for $x \in \mathbb{R}^n$, where $|\cdot|$ denotes the determinant.

We list some important items regarding linear and Gaussian processes.

- If a Gaussian time series, $\{x_t\}$, is weakly stationary, then μ_t is constant and $\gamma(t_i, t_j) = \gamma(|t_i - t_j|)$, so that the vector μ and the matrix Γ are independent of time. These facts imply that all the finite distributions, (1.33), of the series $\{x_t\}$ depend only on time lag and not on the actual times, and hence the series must be strictly stationary. In a sense, weak stationarity and normality go hand-in-hand in that we will base our analyses on the idea that it is enough for the first two moments to behave nicely. We use the multivariate normal density in the form given above as well as in a modified version, applicable to complex random variables throughout the text.
- A result called the *Wold Decomposition* (Theorem B.5) states that a stationary non-deterministic time series is a causal linear process (but with $\sum \psi_j^2 < \infty$). A linear process need not be Gaussian, but if a time series is Gaussian, then it is a causal linear process with $w_t \sim \text{iid } \mathrm{N}(0, \sigma_w^2)$. Hence, stationary Gaussian processes form the basis of modeling many time series.
- It is not enough for the marginal distributions to be Gaussian for the process to be Gaussian. It is easy to construct a situation where X and Y are normal, but (X, Y) is not bivariate normal; e.g., let X and Z be independent normals and let $Y = Z$ if $XZ > 0$ and $Y = -Z$ if $XZ \leq 0$.

1.5 Estimation of Correlation

Although the theoretical autocorrelation and cross-correlation functions are useful for describing the properties of certain hypothesized models, most of the analyses must be performed using sampled data. This limitation means the sampled points $x_1, x_2, \ldots, x_n$ only are available for estimating the mean, autocovariance, and autocorrelation functions. From the point of view of classical statistics, this poses a problem because we will typically not have iid copies of x_t that are available for estimating the covariance and correlation functions. In the usual situation with only one realization, however, the assumption of stationarity becomes critical. Somehow, we must use averages over this single realization to estimate the population means and covariance functions.

Accordingly, if a time series is stationary, the mean function (1.22) $\mu_t = \mu$ is constant so that we can estimate it by the *sample mean*,

$$\bar{x} = \frac{1}{n} \sum_{t=1}^{n} x_t. \tag{1.34}$$

In our case, $\mathrm{E}(\bar{x}) = \mu$, and the standard error of the estimate is the square root of $\mathrm{var}(\bar{x})$, which can be computed using first principles (recall Property 1.1), and is given by

$$\begin{aligned}\operatorname{var}(\bar{x}) &= \operatorname{var}\left(\frac{1}{n}\sum_{t=1}^{n} x_t\right) = \frac{1}{n^2}\operatorname{cov}\left(\sum_{t=1}^{n} x_t, \sum_{s=1}^{n} x_s\right)\\ &= \frac{1}{n^2}\Big(n\gamma_x(0) + (n-1)\gamma_x(1) + (n-2)\gamma_x(2) + \cdots + \gamma_x(n-1)\\ &\qquad + (n-1)\gamma_x(-1) + (n-2)\gamma_x(-2) + \cdots + \gamma_x(1-n)\Big)\\ &= \frac{1}{n}\sum_{h=-n}^{n}\left(1 - \frac{|h|}{n}\right)\gamma_x(h). \end{aligned} \tag{1.35}$$

If the process is white noise, (1.35) reduces to the familiar σ_x^2/n recalling that $\gamma_x(0) = \sigma_x^2$. Note that, in the case of dependence, the standard error of $\bar{x}$ may be smaller or larger than the white noise case depending on the nature of the correlation structure (see Problem 1.19)

The theoretical autocovariance function, (1.23), is estimated by the sample autocovariance function defined as follows.

Definition 1.14 *The* **sample autocovariance function** *is defined as*

$$\hat{\gamma}(h) = n^{-1}\sum_{t=1}^{n-h}(x_{t+h} - \bar{x})(x_t - \bar{x}), \tag{1.36}$$

with $\hat{\gamma}(-h) = \hat{\gamma}(h)$ *for* $h = 0, 1, \ldots, n-1$.

The sum in (1.36) runs over a restricted range because x_{t+h} is not available for $t + h > n$. The estimator in (1.36) is preferred to the one that would be obtained by dividing by $n - h$ because (1.36) is a non-negative definite function. Recall that the autocovariance function of a stationary process is non-negative definite [(1.25); also, see Problem 1.25] ensuring that variances of linear combinations of the variates x_t will never be negative. And because a variance is never negative, the estimate of that variance

$$\widehat{\operatorname{var}}(a_1x_1 + \cdots + a_nx_n) = \sum_{j=1}^{n}\sum_{k=1}^{n} a_j a_k \hat{\gamma}(j-k),$$

should also be non-negative. The estimator in (1.36) guarantees this result, but no such guarantee exists if we divide by $n - h$. Note that neither dividing by n nor $n - h$ in (1.36) yields an unbiased estimator of $\gamma(h)$.

Definition 1.15 *The* **sample autocorrelation function** *is defined, analogously to* (1.24)*, as*

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}. \tag{1.37}$$

The sample autocorrelation function has a sampling distribution that allows us to assess whether the data comes from a completely random or white series or whether correlations are statistically significant at some lags.

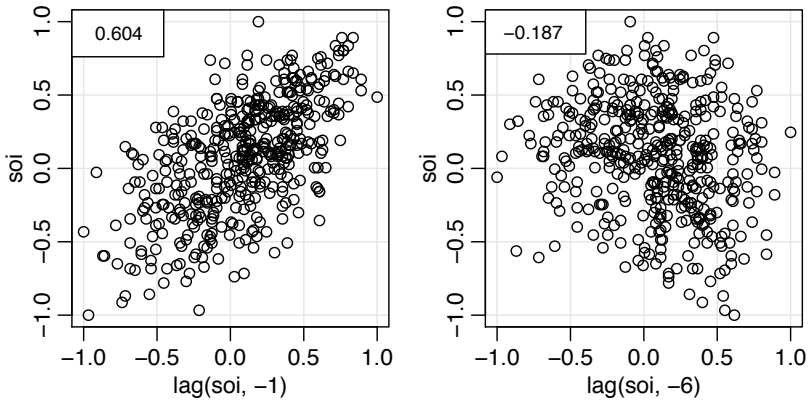

Fig. 1.14. *Display for Example 1.25. For the SOI series, the scatterplots show pairs of values one month apart (left) and six months apart (right). The estimated correlation is displayed in the box.*

