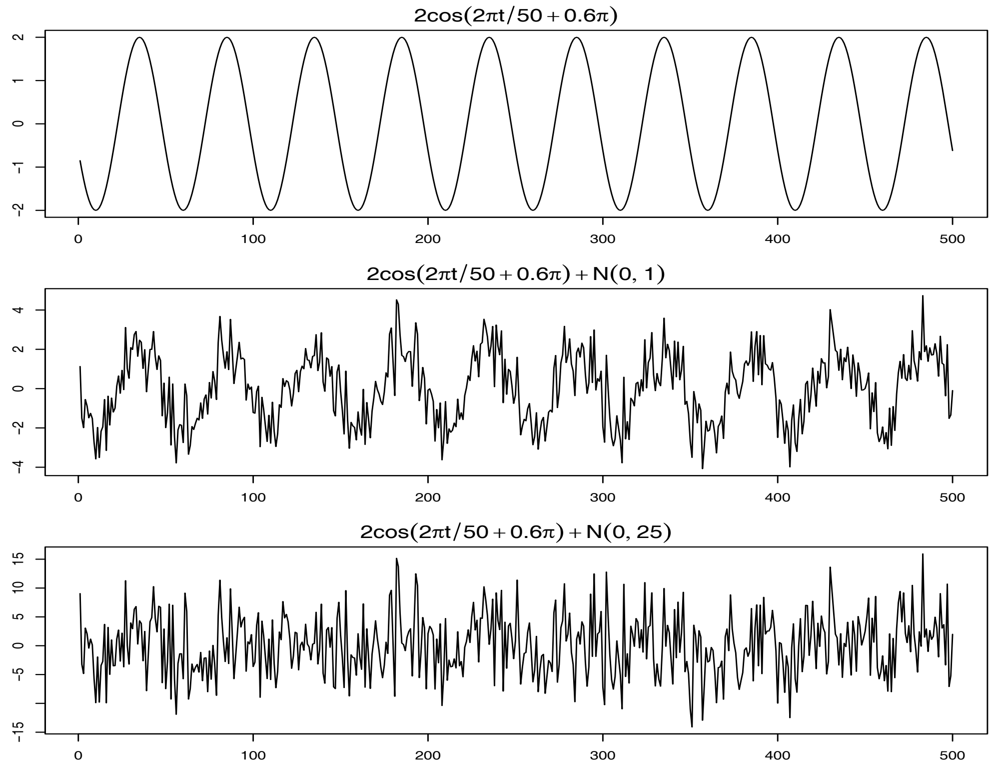

Fig. 1.11. Cosine wave with period 50 points (top panel) compared with the cosine wave contaminated with additive white Gaussian noise, $\sigma_w = 1$ (middle panel) and $\sigma_w = 5$ (bottom panel); see (1.5).

An additive noise term was taken to be white noise with $\sigma_w = 1$ (middle panel) and $\sigma_w = 5$ (bottom panel), drawn from a normal distribution. Adding the two together obscures the signal, as shown in the lower panels of Figure 1.11. Of course, the degree to which the signal is obscured depends on the amplitude of the signal and the size of σ_w. The ratio of the amplitude of the signal to σ_w (or some function of the ratio) is sometimes called the signal-to-noise ratio (SNR); the larger the SNR, the easier it is to detect the signal. Note that the signal is easily discernible in the middle panel of Figure 1.11, whereas the signal is obscured in the bottom panel. Typically, we will not observe the signal but the signal obscured by noise.

