STAT 626: Outline of Lecture 8

Estimation of the Mean, Correlation and ACF (§2.3)

1. A Quick Review of How Statistics Works: **Sample** $\Longrightarrow$ **Population**.

   Estimation of μ, $\gamma(1)$, $\rho(1), \ldots$

2. The Sample Mean: $\bar{x}$

3. Sample Autocovariance Function: $\widehat{\gamma}(1)$

4. Distribution of the Sample Autocorrelation Function (ACF)

5. **The Sample Correlogram and Confidence Interval**

6. Bivariate Time Series: Estimation of Cross-Correlations.

Review of the DEFINITIONS

1. A Time Series $\{x_t\}$ is **stationary** if

   (a) the mean function $E(x_t)$ does not depend on the time t,

   (b) the covariance function $\text{cov}(x_s, x_t)$ depends on the times s, t only through the (time-)lag $|s - t|$.

2. **Autocovariance Function** of a Stationary Time Series:

$$\gamma(h) = \text{cov}(x_{t+h}, x_t), \quad h = 0, 1, \ldots.$$

   NOTE: Setting $h = 0$ it follows that

$$\gamma(0) = \text{cov}(x_t, x_t) = \text{var}(x_t),$$

   so that the variance of the series, just like its mean, is not time-varying.

3. **The Autocorrelation Function (ACF)**

$$\rho(h) = \frac{\text{cov}(x_{t+h}, x_t)}{\sqrt{\text{var}(x_{t+h})\text{var}(x_t)}} = \frac{\gamma(h)}{\gamma(0)}, \quad h = 0, 1, \ldots.$$

4. **Correlogram** is the plot of $\rho(h)$ vs h.

   Its role in identifying TS models is just like that of the histogram in basic statistics.

LINEAR PROCESSES are the most general form of stationary processes, they are formed as linear combinations of a **white noise** $\{w_t\} \sim \mathrm{WN}(0, \sigma_w^2)$.

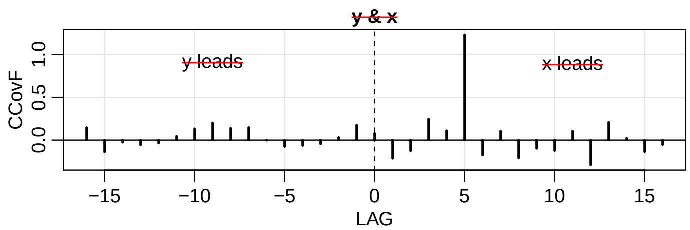
Moving Average of order q or MA(q) Models:

$$x_t = w_t + \theta_1 w_{t-1} + \ldots + \theta_q w_{t-q},$$

where $\theta = (\theta_1, \ldots, \theta_q)$ is the vector of parameters.

What happens when $q = \infty$?
MA(∞) Models or Processes.

Example: Compute the ACF of MA(∞) when $\theta_i = \phi^i,\ i = 1, \ldots,$ for a $|\phi| < 1$.



~~Figure 2.2 *Demonstration of the results of Example 2.25 when* $\ell = 5$*. The title indicates which series is leading.*~~

~~Since the largest value of $|\gamma_x(h-\ell)|$ is $\gamma_x(0)$, i.e., when $h=\ell$, the cross-covariance function will look like the autocovariance of the input series x_t, and it will have an extremum on the positive side if x_t leads y_t and an extremum on the negative side if x_t lags y_t. Below is the R code of an example with a delay of $\ell = 5$ and $\hat{\gamma}_{yx}(h)$, which is defined in Definition 2.30, shown in Figure 2.2.~~

```
x = rnorm(100)
y = lag(x,-5) + rnorm(100)
ccf(y, x, ylab="CCovF", type="covariance", panel.first=grid())
```

◊

2.3 Estimation of Correlation

For data analysis, only the sample values, $x_1, x_2, \dots, x_n$, are available for estimating the mean, autocovariance, and autocorrelation functions. *In this case, the assumption of stationarity becomes critical and allows the use of averaging to estimate the population mean and covariance functions.*

Accordingly, if a time series is stationary, the mean function (2.10) $\mu_t = \mu$ is constant so we can estimate it by the *sample mean*,

$$\bar{x} = \frac{1}{n}\sum_{t=1}^{n} x_t. \tag{2.19}$$

The estimate is unbiased, $E(\bar{x}) = \mu$, and its standard error is the square root of $\mathrm{var}(\bar{x})$, which can be computed using first principles (Property 2.7), and is given by

$$\mathrm{var}(\bar{x}) = \frac{1}{n}\sum_{h=-n}^{n}\Big(1 - \frac{|h|}{n}\Big)\gamma_x(h)\,. \tag{2.20}$$

If the process is white noise, (2.20) 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 2.10).