Example 1.25 Sample ACF and Scatterplots

Estimating autocorrelation is similar to estimating of correlation in the usual setup where we have pairs of observations, say (x_i, y_i), for $i = 1, \ldots, n$. For example, if we have time series data x_t for $t = 1, \ldots, n$, then the pairs of observations for estimating $\rho(h)$ are the $n - h$ pairs given by $\{(x_t, x_{t+h}); t = 1, \ldots, n - h\}$. Figure 1.14 shows an example using the SOI series where $\hat{\rho}(1) = .604$ and $\hat{\rho}(6) = -.187$. The following code was used for Figure 1.14.

```
(r = round(acf(soi, 6, plot=FALSE)$acf[-1], 3))  # first 6 sample acf values
  [1]  0.604  0.374  0.214  0.050 -0.107 -0.187
par(mfrow=c(1,2))
plot(lag(soi,-1), soi); legend('topleft', legend=r[1])
plot(lag(soi,-6), soi); legend('topleft', legend=r[6])
```

Property 1.2 Large-Sample Distribution of the ACF

Under general conditions,[1.5] *if x_t is white noise, then for n large, the sample ACF, $\hat{\rho}_x(h)$, for $h = 1, 2, \ldots, H$, where H is fixed but arbitrary, is approximately normally distributed with zero mean and standard deviation given by*

$$\sigma_{\hat{\rho}_x(h)} = \frac{1}{\sqrt{n}}. \tag{1.38}$$

Based on the previous result, we obtain a rough method of assessing whether peaks in $\hat{\rho}(h)$ are significant by determining whether the observed peak is outside the interval $\pm 2/\sqrt{n}$ (or plus/minus two standard errors); for a white noise sequence, approximately 95% of the sample ACFs should be within these limits. The applications of this property develop because many statistical modeling procedures depend on reducing a time series to a white noise series using various kinds of transformations. After such a procedure is applied, the plotted ACFs of the residuals should then lie roughly within the limits given above.

[1.5] The general conditions are that x_t is iid with finite fourth moment. A sufficient condition for this to hold is that x_t is white Gaussian noise. Precise details are given in Theorem A.7 in Appendix A.

Example 1.26 A Simulated Time Series

To compare the sample ACF for various sample sizes to the theoretical ACF, consider a contrived set of data generated by tossing a fair coin, letting $x_t = 1$ when a head is obtained and $x_t = -1$ when a tail is obtained. Then, construct y_t as

$$y_t = 5 + x_t - .7x_{t-1}. \tag{1.39}$$

To simulate data, we consider two cases, one with a small sample size ($n = 10$) and another with a moderate sample size ($n = 100$).

```
set.seed(101010)
x1 = 2*rbinom(11, 1, .5) - 1    # simulated sequence of coin tosses
x2 = 2*rbinom(101, 1, .5) - 1
y1 = 5 + filter(x1, sides=1, filter=c(1,-.7))[-1]
y2 = 5 + filter(x2, sides=1, filter=c(1,-.7))[-1]
plot.ts(y1, type='s'); plot.ts(y2, type='s')   # plot both series (not shown)
c(mean(y1), mean(y2))           # the sample means
   [1] 5.080   5.002
acf(y1, lag.max=4, plot=FALSE)  # 1/sqrt(10) = .32
   Autocorrelations of series 'y1', by lag
       0      1      2      3      4
   1.000 -0.688  0.425 -0.306 -0.007
acf(y2, lag.max=4, plot=FALSE)  # 1/sqrt(100) = .1
   Autocorrelations of series 'y2', by lag
       0      1      2      3      4
   1.000 -0.480 -0.002 -0.004  0.000
# Note that the sample ACF at lag zero is always 1 (Why?).
```

The theoretical ACF can be obtained from the model (1.39) using the fact that the mean of x_t is zero and the variance of x_t is one. It can be shown that

$$\rho_y(1) = \frac{-.7}{1 + .7^2} = -.47$$

and $\rho_y(h) = 0$ for $|h| > 1$ (Problem 1.24). It is interesting to compare the theoretical ACF with sample ACFs for the realization where $n = 10$ and the other realization where $n = 100$; note the increased variability in the smaller size sample.

Example 1.27 ACF of a Speech Signal

Computing the sample ACF as in the previous example can be thought of as matching the time series h units in the future, say, x_{t+h} against itself, x_t. Figure 1.15 shows the ACF of the speech series of Figure 1.3. The original series appears to contain a sequence of repeating short signals. The ACF confirms this behavior, showing repeating peaks spaced at about 106-109 points. Autocorrelation functions of the short signals appear, spaced at the intervals mentioned above. The distance between the repeating signals is known as the *pitch period* and is a fundamental parameter of interest in systems that encode and decipher speech. Because the series is sampled at 10,000 points per second, the pitch period appears to be between .0106 and .0109 seconds. To compute the sample ACF in R, use `acf(speech, 250)`.

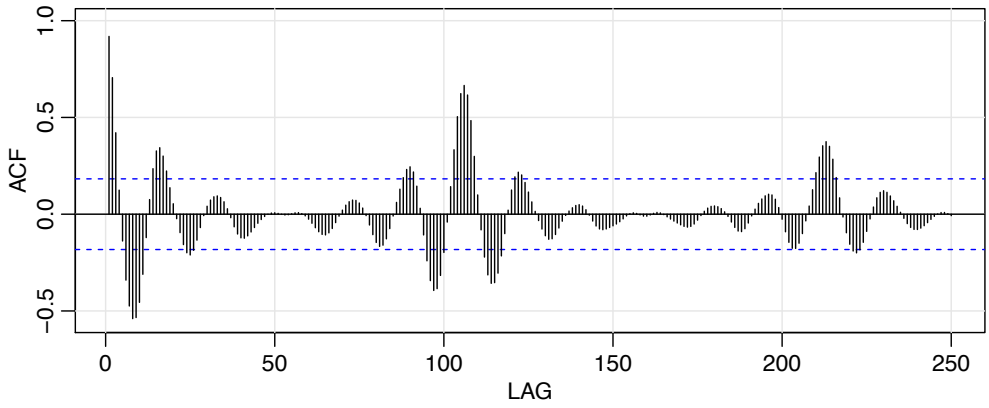

Fig. 1.15. *ACF of the speech series.*

Definition 1.16 *The estimators for the cross-covariance function, $\gamma_{xy}(h)$, as given in (1.28) and the cross-correlation, $\rho_{xy}(h)$, in (1.11) are given, respectively, by the* **sample cross-covariance function**

$$\hat{\gamma}_{xy}(h) = n^{-1} \sum_{t=1}^{n-h} (x_{t+h} - \bar{x})(y_t - \bar{y}), \tag{1.40}$$

where $\hat{\gamma}_{xy}(-h) = \hat{\gamma}_{yx}(h)$ determines the function for negative lags, and the **sample cross-correlation function**

$$\hat{\rho}_{xy}(h) = \frac{\hat{\gamma}_{xy}(h)}{\sqrt{\hat{\gamma}_x(0)\hat{\gamma}_y(0)}}. \tag{1.41}$$

The sample cross-correlation function can be examined graphically as a function of lag h to search for leading or lagging relations in the data using the property mentioned in Example 1.24 for the theoretical cross-covariance function. Because $-1 \leq \hat{\rho}_{xy}(h) \leq 1$, the practical importance of peaks can be assessed by comparing their magnitudes with their theoretical maximum values. Furthermore, for x_t and y_t independent linear processes of the form (1.31), we have the following property.