To reproduce Figure 1.11 in R, use the following commands:

```
cs = 2*cos(2*pi*1:500/50 + .6*pi)
w = rnorm(500,0,1)
par(mfrow=c(3,1), mar=c(3,2,2,1), cex.main=1.5)
plot.ts(cs, main=expression(2*cos(2*pi*t/50+.6*pi)))
plot.ts(cs+w, main=expression(2*cos(2*pi*t/50+.6*pi) + N(0,1)))
plot.ts(cs+5*w, main=expression(2*cos(2*pi*t/50+.6*pi) + N(0,25)))
```

In Chapter 4, we will study the use of spectral analysis as a possible technique for detecting regular or periodic signals, such as the one described in Example 1.12. In general, we would emphasize the importance of simple additive models such as given above in the form

$$x_t = s_t + v_t, \tag{1.7}$$

where s_t denotes some unknown signal and v_t denotes a time series that may be white or correlated over time. The problems of detecting a signal and then in estimating or extracting the waveform of s_t are of great interest in many areas of engineering and the physical and biological sciences. In economics, the underlying signal may be a trend or it may be a seasonal component of a series. Models such as (1.7), where the signal has an autoregressive structure, form the motivation for the state-space model of Chapter 6.

In the above examples, we have tried to motivate the use of various combinations of random variables emulating real time series data. Smoothness characteristics of observed time series were introduced by combining the random variables in various ways. Averaging independent random variables over adjacent time points, as in Example 1.9, or looking at the output of difference equations that respond to white noise inputs, as in Example 1.10, are common ways of generating correlated data. In the next section, we introduce various theoretical measures used for describing how time series behave. As is usual in statistics, the complete description involves the multivariate distribution function of the jointly sampled values $x_1, x_2, \ldots, x_n$, whereas more economical descriptions can be had in terms of the mean and autocorrelation functions. Because correlation is an essential feature of time series analysis, the most useful descriptive measures are those expressed in terms of covariance and correlation functions.

1.4 Measures of Dependence: Autocorrelation and Cross-Correlation

A complete description of a time series, observed as a collection of n random variables at arbitrary integer time points $t_1, t_2, \ldots, t_n$, for any positive integer n, is provided by the joint distribution function, evaluated as the probability that the values of the series are jointly less than the n constants, $c_1, c_2, \ldots, c_n$; i.e.,

$$F(c_1, c_2, \ldots, c_n) = P\big(x_{t_1} \le c_1, x_{t_2} \le c_2, \ldots, x_{t_n} \le c_n\big). \tag{1.8}$$

Unfortunately, the multidimensional distribution function cannot usually be written easily unless the random variables are jointly normal, in which case the joint density has the well-known form displayed in (1.31).

Although the joint distribution function describes the data completely, it is an unwieldy tool for displaying and analyzing time series data. The distribution function (1.8) must be evaluated as a function of n arguments, so any plotting of the corresponding multivariate density functions is virtually impossible. The marginal distribution functions

$$F_t(x) = P\{x_t \le x\}$$

or the corresponding marginal density functions

$$f_t(x) = \frac{\partial F_t(x)}{\partial x},$$

when they exist, are often informative for examining the marginal behavior of a series.[2] Another informative marginal descriptive measure is the mean function.

Definition 1.1 *The* **mean function** *is defined as*

$$\mu_{xt} = E(x_t) = \int_{-\infty}^{\infty} x f_t(x)\, dx, \tag{1.9}$$

provided it exists, where E denotes the usual expected value operator. When no confusion exists about which time series we are referring to, we will drop a subscript and write μ_{xt} as μ_t.

Example 1.13 Mean Function of a Moving Average Series

If w_t denotes a white noise series, then $\mu_{wt} = E(w_t) = 0$ for all t. The top series in Figure 1.8 reflects this, as the series clearly fluctuates around a mean value of zero. Smoothing the series as in Example 1.9 does not change the mean because we can write

$$\mu_{vt} = E(v_t) = \tfrac{1}{3}[E(w_{t-1}) + E(w_t) + E(w_{t+1})] = 0.$$

Example 1.14 Mean Function of a Random Walk with Drift

Consider the random walk with drift model given in (1.4),

$$x_t = \delta\, t + \sum_{j=1}^{t} w_j, \qquad t = 1, 2, \ldots\ .$$

Because $E(w_t) = 0$ for all t, and δ is a constant, we have

$$\mu_{xt} = E(x_t) = \delta\, t + \sum_{j=1}^{t} E(w_j) = \delta\, t$$

which is a straight line with slope δ. A realization of a random walk with drift can be compared to its mean function in Figure 1.10.

