STAT 626: Outline of Lecture 4
Estimation of the Mean, Correlation and ACF (§1.6)

1. Review of Stationary TS and Computing ACF

2. The Sample Mean

3. Sample Autocovariance Function

4. Distribution of the Sample Autocorrelation Function (ACF)

5. The Sample Correlogram and Confidence Interval

Classroom Involvement Is The Key To Understanding

Tell me and I'll forget;

show me and I may remember;

involve me and I'll **understand**.

-Chinese proverb or Benjamin Franklin

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 we need in this course. 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 $|\phi| < 1$.

~~assumed to be positive definite, the multivariate normal density function can be written as~~

$$\cancel{f(\boldsymbol{x}) = (2\pi)^{-n/2}|\Gamma|^{-1/2}\exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'\Gamma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\},} \tag{1.31}$$

~~where $|\cdot|$ denotes the determinant. This distribution forms the basis for solving problems involving statistical inference for time series. If a Gaussian time series, $\{x_t\}$, is weakly stationary, then $\mu_t = \mu$ and $\gamma(t_i, t_j) = \gamma(|t_i - t_j|)$, so that the vector $\boldsymbol{\mu}$ and the matrix Γ are independent of time. These facts imply that all the finite distributions, (1.31), of the series $\{x_t\}$ depend only on time lag and not on the actual times, and hence the series must be strictly stationary.~~

1.6 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 $x_1, x_2, \dots, x_n$ 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 (1.21) $\mu_t = \mu$ is constant so that we can estimate it by the sample mean,

$$\bar{x} = \frac{1}{n}\sum_{t=1}^{n} x_t. \tag{1.32}$$

The standard error of the estimate is the square root of $\text{var}(\bar{x})$, which can be computed using first principles (recall footnote 3 on page 20), and is given by

$$\begin{aligned}
\text{var}(\bar{x}) &= \text{var}\left(\frac{1}{n}\sum_{t=1}^{n} x_t\right) = \frac{1}{n^2}\text{cov}\left(\sum_{t=1}^{n} x_t, \sum_{s=1}^{n} x_s\right) \\
&= \frac{1}{n^2}\Big(n\gamma_x(0) + (n-1)\gamma_x(1) + (n-2)\gamma_x(2) + \cdots + \gamma_x(n-1) \\
&\qquad\qquad + (n-1)\gamma_x(-1) + (n-2)\gamma_x(-2) + \cdots + \gamma_x(1-n)\Big) \\
&= \frac{1}{n}\sum_{h=-n}^{n}\left(1 - \frac{|h|}{n}\right)\gamma_x(h).
\end{aligned} \tag{1.33}$$

If the process is white noise, (1.33) 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, (1.22), is estimated by the sample autocovariance function defined as follows.

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

$$\widehat{\gamma}(h) = n^{-1} \sum_{t=1}^{n-h} (x_{t+h} - \bar{x})(x_t - \bar{x}), \tag{1.34}$$

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

The sum in (1.34) runs over a restricted range because x_{t+h} is not available for $t + h > n$. The estimator in (1.34) is preferred to the one that would be obtained by dividing by $n-h$ because (1.34) is a non-negative definite function. The autocovariance function, $\gamma(h)$, of a stationary process is non-negative definite (see Problem 1.25) ensuring that variances of linear combinations of the variates x_t will never be negative. And, because $\mathrm{var}(a_1 x_{t_1} + \cdots + a_n x_{t_n})$ is never negative, the estimate of that variance should also be non-negative. The estimator in (1.34) guarantees this result, but no such guarantee exists if we divide by $n-h$; this is explored further in Problem 1.25. Note that neither dividing by n nor $n - h$ in (1.34) yields an unbiased estimator of $\gamma(h)$.