Property 1.3 Large-Sample Distribution of Cross-Correlation

The large sample distribution of $\hat{\rho}_{xy}(h)$ is normal with mean zero and

$$\sigma_{\hat{\rho}_{xy}} = \frac{1}{\sqrt{n}} \tag{1.42}$$

if at least one of the processes is independent white noise (see Theorem A.8).

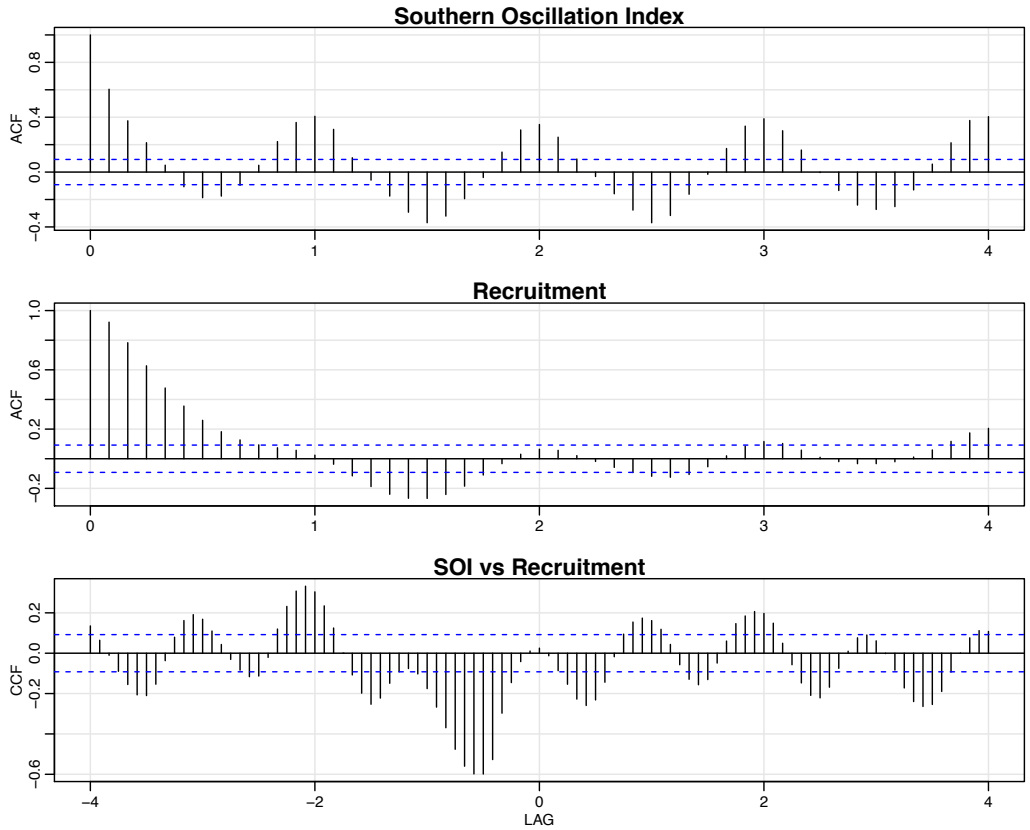

Fig. 1.16. *Sample ACFs of the SOI series (top) and of the Recruitment series (middle), and the sample CCF of the two series (bottom); negative lags indicate SOI leads Recruitment. The lag axes are in terms of seasons (12 months).*

Example 1.28 SOI and Recruitment Correlation Analysis

The autocorrelation and cross-correlation functions are also useful for analyzing the joint behavior of two stationary series whose behavior may be related in some unspecified way. In Example 1.5 (see Figure 1.5), we have considered simultaneous monthly readings of the SOI and the number of new fish (Recruitment) computed from a model. Figure 1.16 shows the autocorrelation and cross-correlation functions (ACFs and CCF) for these two series. Both of the ACFs exhibit periodicities corresponding to the correlation between values separated by 12 units. Observations 12 months or one year apart are strongly positively correlated, as are observations at multiples such as 24, 36, 48, . . . Observations separated by six months are negatively correlated, showing that positive excursions tend to be associated with negative excursions six months removed.

The sample CCF in Figure 1.16, however, shows some departure from the cyclic component of each series and there is an obvious peak at $h = -6$. This result implies that SOI measured at time $t - 6$ months is associated with the Recruitment series at time t. We could say the SOI leads the Recruitment series by six months. The sign of the CCF is negative, leading to the conclusion that the two series move in different directions; that is, increases in SOI lead to decreases in Recruitment

and vice versa. We will discover in Chapter 2 that there is a relationship between the series, but the relationship is nonlinear. The dashed lines shown on the plots indicate $\pm 2/\sqrt{453}$ [see (1.42)], but since neither series is noise, these lines do not apply. To reproduce Figure 1.16 in R, use the following commands:

```
par(mfrow=c(3,1))
acf(soi, 48, main="Southern Oscillation Index")
acf(rec, 48, main="Recruitment")
ccf(soi, rec, 48, main="SOI vs Recruitment", ylab="CCF")
```

Example 1.29 Prewhitening and Cross Correlation Analysis

Although we do not have all the tools necessary yet, it is worthwhile to discuss the idea of prewhitening a series prior to a cross-correlation analysis. The basic idea is simple; in order to use Property 1.3, at least one of the series must be white noise. If this is not the case, there is no simple way to tell if a cross-correlation estimate is significantly different from zero. Hence, in Example 1.28, we were only guessing at the linear dependence relationship between SOI and Recruitment.

For example, in Figure 1.17 we generated two series, x_t and y_t, for $t = 1, \ldots, 120$ independently as

$$x_t = 2\cos(2\pi\, t\, \tfrac{1}{12}) + w_{t1} \quad \text{and} \quad y_t = 2\cos(2\pi\, [t+5]\, \tfrac{1}{12}) + w_{t2}$$

where $\{w_{t1}, w_{t2};\ t = 1, \ldots, 120\}$ are all independent standard normals. The series are made to resemble SOI and Recruitment. The generated data are shown in the top row of the figure. The middle row of Figure 1.17 shows the sample ACF of each series, each of which exhibits the cyclic nature of each series. The bottom row (left) of Figure 1.17 shows the sample CCF between x_t and y_t, which appears to show cross-correlation even though the series are independent. The bottom row (right) also displays the sample CCF between x_t and the prewhitened y_t, which shows that the two sequences are uncorrelated. By prewhtiening y_t, we mean that the signal has been removed from the data by running a regression of y_t on $\cos(2\pi t)$ and $\sin(2\pi t)$ [see Example 2.10] and then putting $\tilde{y}_t = y_t - \hat{y}_t$, where $\hat{y}_t$ are the predicted values from the regression.