[2] If x_t is Gaussian with mean μ_t and variance σ_t^2, abbreviated as $x_t \sim \mathrm{N}(\mu_t, \sigma_t^2)$, the marginal density is given by $f_t(x) = \frac{1}{\sigma_t\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma_t^2}(x-\mu_t)^2\right\}$.

Example 1.15 Mean Function of Signal Plus Noise

A great many practical applications depend on assuming the observed data have been generated by a fixed signal waveform superimposed on a zero-mean noise process, leading to an additive signal model of the form (1.5). It is clear, because the signal in (1.5) is a fixed function of time, we will have

$$\begin{aligned}\mu_{xt} &= E(x_t) = E\big[2\cos(2\pi t/50 + .6\pi) + w_t\big] \\ &= 2\cos(2\pi t/50 + .6\pi) + E(w_t) \\ &= 2\cos(2\pi t/50 + .6\pi),\end{aligned}$$

and the mean function is just the cosine wave.

The lack of independence between two adjacent values x_s and x_t can be assessed numerically, as in classical statistics, using the notions of covariance and correlation. Assuming the variance of x_t is finite, we have the following definition.

Definition 1.2 *The* **autocovariance function** *is defined as the second moment product*

$$\gamma_x(s,t) = \text{cov}(x_s, x_t) = E[(x_s - \mu_s)(x_t - \mu_t)], \tag{1.10}$$

for all s and t. When no possible confusion exists about which time series we are referring to, we will drop the subscript and write $\gamma_x(s,t)$ as $\gamma(s,t)$.

Note that $\gamma_x(s,t) = \gamma_x(t,s)$ for all time points s and t. The autocovariance measures the *linear* dependence between two points on the same series observed at different times. Very smooth series exhibit autocovariance functions that stay large even when the t and s are far apart, whereas choppy series tend to have autocovariance functions that are nearly zero for large separations. The autocovariance (1.10) is the average cross-product relative to the joint distribution $F(x_s, x_t)$. Recall from classical statistics that if $\gamma_x(s,t) = 0$, x_s and x_t are not linearly related, but there still may be some dependence structure between them. If, however, x_s and x_t are bivariate normal, $\gamma_x(s,t) = 0$ ensures their independence. It is clear that, for $s = t$, the autocovariance reduces to the (assumed finite) variance, because

$$\gamma_x(t,t) = E[(x_t - \mu_t)^2] = \text{var}(x_t). \tag{1.11}$$

Example 1.16 Autocovariance of White Noise

The white noise series w_t has $E(w_t) = 0$ and

$$\gamma_w(s,t) = \text{cov}(w_s, w_t) = \begin{cases} \sigma_w^2 & s = t, \\ 0 & s \neq t. \end{cases} \tag{1.12}$$

A realization of white noise with $\sigma_w^2 = 1$ is shown in the top panel of Figure 1.8.

Example 1.17 Autocovariance of a Moving Average

Consider applying a three-point moving average to the white noise series w_t of the previous example as in Example 1.9. In this case,

$$\gamma_v(s,t) = \text{cov}(v_s, v_t) = \text{cov}\left\{\tfrac{1}{3}(w_{s-1}+w_s+w_{s+1}), \tfrac{1}{3}(w_{t-1}+w_t+w_{t+1})\right\}.$$

When $s = t$ we have[3]

$$\begin{aligned}\gamma_v(t,t) &= \tfrac{1}{9}\text{cov}\{(w_{t-1}+w_t+w_{t+1}), (w_{t-1}+w_t+w_{t+1})\}\\ &= \tfrac{1}{9}[\text{cov}(w_{t-1},w_{t-1}) + \text{cov}(w_t,w_t) + \text{cov}(w_{t+1},w_{t+1})]\\ &= \tfrac{3}{9}\sigma_w^2.\end{aligned}$$