The theoretical autocorrelation function, (2.12), is estimated by the sample ACF as follows.

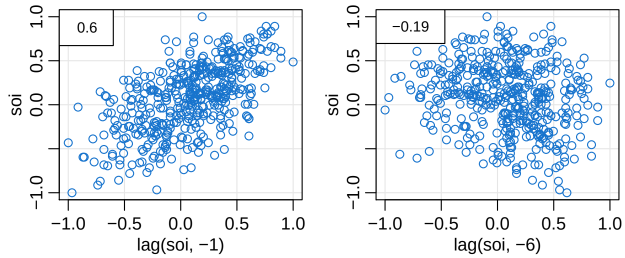


Figure 2.3 *Display for Example 2.27. For the SOI series, we have a scatterplot of pairs of values one month apart (left) and six months apart (right). The estimated autocorrelation is displayed in the box.*

Definition 2.26. *The* **sample autocorrelation function (ACF)** *is defined as*

$$\widehat{\rho}(h) = \frac{\widehat{\gamma}(h)}{\widehat{\gamma}(0)} = \frac{\sum_{t=1}^{n-h}(x_{t+h} - \bar{x})(x_t - \bar{x})}{\sum_{t=1}^{n}(x_t - \bar{x})^2} \tag{2.21}$$

for $h = 0, 1, \ldots, n-1$.

The sum in the numerator of (2.21) runs over a restricted range because x_{t+h} is not available for $t + h > n$. Note that we are in fact estimating the autocovariance function by

$$\widehat{\gamma}(h) = n^{-1} \sum_{t=1}^{n-h} (x_{t+h} - \bar{x})(x_t - \bar{x}), \tag{2.22}$$

with $\widehat{\gamma}(-h) = \widehat{\gamma}(h)$ for $h = 0, 1, \ldots, n-1$. That is, we divide by n even though there are only $n - h$ pairs of observations at lag h,

$$\{(x_{t+h}, x_t);\ t = 1, \ldots, n-h\}\,. \tag{2.23}$$

This assures that the sample autocovariance function will behave as a true autocovariance function, and for example, will not give negative values when estimating $\mathrm{var}(\bar{x})$ by replacing $\gamma_x(h)$ with $\widehat{\gamma}_x(h)$ in (2.20).

Example 2.27. Sample ACF and Scatterplots

Estimating autocorrelation is similar to estimating of correlation in the classical case, but we use (2.21) instead of the sample correlation coefficient you learned in a course on regression. Figure 2.3 shows an example using the SOI series where $\widehat{\rho}(1) = .60$ and $\widehat{\rho}(6) = -.19$. The following code was used for Figure 2.3.

```
(r = acf1(soi, 6, plot=FALSE))  # sample acf values
  [1]  0.60  0.37  0.21  0.05 -0.11 -0.19
par(mfrow=c(1,2), mar=c(2.5,2.5,0,0)+.5, mgp=c(1.6,.6,0))
plot(lag(soi,-1), soi, col="dodgerblue3", panel.first=grid(lty=1))
legend("topleft", legend=r[1], bg="white", adj=.45, cex = 0.85)
plot(lag(soi,-6), soi, col="dodgerblue3", panel.first=grid(lty=1))
legend("topleft", legend=r[6], bg="white", adj=.25, cex = 0.8)
```

◊

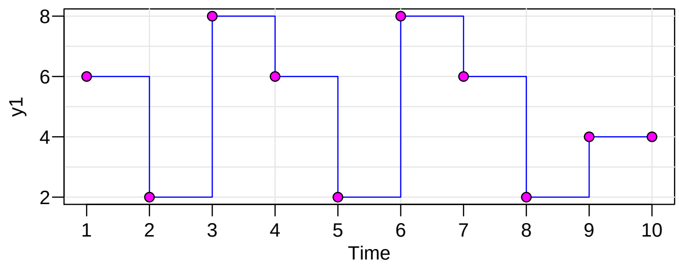


Figure 2.4: *Realization of* (2.24), $n = 10$.