The following code will reproduce Figure 1.17.

```
set.seed(1492)
num=120; t=1:num
X = ts(2*cos(2*pi*t/12) + rnorm(num), freq=12)
Y = ts(2*cos(2*pi*(t+5)/12) + rnorm(num), freq=12)
Yw = resid( lm(Y~ cos(2*pi*t/12) + sin(2*pi*t/12), na.action=NULL) )
par(mfrow=c(3,2), mgp=c(1.6,.6,0), mar=c(3,3,1,1) )
plot(X)
plot(Y)
acf(X,48, ylab='ACF(X)')
acf(Y,48, ylab='ACF(Y)')
ccf(X,Y,24, ylab='CCF(X,Y)')
ccf(X,Yw,24, ylab='CCF(X,Yw)', ylim=c(-.6,.6))
```

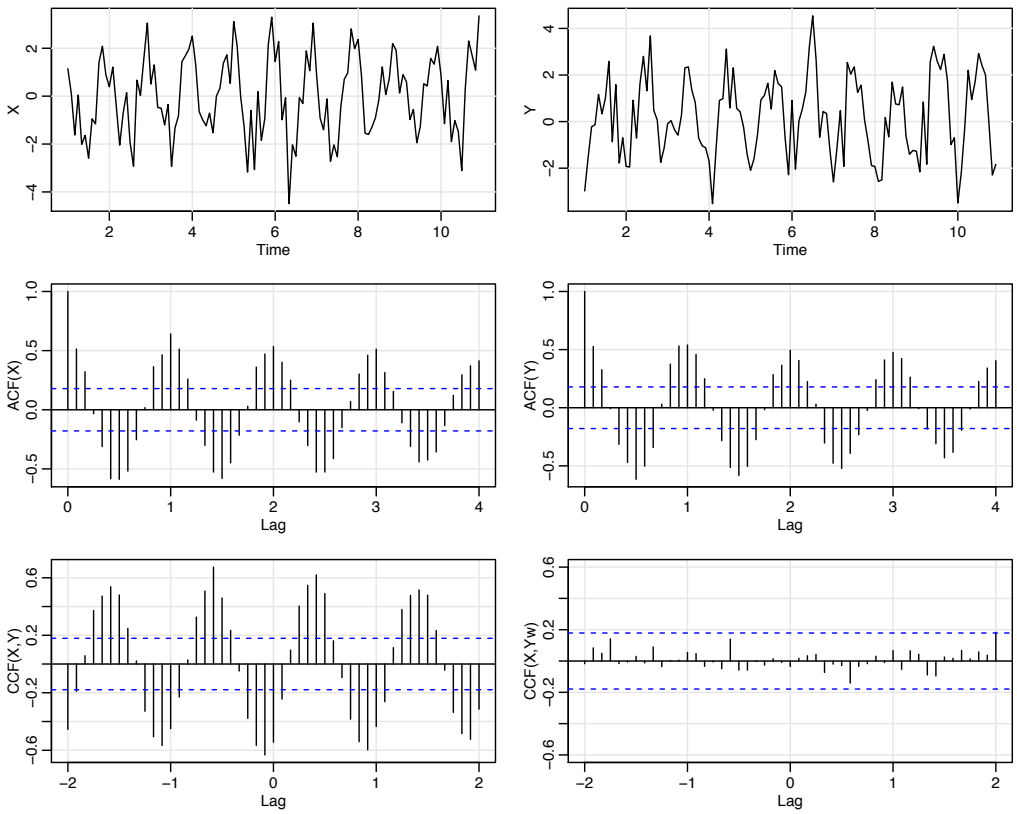

Fig. 1.17. *Display for Example 1.29.* Top row; *The generated series.* Middle row: *The sample ACF of each series.* Bottom row; *The sample CCF of the series (left) and the sample CCF of the first series with the prewhitened second series (right).*

1.6 Vector-Valued and Multidimensional Series

We frequently encounter situations in which the relationships between a number of jointly measured time series are of interest. For example, in the previous sections, we considered discovering the relationships between the SOI and Recruitment series. Hence, it will be useful to consider the notion of a *vector time series* $x_t = (x_{t1}, x_{t2}, \ldots, x_{tp})'$, which contains as its components p univariate time series. We denote the $p \times 1$ column vector of the observed series as x_t. The row vector x_t' is its transpose. For the stationary case, the $p \times 1$ mean vector

$$\mu = \mathrm{E}(x_t) \tag{1.43}$$

of the form $\mu = (\mu_{t1}, \mu_{t2}, \ldots, \mu_{tp})'$ and the $p \times p$ autocovariance matrix

$$\Gamma(h) = \mathrm{E}[(x_{t+h} - \mu)(x_t - \mu)'] \tag{1.44}$$

can be defined, where the elements of the matrix $\Gamma(h)$ are the cross-covariance functions

$$\gamma_{ij}(h) = \mathrm{E}[(x_{t+h,i} - \mu_i)(x_{tj} - \mu_j)] \tag{1.45}$$

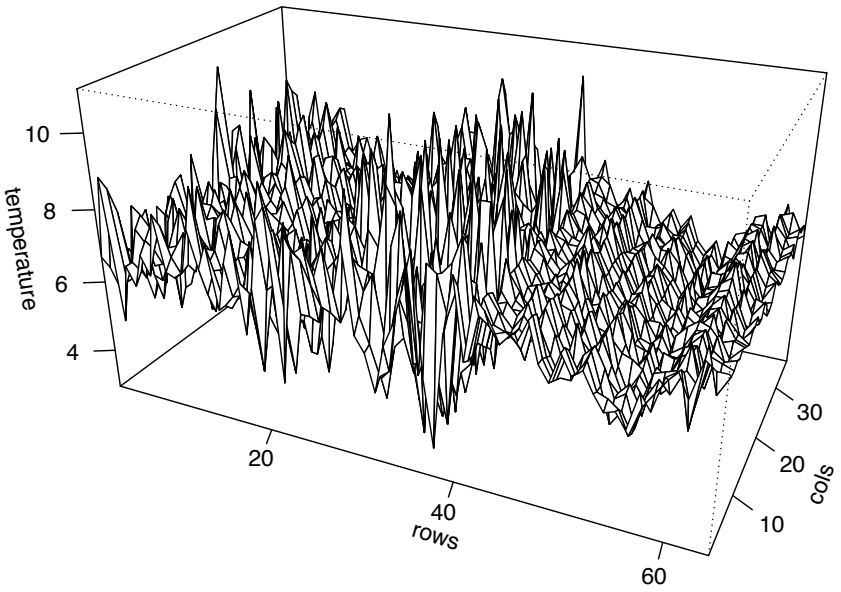

Fig. 1.18. *Two-dimensional time series of temperature measurements taken on a rectangular field* (64×36 *with 17-foot spacing). Data are from Bazza et al. (1988).*

for $i, j = 1, \ldots, p$. Because $\gamma_{ij}(h) = \gamma_{ji}(-h)$, it follows that