When $s = t+1$,

$$\begin{aligned}\gamma_v(t+1,t) &= \tfrac{1}{9}\text{cov}\{(w_t+w_{t+1}+w_{t+2}), (w_{t-1}+w_t+w_{t+1})\}\\ &= \tfrac{1}{9}[\text{cov}(w_t,w_t) + \text{cov}(w_{t+1},w_{t+1})]\\ &= \tfrac{2}{9}\sigma_w^2,\end{aligned}$$

using (1.12). Similar computations give $\gamma_v(t-1,t) = 2\sigma_w^2/9$, $\gamma_v(t+2,t) = \gamma_v(t-2,t) = \sigma_w^2/9$, and 0 when $|t-s| > 2$. We summarize the values for all s and t as

$$\gamma_v(s,t) = \begin{cases} \frac{3}{9}\sigma_w^2 & s = t,\\ \frac{2}{9}\sigma_w^2 & |s-t| = 1,\\ \frac{1}{9}\sigma_w^2 & |s-t| = 2,\\ 0 & |s-t| > 2. \end{cases} \tag{1.13}$$

Example 1.17 shows clearly that the smoothing operation introduces a covariance function that decreases as the separation between the two time points increases and disappears completely when the time points are separated by three or more time points. This particular autocovariance is interesting because it only depends on the time separation or lag and not on the absolute location of the points along the series. We shall see later that this dependence suggests a mathematical model for the concept of weak stationarity.

Example 1.18 Autocovariance of a Random Walk

For the random walk model, $x_t = \sum_{j=1}^t w_j$, we have

$$\gamma_x(s,t) = \text{cov}(x_s, x_t) = \text{cov}\left(\sum_{j=1}^s w_j, \sum_{k=1}^t w_k\right) = \min\{s,t\}\,\sigma_w^2,$$

because the w_t are uncorrelated random variables. Note that, as opposed to the previous examples, the autocovariance function of a random walk

[3] If the random variables $U = \sum_{j=1}^m a_j X_j$ and $V = \sum_{k=1}^r b_k Y_k$ are linear combinations of random variables $\{X_j\}$ and $\{Y_k\}$, respectively, then $\text{cov}(U,V) = \sum_{j=1}^m \sum_{k=1}^r a_j b_k \text{cov}(X_j, Y_k)$. Furthermore, $\text{var}(U) = \text{cov}(U,U)$.

depends on the particular time values s and t, and not on the time separation or lag. Also, notice that the variance of the random walk, $\text{var}(x_t) = \gamma_x(t,t) = t\,\sigma_w^2$, increases without bound as time t increases. The effect of this variance increase can be seen in Figure 1.10 where the processes start to move away from their mean functions $\delta\,t$ (note that $\delta = 0$ and .2 in that example).

As in classical statistics, it is more convenient to deal with a measure of association between -1 and 1, and this leads to the following definition.

Definition 1.3 *The* **autocorrelation function (ACF)** *is defined as*

$$\rho(s,t) = \frac{\gamma(s,t)}{\sqrt{\gamma(s,s)\gamma(t,t)}}. \tag{1.14}$$

The ACF measures the linear predictability of the series at time t, say x_t, using only the value x_s. We can show easily that $-1 \le \rho(s,t) \le 1$ using the Cauchy–Schwarz inequality.[4] If we can predict x_t *perfectly* from x_s through a linear relationship, $x_t = \beta_0 + \beta_1 x_s$, then the correlation will be $+1$ when $\beta_1 > 0$, and -1 when $\beta_1 < 0$. Hence, we have a rough measure of the ability to forecast the series at time t from the value at time s.

Often, we would like to measure the predictability of another series y_t from the series x_s. Assuming both series have finite variances, we have the following definition.

Definition 1.4 *The* **cross-covariance function** *between two series, x_t and y_t, is*

$$\gamma_{xy}(s,t) = \text{cov}(x_s, y_t) = E[(x_s - \mu_{xs})(y_t - \mu_{yt})]. \tag{1.15}$$

There is also a scaled version of the cross-covariance function.