Remark. It is important to note that this approach to estimating correlation ***makes sense only if the data are stationary.*** If the data were not stationary, each point in the graph could be an observation from a different correlation structure.

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.

Property 2.28 (Large-Sample Distribution of the ACF). *If x_t is white noise, then for n large and under mild conditions, the sample ACF, $\widehat{\rho}_x(h)$, for $h = 1, 2, \dots, H$, where H is fixed but arbitrary, is approximately normal with zero mean and standard deviation given by of $1/\sqrt{n}$.*

Based on Property 2.28, we obtain a rough method for assessing whether a series is white noise by determining how many values of $\widehat{\rho}(h)$ are outside the interval $\pm 2/\sqrt{n}$ (two standard errors); for white noise, approximately 95% of the sample ACFs should be within these limits.[2]

Example 2.29. 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 $x_t = 2$ when a head is obtained and $x_t = -2$ when a tail is obtained. Then, because we can only appreciate 2, 4, 6, or 8, we let

$$y_t = 5 + x_t - .5x_{t-1}\,. \tag{2.24}$$

We consider two cases, one with a small sample size ($n = 10$; see Figure 2.4) and another with a moderate sample size ($n = 100$).

```
set.seed(101011)
x1 = sample(c(-2,2),  11, replace=TRUE) # simulated coin tosses
x2 = sample(c(-2,2), 101, replace=TRUE)
y1 = 5 + filter(x1, sides=1, filter=c(1,-.5))[-1]
y2 = 5 + filter(x2, sides=1, filter=c(1,-.5))[-1]
tsplot(y1, type="s", col=4, xaxt="n", yaxt="n")  # y2 not shown
axis(1, 1:10); axis(2, seq(2,8,2), las=1)
```

[2]In this text, $z_{.025} = 1.95996398454\ldots$ of normal fame, often rounded to 1.96, is rounded to 2.

```
points(y1, pch=21, cex=1.1, bg=6)
acf(y1, lag.max=4, plot=FALSE)  # 1/√10 =.32
      0      1      2      3      4
  1.000 -0.352 -0.316  0.510 -0.245
acf(y2, lag.max=4, plot=FALSE)  # 1/√100 =.1
      0      1      2      3      4
  1.000 -0.496  0.067  0.087  0.063
```

The theoretical ACF can be obtained from the model (2.24) using first principles so that

$$\rho_y(1) = \frac{-.5}{1 + .5^2} = -.4$$

and $\rho_y(h) = 0$ for $|h| > 1$ (do Problem 2.15 now). It is interesting to compare the theoretical ACF with sample ACFs for the realization where $n = 10$ and where $n = 100$; note that small sample size means increased variability. $\lozenge$

Definition 2.30. *The estimators for the cross-covariance function, $\gamma_{xy}(h)$, as given in (2.17) and the cross-correlation, $\rho_{xy}(h)$, in (2.18) are given, respectively, by the* **sample cross-covariance function**

$$\widehat{\gamma}_{xy}(h) = n^{-1} \sum_{t=1}^{n-h} (x_{t+h} - \bar{x})(y_t - \bar{y}), \tag{2.25}$$

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

$$\widehat{\rho}_{xy}(h) = \frac{\widehat{\gamma}_{xy}(h)}{\sqrt{\widehat{\gamma}_x(0)\widehat{\gamma}_y(0)}}. \tag{2.26}$$

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 2.25 for the theoretical cross-covariance function. Because $-1 \le \widehat{\rho}_{xy}(h) \le 1$, the practical importance of peaks can be assessed by comparing their magnitudes with their theoretical maximum values.

Property 2.31 (Large-Sample Distribution of Cross-Correlation). *If x_t and y_t are independent processes, then under mild conditions, the large sample distribution of $\widehat{\rho}_{xy}(h)$ is normal with mean zero and standard deviation $1/\sqrt{n}$ if at least one of the processes is independent white noise.*

Example 2.32. 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 1.4 (see Figure 1.5), we have considered simultaneous monthly readings of the SOI and an index for the number of new fish (Recruitment).