$$\Gamma(-h) = \Gamma'(h). \tag{1.46}$$

Now, the *sample autocovariance matrix* of the vector series x_t is the $p \times p$ matrix of sample cross-covariances, defined as

$$\hat{\Gamma}(h) = n^{-1} \sum_{t=1}^{n-h} (x_{t+h} - \bar{x})(x_t - \bar{x})', \tag{1.47}$$

where

$$\bar{x} = n^{-1} \sum_{t=1}^{n} x_t \tag{1.48}$$

denotes the $p \times 1$ *sample mean vector*. The symmetry property of the theoretical autocovariance (1.46) extends to the sample autocovariance (1.47), which is defined for negative values by taking

$$\hat{\Gamma}(-h) = \hat{\Gamma}(h)'. \tag{1.49}$$

In many applied problems, an observed series may be indexed by more than time alone. For example, the position in space of an experimental unit might be described by two coordinates, say, s_1 and s_2. We may proceed in these cases by defining a *multidimensional process* x_s as a function of the $r \times 1$ vector $s = (s_1, s_2, \ldots, s_r)'$, where s_i denotes the coordinate of the ith index.

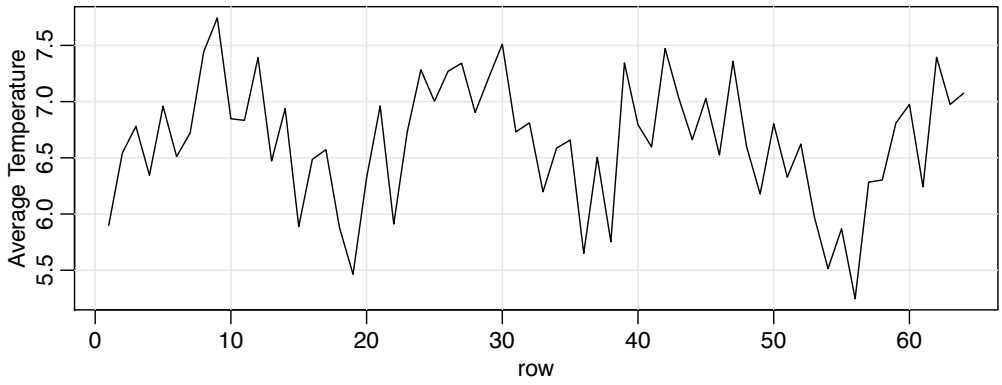

Fig. 1.19. *Row averages of the two-dimensional soil temperature profile.* $\bar{x}_{s_1,\cdot} = \sum_{s_2} x_{s_1,s_2}/36$.

Example 1.30 Soil Surface Temperatures

As an example, the two-dimensional ($r = 2$) temperature series x_{s_1,s_2} in Figure 1.18 is indexed by a row number s_1 and a column number s_2 that represent positions on a 64×36 spatial grid set out on an agricultural field. The value of the temperature measured at row s_1 and column s_2, is denoted by $x_s = x_{s1,s2}$. We can note from the two-dimensional plot that a distinct change occurs in the character of the two-dimensional surface starting at about row 40, where the oscillations along the row axis become fairly stable and periodic. For example, averaging over the 36 columns, we may compute an average value for each s_1 as in Figure 1.19. It is clear that the noise present in the first part of the two-dimensional series is nicely averaged out, and we see a clear and consistent temperature signal.

To generate Figure 1.18 and Figure 1.19 in R, use the following commands:

```
persp(1:64, 1:36, soiltemp, phi=25, theta=25, scale=FALSE, expand=4,
          ticktype="detailed", xlab="rows", ylab="cols", zlab="temperature")
plot.ts(rowMeans(soiltemp), xlab="row", ylab="Average Temperature")
```

The *autocovariance function* of a stationary multidimensional process, x_s, can be defined as a function of the multidimensional lag vector, say, $h = (h_1, h_2, \ldots, h_r)'$, as

$$\gamma(h) = \mathrm{E}[(x_{s+h} - \mu)(x_s - \mu)], \tag{1.50}$$

where

$$\mu = \mathrm{E}(x_s) \tag{1.51}$$

does not depend on the spatial coordinate s. For the two dimensional temperature process, (1.50) becomes

$$\gamma(h_1, h_2) = \mathrm{E}[(x_{s_1+h_1,s_2+h_2} - \mu)(x_{s_1,s_2} - \mu)], \tag{1.52}$$

which is a function of lag, both in the row (h_1) and column (h_2) directions.

The *multidimensional sample autocovariance function* is defined as

$$\hat{\gamma}(h) = (S_1 S_2 \cdots S_r)^{-1} \sum_{s_1} \sum_{s_2} \cdots \sum_{s_r} (x_{s+h} - \bar{x})(x_s - \bar{x}), \tag{1.53}$$

where $s = (s_1, s_2, \ldots, s_r)'$ and the range of summation for each argument is $1 \le s_i \le S_i - h_i$, for $i = 1, \ldots, r$. The mean is computed over the r-dimensional array, that is,

$$\bar{x} = (S_1 S_2 \cdots S_r)^{-1} \sum_{s_1} \sum_{s_2} \cdots \sum_{s_r} x_{s_1, s_2, \cdots, s_r}, \tag{1.54}$$

where the arguments s_i are summed over $1 \le s_i \le S_i$. The multidimensional sample autocorrelation function follows, as usual, by taking the scaled ratio

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}. \tag{1.55}$$

Example 1.31 Sample ACF of the Soil Temperature Series

The autocorrelation function of the two-dimensional (2d) temperature process can be written in the form

$$\hat{\rho}(h_1, h_2) = \frac{\hat{\gamma}(h_1, h_2)}{\hat{\gamma}(0, 0)},$$

where

$$\hat{\gamma}(h_1, h_2) = (S_1 S_2)^{-1} \sum_{s_1} \sum_{s_2} (x_{s_1+h_1, s_2+h_2} - \bar{x})(x_{s_1, s_2} - \bar{x})$$

Figure 1.20 shows the autocorrelation function for the temperature data, and we note the systematic periodic variation that appears along the rows. The autocovariance over columns seems to be strongest for $h_1 = 0$, implying columns may form replicates of some underlying process that has a periodicity over the rows. This idea can be investigated by examining the mean series over columns as shown in Figure 1.19.