Definition 1.5 *The* **cross-correlation function (CCF)** *is given by*

$$\rho_{xy}(s,t) = \frac{\gamma_{xy}(s,t)}{\sqrt{\gamma_x(s,s)\gamma_y(t,t)}}. \tag{1.16}$$

We may easily extend the above ideas to the case of more than two series, say, $x_{t1}, x_{t2}, \ldots, x_{tr}$; that is, multivariate time series with r components. For example, the extension of (1.10) in this case is

$$\gamma_{jk}(s,t) = E[(x_{sj} - \mu_{sj})(x_{tk} - \mu_{tk})] \qquad j,k = 1,2,\ldots,r. \tag{1.17}$$

In the definitions above, the autocovariance and cross-covariance functions may change as one moves along the series because the values depend on both s

[4] The Cauchy–Schwarz inequality implies $|\gamma(s,t)|^2 \le \gamma(s,s)\gamma(t,t)$.

and t, the locations of the points in time. In Example 1.17, the autocovariance function depends on the separation of x_s and x_t, say, $h = |s - t|$, and not on where the points are located in time. As long as the points are separated by h units, the location of the two points does not matter. This notion, called weak stationarity, when the mean is constant, is fundamental in allowing us to analyze sample time series data when only a single series is available.

1.5 Stationary Time Series

The preceding definitions of the mean and autocovariance functions are completely general. Although we have not made any special assumptions about the behavior of the time series, many of the preceding examples have hinted that a sort of regularity may exist over time in the behavior of a time series. We introduce the notion of regularity using a concept called stationarity.

Definition 1.6 *A* **strictly stationary** *time series is one for which the probabilistic behavior of every collection of values*

$$\{x_{t_1}, x_{t_2}, \ldots, x_{t_k}\}$$

is identical to that of the time shifted set

$$\{x_{t_1+h}, x_{t_2+h}, \ldots, x_{t_k+h}\}.$$

That is,

$$P\{x_{t_1} \le c_1, \ldots, x_{t_k} \le c_k\} = P\{x_{t_1+h} \le c_1, \ldots, x_{t_k+h} \le c_k\} \tag{1.18}$$

for all $k = 1, 2, \ldots$, all time points $t_1, t_2, \ldots, t_k$, all numbers $c_1, c_2, \ldots, c_k$, and all time shifts $h = 0, \pm 1, \pm 2, \ldots$.

If a time series is strictly stationary, then all of the multivariate distribution functions for subsets of variables must agree with their counterparts in the shifted set for all values of the shift parameter h. For example, when $k = 1$, (1.18) implies that

$$P\{x_s \le c\} = P\{x_t \le c\} \tag{1.19}$$

for any time points s and t. This statement implies, for example, that the probability that the value of a time series sampled hourly is negative at 1 AM is the same as at 10 AM. In addition, if the mean function, μ_t, of the series x_t exists, (1.19) implies that $\mu_s = \mu_t$ for all s and t, and hence μ_t must be constant. Note, for example, that a random walk process with drift is *not* strictly stationary because its mean function changes with time; see Example 1.14 on page 18.

When $k = 2$, we can write (1.18) as

$$P\{x_s \le c_1, x_t \le c_2\} = P\{x_{s+h} \le c_1, x_{t+h} \le c_2\} \tag{1.20}$$

for any time points s and t and shift h. Thus, if the variance function of the process exists, (1.20) implies that the autocovariance function of the series x_t satisfies

$$\gamma(s,t) = \gamma(s+h, t+h)$$

for all s and t and h. We may interpret this result by saying the autocovariance function of the process depends only on the time difference between s and t, and not on the actual times.

The version of stationarity in Definition 1.6 is too strong for most applications. Moreover, it is difficult to assess strict stationarity from a single data set. Rather than imposing conditions on all possible distributions of a time series, we will use a milder version that imposes conditions only on the first two moments of the series. We now have the following definition.