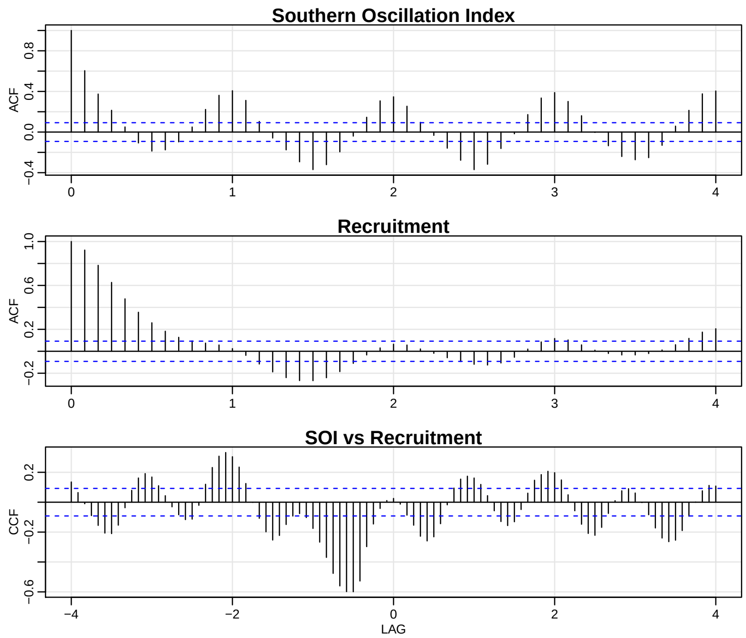


Figure 2.5 *Sample ACFs of the SOI series (top) and of the Recruitment series (middle), and the sample CCF of the two series (bottom); negative lags indicate SOI leads Recruitment. The lag axes are in terms of seasons (12 months).*

Figure 2.5 shows the sample 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 $24, 36, 48, \ldots$ Observations separated by six months are negatively correlated, showing that positive excursions tend to be associated with negative excursions six months removed. This appearance is rather characteristic of the pattern that would be produced by a sinusoidal component with a period of 12 months; see Example 2.33. The cross-correlation function peaks at $h = -6$, showing that the 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 at $h = -6$ 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. Again, note the periodicity of 12 months in the CCF.

The flat lines shown on the plots indicate $\pm 2/\sqrt{453}$, so that upper values would be exceeded about 2.5% of the time if the noise were white as specified in Property 2.28 and Property 2.31. Of course, neither series is noise, so we can ignore these lines. To reproduce Figure 2.5 in R, use the following commands:

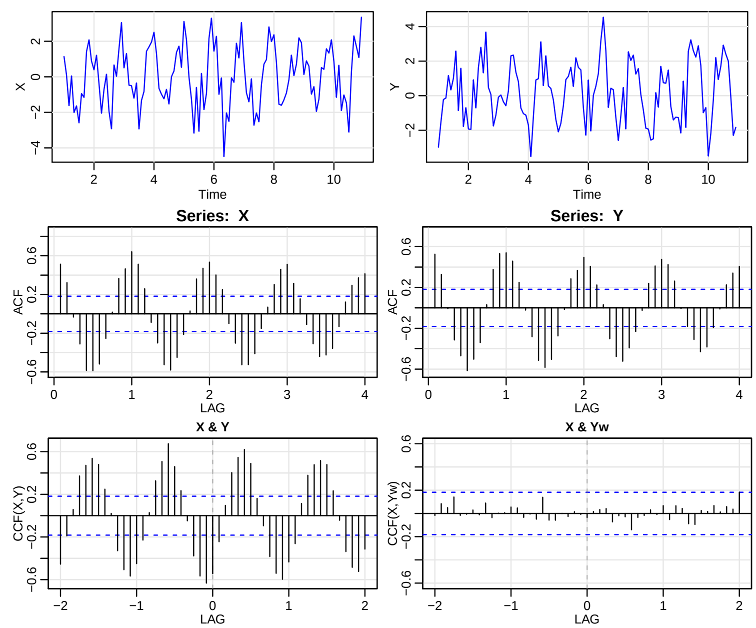


Figure 2.6: *Display for Example 2.33*

```
par(mfrow=c(3,1))
acf1(soi, 48, main="Southern Oscillation Index")
acf1(rec, 48, main="Recruitment")
ccf2(soi, rec, 48, main="SOI vs Recruitment")
```

◊

Example 2.33. Prewhitening and Cross Correlation Analysis *
Although we do not have all the tools necessary yet, it is worthwhile discussing the idea of prewhitening a series prior to a cross-correlation analysis. The basic idea is simple, to use Property 2.31, at least one of the series must be white noise. If this is not the case, there is no simple way of telling if a cross-correlation estimate is significantly different from zero. Hence, in Example 2.32, we were only guessing at the linear dependence relationship between SOI and Recruitment. The preferred method of prewhitening a time series is discussed in Section 8.5.