The easiest way (that we know of) to calculate a 2d ACF in R is by using the fast Fourier transform (FFT) as shown below. Unfortunately, the material needed to understand this approach is given in Chapter 4, Section 4.3. The 2d autocovariance function is obtained in two steps and is contained in `cs` below; $\hat{\gamma}(0, 0)$ is the (1,1) element so that $\hat{\rho}(h_1, h_2)$ is obtained by dividing each element by that value. The 2d ACF is contained in `rs` below, and the rest of the code is simply to arrange the results to yield a nice display.

```
fs = Mod(fft(soiltemp-mean(soiltemp)))^2/(64*36)
cs = Re(fft(fs, inverse=TRUE)/sqrt(64*36))  # ACovF
rs = cs/cs[1,1]                             # ACF
rs2 = cbind(rs[1:41,21:2], rs[1:41,1:21])
rs3 = rbind(rs2[41:2,], rs2)
par(mar = c(1,2.5,0,0)+.1)
persp(-40:40, -20:20, rs3, phi=30, theta=30, expand=30, scale="FALSE",
          ticktype="detailed", xlab="row lags", ylab="column lags",
          zlab="ACF")
```

The sampling requirements for multidimensional processes are rather severe because values must be available over some uniform grid in order to compute the ACF. In some areas of application, such as in soil science, we may prefer to sample a limited number of rows or *transects* and hope these are essentially replicates of the

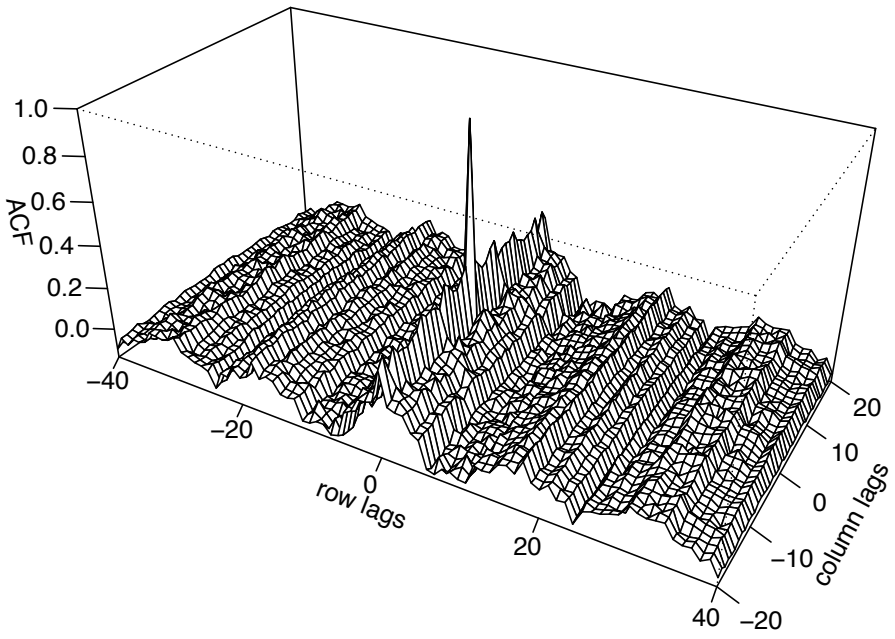

Fig. 1.20. *Two-dimensional autocorrelation function for the soil temperature data.*

basic underlying phenomenon of interest. One-dimensional methods can then be applied. When observations are irregular in time space, modifications to the estimators need to be made. Systematic approaches to the problems introduced by irregularly spaced observations have been developed by Journel and Huijbregts (1978) or Cressie (1993). We shall not pursue such methods in detail here, but it is worth noting that the introduction of the *variogram*

$$2V_x(h) = \text{var}\{x_{s+h} - x_s\} \tag{1.56}$$

and its sample estimator

$$2\hat{V}_x(h) = \frac{1}{N(h)} \sum_s (x_{s+h} - x_s)^2 \tag{1.57}$$

play key roles, where $N(h)$ denotes both the number of points located within h, and the sum runs over the points in the neighborhood. Clearly, substantial indexing difficulties will develop from estimators of the kind, and often it will be difficult to find non-negative definite estimators for the covariance function. Problem 1.27 investigates the relation between the variogram and the autocovariance function in the stationary case.

Problems

Section 1.1

1.1 To compare the earthquake and explosion signals, plot the data displayed in Figure 1.7 on the same graph using different colors or different line types and comment on the results. (The R code in Example 1.11 may be of help on how to add lines to existing plots.)

1.2 Consider a signal-plus-noise model of the general form $x_t = s_t + w_t$, where w_t is Gaussian white noise with $\sigma_w^2 = 1$. Simulate and plot $n = 200$ observations from each of the following two models.

(a) $x_t = s_t + w_t$, for $t = 1, ..., 200$, where

$$s_t = \begin{cases} 0, & t = 1, \ldots, 100 \\ 10 \exp\{-\frac{(t-100)}{20}\} \cos(2\pi t/4), & t = 101, \ldots, 200. \end{cases}$$

Hint:

```
s = c(rep(0,100), 10*exp(-(1:100)/20)*cos(2*pi*1:100/4))
x = s + rnorm(200)
plot.ts(x)
```

(b) $x_t = s_t + w_t$, for $t = 1, \ldots, 200$, where

$$s_t = \begin{cases} 0, & t = 1, \ldots, 100 \\ 10 \exp\{-\frac{(t-100)}{200}\} \cos(2\pi t/4), & t = 101, \ldots, 200. \end{cases}$$

(c) Compare the general appearance of the series (a) and (b) with the earthquake series and the explosion series shown in Figure 1.7. In addition, plot (or sketch) and compare the signal modulators (a) $\exp\{-t/20\}$ and (b) $\exp\{-t/200\}$, for $t = 1, 2, \ldots, 100$.

Section 1.2

1.3 (a) Generate $n = 100$ observations from the autoregression

$$x_t = -.9x_{t-2} + w_t$$

with $\sigma_w = 1$, using the method described in Example 1.10. Next, apply the moving average filter

$$v_t = (x_t + x_{t-1} + x_{t-2} + x_{t-3})/4$$

to x_t, the data you generated. Now plot x_t as a line and superimpose v_t as a dashed line. Comment on the behavior of x_t and how applying the moving average filter changes that behavior. [*Hints:* Use `v = filter(x, rep(1/4, 4), sides = 1)` for the filter and note that the R code in Example 1.11 may be of help on how to add lines to existing plots.]

(b) Repeat (a) but with

$$x_t = \cos(2\pi t/4).$$

(c) Repeat (b) but with added N(0, 1) noise,

$$x_t = \cos(2\pi t/4) + w_t.$$

(d) Compare and contrast (a)–(c); i.e., how does the moving average change each series.

Section 1.3

1.4 Show that the autocovariance function can be written as

$$\gamma(s,t) = \mathrm{E}[(x_s - \mu_s)(x_t - \mu_t)] = \mathrm{E}(x_s x_t) - \mu_s \mu_t,$$

where $\mathrm{E}[x_t] = \mu_t$.

1.5 For the two series, x_t, in Problem 1.2 (a) and (b):

(a) Compute and plot the mean functions $\mu_x(t)$, for $t = 1, \ldots, 200$.
(b) Calculate the autocovariance functions, $\gamma_x(s,t)$, for $s, t = 1, \ldots, 200$.

Section 1.4

1.6 Consider the time series

$$x_t = \beta_1 + \beta_2 t + w_t,$$

where β_1 and β_2 are known constants and w_t is a white noise process with variance σ_w^2.