Definition 1.7 *A* **weakly stationary** *time series, x_t, is a finite variance process such that*

(i) the mean value function, μ_t, defined in (1.9) is constant and does not depend on time t, and
(ii) the autocovariance function, $\gamma(s,t)$, defined in (1.10) depends on s and t only through their difference $|s-t|$.

Henceforth, we will use the term **stationary** *to mean weakly stationary; if a process is stationary in the strict sense, we will use the term strictly stationary.*

It should be clear from the discussion of strict stationarity following Definition 1.6 that a strictly stationary, finite variance, time series is also stationary. The converse is not true unless there are further conditions. One important case where stationarity implies strict stationarity is if the time series is Gaussian [meaning all finite distributions, (1.18), of the series are Gaussian]. We will make this concept more precise at the end of this section.

Because the mean function, $E(x_t) = \mu_t$, of a stationary time series is independent of time t, we will write

$$\mu_t = \mu. \tag{1.21}$$

Also, because the autocovariance function, $\gamma(s,t)$, of a stationary time series, x_t, depends on s and t only through their difference $|s-t|$, we may simplify the notation. Let $s = t+h$, where h represents the time shift or lag. Then

$$\gamma(t+h, t) = \text{cov}(x_{t+h}, x_t) = \text{cov}(x_h, x_0) = \gamma(h, 0)$$

because the time difference between times $t+h$ and t is the same as the time difference between times h and 0. Thus, the autocovariance function of a stationary time series does not depend on the time argument t. Henceforth, for convenience, we will drop the second argument of $\gamma(h,0)$.

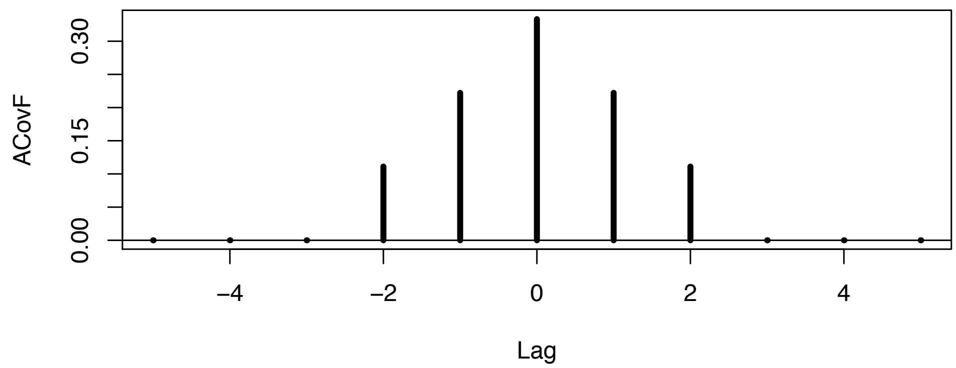

Fig. 1.12. Autocovariance function of a three-point moving average.

Definition 1.8 *The* **autocovariance function of a stationary time series** *will be written as*

$$\gamma(h) = \text{cov}(x_{t+h}, x_t) = E[(x_{t+h} - \mu)(x_t - \mu)]. \tag{1.22}$$

Definition 1.9 *The* **autocorrelation function (ACF) of a stationary time series** *will be written using* (1.14) *as*

$$\rho(h) = \frac{\gamma(t+h, t)}{\sqrt{\gamma(t+h, t+h)\gamma(t,t)}} = \frac{\gamma(h)}{\gamma(0)}. \tag{1.23}$$

The Cauchy–Schwarz inequality shows again that $-1 \leq \rho(h) \leq 1$ for all h, enabling one to assess the relative importance of a given autocorrelation value by comparing with the extreme values -1 and 1.