For example, in Figure 2.6 we generated two series, x_t and y_t, for $t = 1, \dots, 120$ 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 $\{w_{t1}, w_{t2};\, t = 1, \dots, 120\}$ 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 2.6 shows the sample ACF of each series, each of which exhibits the cyclic nature of each series. The bottom row (left) of Figure 2.6 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 $\cos(2\pi t/12)$ and $\sin(2\pi t/12)$ (both are needed to capture the phase; see Example 3.15) 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 2.6.

```
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))
tsplot(X, col=4);  tsplot(Y, col=4)
acf1(X, 48); acf1(Y, 48)
ccf2(X, Y, 24);  ccf2(X, Yw, 24, ylim=c(-.6,.6))
```

◊

Problems

2.1. In 25 words or less, and without using symbols, why is stationarity important?

2.2. Consider the time series

$$x_t = \beta_0 + \beta_1 t + w_t,$$

where β_0 and β_1 are regression coefficients, 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 two-sided moving average

$$v_t = \tfrac{1}{3}(x_{t-1} + x_t + x_{t+1})$$

is $\beta_0 + \beta_1 t$.

2.3. When smoothing time series data, it is sometimes advantageous to give decreasing amounts of weights to values farther away from the center. Consider the simple two-sided moving average smoother of the form

$$x_t = \tfrac{1}{4}(w_{t-1} + 2w_t + w_{t+1}),$$

where w_t are independent with zero mean and variance σ_w^2. Determine the autocovariance and autocorrelation functions as a function of lag h and sketch the ACF as a function of h.

2.4. We have not discussed the stationarity of autoregressive models, and we will do that in Chapter 4. But for now, let $x_t = \phi x_{t-1} + w_t$ where $w_t \sim wn(0,1)$ and ϕ is a constant. Assume x_t is stationary and x_{t-1} is uncorrelated with the noise term w_t.

(a) Show that mean function of x_t is $\mu_{xt} = 0$.

(b) Show $\gamma_x(0) = \text{var}(x_t) = 1/(1-\phi^2)$.

(c) For which values of ϕ does the solution to part (b) make sense?

(d) Find the lag-one autocorrelation, $\rho_x(1)$.

2.5. Consider the random walk with drift model

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

for $t = 1, 2, \dots$, 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.

2.6. Would you treat the global temperature data discussed in Example 1.2 and shown in Figure 1.2 as stationary or non-stationary? Support your answer.

2.7. 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 $E(U_1^2) = E(U_2^2) = \sigma^2$. 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

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

2.8. 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 $h = 0, \pm 1, \pm 2, \dots$ 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.

2.9. 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.

2.10. 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 $E(x_t) = \mu$.

(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 (2.20) to calculate $\text{var}(\bar{x})$ 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?

2.11. (a) Simulate a series of $n = 500$ Gaussian white noise observations as in Example 1.7 and compute the sample ACF, $\widehat{\rho}(h)$, to lag 20. Compare the sample ACF you obtain to the actual ACF, $\rho(h)$. [Recall Example 2.17.]

(b) Repeat part (a) using only $n = 50$. How does changing n affect the results?

2.12. (a) Simulate a series of $n = 500$ moving average observations as in Example 1.8 and compute the sample ACF, $\widehat{\rho}(h)$, to lag 20. Compare the sample ACF you obtain to the actual ACF, $\rho(h)$. [Recall Example 2.18.]

(b) Repeat part (a) using only $n = 50$. How does changing n affect the results?

2.13. Simulate 500 observations from the AR model specified in Example 1.9 and then plot the sample ACF to lag 50. What does the sample ACF tell you about the approximate cyclic behavior of the data? Hint: Recall Example 2.32.

2.14. Simulate a series of $n = 500$ observations from the signal-plus-noise model presented in Example 1.11 with (a) $\sigma_w = 0$, (b) $\sigma_w = 1$ and (c) $\sigma_w = 5$. Compute the sample ACF to lag 100 of the three series you generated and comment.

2.15. For the time series y_t described in Example 2.29, verify the stated result that $\rho_y(1) = -.4$ and $\rho_y(h) = 0$ for $h > 1$.