(a) Determine whether x_t is stationary.
(b) Show that the process $y_t = x_t - x_{t-1}$ is stationary.
(c) Show that the mean of the moving average

$$v_t = \frac{1}{2q+1} \sum_{j=-q}^{q} x_{t-j}$$

is $\beta_1 + \beta_2 t$, and give a simplified expression for the autocovariance function.

1.7 For a moving average process of the form

$$x_t = w_{t-1} + 2w_t + w_{t+1},$$

where w_t are independent with zero means and variance σ_w^2, determine the autocovariance and autocorrelation functions as a function of lag $h = s - t$ and plot the ACF as a function of h.

1.8 Consider the random walk with drift model

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

for $t = 1, 2, \ldots$, with $x_0 = 0$, where w_t is white noise with variance σ_w^2.

(a) Show that the model can be written as $x_t = \delta t + \sum_{k=1}^{t} w_k$.
(b) Find the mean function and the autocovariance function of x_t.
(c) Argue that x_t is not stationary.
(d) Show $\rho_x(t-1, t) = \sqrt{\frac{t-1}{t}} \to 1$ as $t \to \infty$. What is the implication of this result?
(e) Suggest a transformation to make the series stationary, and prove that the transformed series is stationary. (Hint: See Problem 1.6b.)

1.9 A time series with a periodic component can be constructed from

$$x_t = U_1 \sin(2\pi\omega_0 t) + U_2 \cos(2\pi\omega_0 t),$$

where U_1 and U_2 are independent random variables with zero means and $\mathrm{E}(U_1^2) = \mathrm{E}(U_2^2) = \sigma^2$. The constant ω_0 determines the period or time it takes the process to make one complete cycle. Show that this series is weakly stationary with autocovariance function

$$\gamma(h) = \sigma^2 \cos(2\pi\omega_0 h).$$

1.10 Suppose we would like to predict a single stationary series x_t with zero mean and autocorrelation function $\gamma(h)$ at some time in the future, say, $t + \ell$, for $\ell > 0$.

(a) If we predict using only x_t and some scale multiplier A, show that the mean-square prediction error

$$MSE(A) = \mathrm{E}[(x_{t+\ell} - Ax_t)^2]$$

is minimized by the value

$$A = \rho(\ell).$$

(b) Show that the minimum mean-square prediction error is

$$MSE(A) = \gamma(0)[1 - \rho^2(\ell)].$$

(c) Show that if $x_{t+\ell} = Ax_t$, then $\rho(\ell) = 1$ if $A > 0$, and $\rho(\ell) = -1$ if $A < 0$.

1.11 Consider the linear process defined in (1.31).

(a) Verify that the autocovariance function of the process is given by (1.32). Use the result to verify your answer to Problem 1.7. *Hint:* For $h \geq 0$, $\mathrm{cov}(x_{t+h}, x_t) = \mathrm{cov}(\sum_k \psi_k w_{t+h-k}, \sum_j \psi_j w_{t-j})$. For each $j \in \mathbb{Z}$, the only "survivor" will be when $k = h + j$.
(b) Show that x_t exists as a limit in mean square (see Appendix A).

1.12 For two weakly stationary series x_t and y_t, verify (1.30).

1.13 Consider the two series

$$x_t = w_t$$

$$y_t = w_t - \theta w_{t-1} + u_t,$$

where w_t and u_t are independent white noise series with variances σ_w^2 and σ_u^2, respectively, and θ is an unspecified constant.

(a) Express the ACF, $\rho_y(h)$, for $h = 0, \pm 1, \pm 2, \ldots$ of the series y_t as a function of σ_w^2, σ_u^2, and θ.
(b) Determine the CCF, $\rho_{xy}(h)$ relating x_t and y_t.
(c) Show that x_t and y_t are jointly stationary.

1.14 Let x_t be a stationary normal process with mean μ_x and autocovariance function $\gamma(h)$. Define the nonlinear time series

$$y_t = \exp\{x_t\}.$$

(a) Express the mean function $\mathrm{E}(y_t)$ in terms of μ_x and $\gamma(0)$. The moment generating function of a normal random variable x with mean μ and variance σ^2 is

$$M_x(\lambda) = \mathrm{E}[\exp\{\lambda x\}] = \exp\left\{\mu\lambda + \frac{1}{2}\sigma^2\lambda^2\right\}.$$

(b) Determine the autocovariance function of y_t. The sum of the two normal random variables $x_{t+h} + x_t$ is still a normal random variable.

1.15 Let w_t, for $t = 0, \pm 1, \pm 2, \ldots$ be a normal white noise process, and consider the series

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

Determine the mean and autocovariance function of x_t, and state whether it is stationary.

1.16 Consider the series

$$x_t = \sin(2\pi U t),$$

$t = 1, 2, \ldots$, where U has a uniform distribution on the interval $(0, 1)$.

(a) Prove x_t is weakly stationary.
(b) Prove x_t is not strictly stationary.

1.17 Suppose we have the linear process x_t generated by

$$x_t = w_t - \theta w_{t-1},$$

$t = 0, 1, 2, \ldots$, where $\{w_t\}$ is independent and identically distributed with characteristic function $\phi_w(\cdot)$, and θ is a fixed constant. [Replace "characteristic function" with "moment generating function" if instructed to do so.]

(a) Express the joint characteristic function of $x_1, x_2, \ldots, x_n$, say,

$$\phi_{x_1,x_2,\ldots,x_n}(\lambda_1, \lambda_2, \ldots, \lambda_n),$$

in terms of $\phi_w(\cdot)$.
(b) Deduce from (a) that x_t is strictly stationary.

1.18 Suppose that x_t is a linear process of the form (1.31). Prove

$$\sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty.$$

Section 1.5

1.19 Suppose $x_t = \mu + w_t + \theta w_{t-1}$, where $w_t \sim wn(0, \sigma_w^2)$.

(a) Show that mean function is $\mathrm{E}(x_t) = \mu$.
(b) Show that the autocovariance function of x_t is given by $\gamma_x(0) = \sigma_w^2(1 + \theta^2)$, $\gamma_x(\pm 1) = \sigma_w^2\theta$, and $\gamma_x(h) = 0$ otherwise.
(c) Show that x_t is stationary for all values of $\theta \in \mathbb{R}$.
(d) Use (1.35) to calculate $\mathrm{var}(\bar{x})$ for estimating μ when (i) $\theta = 1$, (ii) $\theta = 0$, and (iii) $\theta = -1$
(e) In time series, the sample size n is typically large, so that $\frac{(n-1)}{n} \approx 1$. With this as a consideration, comment on the results of part (d); in particular, how does the accuracy in the estimate of the mean μ change for the three different cases?

1.20 (a) Simulate a series of $n = 500$ Gaussian white noise observations as in Example 1.8 and compute the sample ACF, $\hat{\rho}(h)$, to lag 20. Compare the sample ACF you obtain to the actual ACF, $\rho(h)$. [Recall Example 1.19.]
(b) Repeat part (a) using only $n = 50$. How does changing n affect the results?