Example 1.19 Stationarity of White Noise

The mean and autocovariance functions of the white noise series discussed in Example 1.8 and Example 1.16 are easily evaluated as $\mu_{wt} = 0$ and

$$\gamma_w(h) = \text{cov}(w_{t+h}, w_t) = \begin{cases} \sigma_w^2 & h = 0, \\ 0 & h \neq 0. \end{cases}$$

Thus, white noise satisfies the conditions of Definition 1.7 and is weakly stationary or stationary. If the white noise variates are also normally distributed or Gaussian, the series is also strictly stationary, as can be seen by evaluating (1.18) using the fact that the noise would also be iid.

Example 1.20 Stationarity of a Moving Average

The three-point moving average process of Example 1.9 is stationary because, from Example 1.13 and Example 1.17, the mean and autocovariance functions $\mu_{vt} = 0$, and

$$\gamma_v(h) = \begin{cases} \frac{3}{9}\sigma_w^2 & h = 0, \\ \frac{2}{9}\sigma_w^2 & h = \pm 1, \\ \frac{1}{9}\sigma_w^2 & h = \pm 2, \\ 0 & |h| > 2 \end{cases}$$

are independent of time t, satisfying the conditions of Definition 1.7. Figure 1.12 shows a plot of the autocovariance as a function of lag h with $\sigma_w^2 = 1$. Interestingly, the autocovariance function is symmetric about lag zero and decays as a function of lag.

The autocovariance function of a stationary process has several useful properties (also, see Problem 1.25). First, the value at $h = 0$, namely

$$\gamma(0) = E[(x_t - \mu)^2] \tag{1.24}$$

is the variance of the time series; note that the Cauchy–Schwarz inequality implies

$$|\gamma(h)| \leq \gamma(0).$$

A final useful property, noted in the previous example, is that the autocovariance function of a stationary series is symmetric around the origin; that is,

$$\gamma(h) = \gamma(-h) \tag{1.25}$$

for all h. This property follows because shifting the series by h means that

$$\begin{aligned} \gamma(h) &= \gamma(t + h - t) \\ &= E[(x_{t+h} - \mu)(x_t - \mu)] \\ &= E[(x_t - \mu)(x_{t+h} - \mu)] \\ &= \gamma(t - (t + h)) \\ &= \gamma(-h), \end{aligned}$$

which shows how to use the notation as well as proving the result.

When several series are available, a notion of stationarity still applies with additional conditions.

Definition 1.10 *Two time series, say, x_t and y_t, are said to be* **jointly stationary** *if they are each stationary, and the cross-covariance function*

$$\gamma_{xy}(h) = \text{cov}(x_{t+h}, y_t) = E[(x_{t+h} - \mu_x)(y_t - \mu_y)] \tag{1.26}$$

is a function only of lag h.

Definition 1.11 *The* **cross-correlation function (CCF)** *of jointly stationary time series x_t and y_t is defined as*

$$\rho_{xy}(h) = \frac{\gamma_{xy}(h)}{\sqrt{\gamma_x(0)\gamma_y(0)}}. \tag{1.27}$$

Again, we have the result $-1 \leq \rho_{xy}(h) \leq 1$ which enables comparison with the extreme values -1 and 1 when looking at the relation between x_{t+h} and y_t. The cross-correlation function is not generally symmetric about zero [i.e., typically $\rho_{xy}(h) \neq \rho_{xy}(-h)$]; however, it is the case that

$$\rho_{xy}(h) = \rho_{yx}(-h), \tag{1.28}$$

which can be shown by manipulations similar to those used to show (1.25).

Example 1.21 Joint Stationarity

Consider the two series, x_t and y_t, formed from the sum and difference of two successive values of a white noise process, say,

$$x_t = w_t + w_{t-1}$$

and

$$y_t = w_t - w_{t-1},$$

where w_t are independent random variables with zero means and variance σ_w^2. It is easy to show that $\gamma_x(0) = \gamma_y(0) = 2\sigma_w^2$ and $\gamma_x(1) = \gamma_x(-1) = \sigma_w^2, \gamma_y(1) = \gamma_y(-1) = -\sigma_w^2$. Also,

$$\gamma_{xy}(1) = \text{cov}(x_{t+1}, y_t) = \text{cov}(w_{t+1} + w_t, w_t - w_{t-1}) = \sigma_w^2$$

because only one term is nonzero (recall footnote 3 on page 20). Similarly, $\gamma_{xy}(0) = 0, \gamma_{xy}(-1) = -\sigma_w^2$. We obtain, using (1.11),

$$\rho_{xy}(h) = \begin{cases} 0 & h = 0, \\ 1/2 & h = 1, \\ -1/2 & h = -1, \\ 0 & |h| \geq 2. \end{cases}$$

Clearly, the autocovariance and cross-covariance functions depend only on the lag separation, h, so the series are jointly stationary.

Example 1.22 Prediction Using Cross-Correlation

As a simple example of cross-correlation, consider the problem of determining possible leading or lagging relations between two series x_t and y_t. If the model

$$y_t = Ax_{t-\ell} + w_t$$

holds, the series x_t is said to lead y_t for $\ell > 0$ and is said to lag y_t for $\ell < 0$. Hence, the analysis of leading and lagging relations might be important in predicting the value of y_t from x_t. Assuming, for convenience, that x_t and y_t have zero means, and the noise w_t is uncorrelated with the x_t series, the cross-covariance function can be computed as

$$\gamma_{yx}(h) = \text{cov}(y_{t+h}, x_t) = \text{cov}(Ax_{t+h-\ell} + w_{t+h}, x_t)$$
$$= \text{cov}(Ax_{t+h-\ell}, x_t) = A\gamma_x(h-\ell).$$

The cross-covariance function will look like the autocovariance of the input series x_t, with a peak on the positive side if x_t leads y_t and a peak on the negative side if x_t lags y_t.

The concept of weak stationarity forms the basis for much of the analysis performed with time series. The fundamental properties of the mean and autocovariance functions (1.21) and (1.22) are satisfied by many theoretical models that appear to generate plausible sample realizations. In Example 1.9 and Example 1.10, two series were generated that produced stationary looking realizations, and in Example 1.20, we showed that the series in Example 1.9 was, in fact, weakly stationary. Both examples are special cases of the so-called linear process.

Definition 1.12 *A* **linear process**, *x_t, is defined to be a linear combination of white noise variates w_t, and is given by*

$$x_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j w_{t-j}, \qquad \sum_{j=-\infty}^{\infty} |\psi_j| < \infty. \tag{1.29}$$

For the linear process (see Problem 1.11), we may show that the autocovariance function is given by

$$\gamma(h) = \sigma_w^2 \sum_{j=-\infty}^{\infty} \psi_{j+h}\psi_j \tag{1.30}$$

for $h \geq 0$; recall that $\gamma(-h) = \gamma(h)$. This method exhibits the autocovariance function of the process in terms of the lagged products of the coefficients. Note that, for Example 1.9, we have $\psi_0 = \psi_{-1} = \psi_1 = 1/3$ and the result in Example 1.20 comes out immediately. The autoregressive series in Example 1.10 can also be put in this form, as can the general autoregressive moving average processes considered in Chapter 3.

Finally, as previously mentioned, an important case in which a weakly stationary series is also strictly stationary is the normal or Gaussian series.

Definition 1.13 *A process, $\{x_t\}$, is said to be a* **Gaussian process** *if the n-dimensional vectors $\boldsymbol{x} = (x_{t_1}, x_{t_2}, \ldots, x_{t_n})'$, for every collection of time points $t_1, t_2, \ldots, t_n$, and every positive integer n, have a multivariate normal distribution.*

Defining the $n \times 1$ mean vector $E(\boldsymbol{x}) \equiv \boldsymbol{\mu} = (\mu_{t_1}, \mu_{t_2}, \ldots, \mu_{t_n})'$ and the $n \times n$ covariance matrix as $\text{var}(\boldsymbol{x}) \equiv \Gamma = \{\gamma(t_i, t_j);\ i, j = 1, \ldots, n\}$, which is assumed to be positive definite, the multivariate normal density function can be written as