1.21 (a) Simulate a series of $n = 500$ moving average observations as in Example 1.9 and compute the sample ACF, $\hat{\rho}(h)$, to lag 20. Compare the sample ACF you obtain to the actual ACF, $\rho(h)$. [Recall Example 1.20.]
(b) Repeat part (a) using only $n = 50$. How does changing n affect the results?

1.22 Although the model in Problem 1.2(a) is not stationary (Why?), the sample ACF can be informative. For the data you generated in that problem, calculate and plot the sample ACF, and then comment.

1.23 Simulate a series of $n = 500$ observations from the signal-plus-noise model presented in Example 1.12 with $\sigma_w^2 = 1$. Compute the sample ACF to lag 100 of the data you generated and comment.

1.24 For the time series y_t described in Example 1.26, verify the stated result that $\rho_y(1) = -.47$ and $\rho_y(h) = 0$ for $h > 1$.

1.25 A real-valued function $g(t)$, defined on the integers, is non-negative definite if and only if

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_i g(t_i - t_j) a_j \geq 0$$

for all positive integers n and for all vectors $a = (a_1, a_2, \ldots, a_n)'$ and $t = (t_1, t_2, \ldots, t_n)'$. For the matrix $G = \{g(t_i - t_j); i, j = 1, 2, \ldots, n\}$, this implies that $a'Ga \geq 0$ for all vectors a. It is called positive definite if we can replace '$\geq$' with '$>$' for all $a \neq 0$, the zero vector.

(a) Prove that $\gamma(h)$, the autocovariance function of a stationary process, is a non-negative definite function.
(b) Verify that the sample autocovariance $\hat{\gamma}(h)$ is a non-negative definite function.

Section 1.6

1.26 Consider a collection of time series $x_{1t}, x_{2t}, \ldots, x_{Nt}$ that are observing some common signal μ_t observed in noise processes $e_{1t}, e_{2t}, \ldots, e_{Nt}$, with a model for the j-th observed series given by

$$x_{jt} = \mu_t + e_{jt}.$$

Suppose the noise series have zero means and are uncorrelated for different j. The common autocovariance functions of all series are given by $\gamma_e(s, t)$. Define the sample mean

$$\bar{x}_t = \frac{1}{N}\sum_{j=1}^{N} x_{jt}.$$

(a) Show that $\mathrm{E}[\bar{x}_t] = \mu_t$.
(b) Show that $\mathrm{E}[(\bar{x}_t - \mu)^2)] = N^{-1}\gamma_e(t, t)$.
(c) How can we use the results in estimating the common signal?

1.27 A concept used in *geostatistics*, see Journel and Huijbregts (1978) or Cressie (1993), is that of the *variogram*, defined for a spatial process x_s, $s = (s_1, s_2)$, for $s_1, s_2 = 0, \pm 1, \pm 2, \ldots$, as

$$V_x(h) = \frac{1}{2}\mathrm{E}[(x_{s+h} - x_s)^2],$$

where $h = (h_1, h_2)$, for $h_1, h_2 = 0, \pm 1, \pm 2, \ldots$ Show that, for a stationary process, the variogram and autocovariance functions can be related through

$$V_x(h) = \gamma(0) - \gamma(h),$$

where $\gamma(h)$ is the usual lag h covariance function and $0 = (0, 0)$. Note the easy extension to any spatial dimension.

The following problems require the material given in Appendix A

1.28 Suppose $x_t = \beta_0 + \beta_1 t$, where β_0 and β_1 are constants. Prove as $n \to \infty$, $\hat{\rho}_x(h) \to 1$ for fixed h, where $\hat{\rho}_x(h)$ is the ACF (1.37).

1.29 (a) Suppose x_t is a weakly stationary time series with mean zero and with absolutely summable autocovariance function, $\gamma(h)$, such that

$$\sum_{h=-\infty}^{\infty} \gamma(h) = 0.$$

Prove that $\sqrt{n}\,\bar{x} \xrightarrow{p} 0$, where $\bar{x}$ is the sample mean (1.34).

(b) Give an example of a process that satisfies the conditions of part (a). What is special about this process?

1.30 Let x_t be a linear process of the form (A.43)–(A.44). If we define

$$\tilde{\gamma}(h) = n^{-1} \sum_{t=1}^{n} (x_{t+h} - \mu_x)(x_t - \mu_x),$$

show that

$$n^{1/2}\big(\tilde{\gamma}(h) - \hat{\gamma}(h)\big) = o_p(1).$$

Hint: The Markov Inequality

$$\Pr\{|x| \geq \epsilon\} < \frac{\mathrm{E}|x|}{\epsilon}$$

can be helpful for the cross-product terms.

1.31 For a linear process of the form

$$x_t = \sum_{j=0}^{\infty} \phi^j w_{t-j},$$

where $\{w_t\}$ satisfies the conditions of Theorem A.7 and $|\phi| < 1$, show that

$$\sqrt{n}\frac{(\hat{\rho}_x(1) - \rho_x(1))}{\sqrt{1 - \rho_x^2(1)}} \xrightarrow{d} N(0, 1),$$

and construct a 95% confidence interval for ϕ when $\hat{\rho}_x(1) = .64$ and $n = 100$.

1.32 Let $\{x_t;\ t = 0, \pm 1, \pm 2, \ldots\}$ be iid$(0, \sigma^2)$.

(a) For $h \geq 1$ and $k \geq 1$, show that $x_t x_{t+h}$ and $x_s x_{s+k}$ are uncorrelated for all $s \neq t$.

(b) For fixed $h \geq 1$, show that the $h \times 1$ vector

$$\sigma^{-2} n^{-1/2} \sum_{t=1}^{n} (x_t x_{t+1}, \ldots, x_t x_{t+h})' \xrightarrow{d} (z_1, \ldots, z_h)'$$

where $z_1, \ldots, z_h$ are iid N$(0, 1)$ random variables. [Hint: Use the Cramér-Wold device.]

(c) Show, for each $h \geq 1$,

$$n^{-1/2}\left[\sum_{t=1}^{n} x_t x_{t+h} - \sum_{t=1}^{n-h}(x_t - \bar{x})(x_{t+h} - \bar{x})\right] \xrightarrow{p} 0 \qquad \text{as } n \to \infty$$

where $\bar{x} = n^{-1}\sum_{t=1}^{n} x_t$.

(d) Noting that $n^{-1}\sum_{t=1}^{n} x_t^2 \xrightarrow{p} \sigma^2$ by the WLLN, conclude that

$$n^{1/2}\,[\hat{\rho}(1), \ldots, \hat{\rho}(h)]' \xrightarrow{d} (z_1, \ldots, z_h)'$$

where $\hat{\rho}(h)$ is the sample ACF of the data $x_1, \ldots, x_n$.