Definition 1.15 *The* **sample autocorrelation function** *is defined, analogously to (1.23), as*

$$\widehat{\rho}(h) = \frac{\widehat{\gamma}(h)}{\widehat{\gamma}(0)}. \tag{1.35}$$

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 1.1 Large-Sample Distribution of the ACF

Under general conditions,[5] *if* x_t *is white noise, then for* n *large, the sample ACF,* $\widehat{\rho}_x(h)$, *for* $h = 1, 2, \ldots, H$, *where* H *is fixed but arbitrary, is approximately normally distributed with zero mean and standard deviation given by*

$$\sigma_{\hat{\rho}_x(h)} = \frac{1}{\sqrt{n}}. \tag{1.36}$$

[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.

Based on the previous result, we obtain a rough method of assessing whether peaks in $\widehat{\rho}(h)$ 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.

Definition 1.16 *The estimators for the cross-covariance function, $\gamma_{xy}(h)$, as given in (1.26) and the cross-correlation, $\rho_{xy}(h)$, in (1.27) are given, respectively, by the* **sample cross-covariance function**

$$\widehat{\gamma}_{xy}(h) = n^{-1} \sum_{t=1}^{n-h} (x_{t+h} - \bar{x})(y_t - \bar{y}), \tag{1.37}$$

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

$$\widehat{\rho}_{xy}(h) = \frac{\widehat{\gamma}_{xy}(h)}{\sqrt{\widehat{\gamma}_x(0)\widehat{\gamma}_y(0)}}. \tag{1.38}$$

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 1.22 for the theoretical cross-covariance function. Because $-1 \leq \widehat{\rho}_{xy}(h) \leq 1$, 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 (1.29), we have the following property.

Property 1.2 Large-Sample Distribution of Cross-Correlation Under Independence

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

$$\sigma_{\hat{\rho}_{xy}} = \frac{1}{\sqrt{n}} \tag{1.39}$$

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

Example 1.23 A Simulated Time Series

To give an example of the procedure for calculating numerically the autocovariance and cross-covariance functions, consider a contrived set of data

Table 1.1. Sample Realization of the Contrived Series y_t

t	1	2	3	4	5	6	7	8	9	10
Coin	H	H	T	H	T	T	T	H	T	H
x_t	1	1	-1	1	-1	-1	-1	1	-1	1
y_t	6.7	5.3	3.3	6.7	3.3	4.7	4.7	6.7	3.3	6.7
$y_t - \bar{y}$	1.56	.16	-1.84	1.56	-1.84	$-.44$	$-.44$	1.56	-1.84	1.56

generated by tossing a fair coin, letting $x_t = 1$ when a head is obtained and $x_t = -1$ when a tail is obtained. Construct y_t as

$$y_t = 5 + x_t - .7x_{t-1}. \tag{1.40}$$

Table 1.1 shows sample realizations of the appropriate processes with $x_0 = -1$ and $n = 10$.

The sample autocorrelation for the series y_t can be calculated using (1.34) and (1.35) for $h = 0, 1, 2, \ldots$. It is not necessary to calculate for negative values because of the symmetry. For example, for $h = 3$, the autocorrelation becomes the ratio of

$$\begin{aligned}\widehat{\gamma}_y(3) &= \tfrac{1}{10}\sum_{t=1}^{7}(y_{t+3} - \bar{y})(y_t - \bar{y}) \\ &= \tfrac{1}{10}\Big[(1.56)(1.56) + (-1.84)(.16) + (-.44)(-1.84) + (-.44)(1.56) \\ &\qquad + (1.56)(-1.84) + (-1.84)(-.44) + (1.56)(-.44)\Big] = -.048\end{aligned}$$

to

$$\widehat{\gamma}_y(0) = \tfrac{1}{10}[(1.56)^2 + (.16)^2 + \cdots + (1.56)^2] = 2.030$$

so that

$$\widehat{\rho}_y(3) = \frac{-.048}{2.030} = -.024.$$

The theoretical ACF can be obtained from the model (1.40) 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). Table 1.2 compares the theoretical ACF with sample ACFs for a realization where $n = 10$ and another realization where $n = 100$; we note the increased variability in the smaller size sample.

Table 1.2. Theoretical and Sample ACFs for $n = 10$ and $n = 100$

		$n = 10$	$n = 100$
h	$\rho_y(h)$	$\widehat{\rho}_y(h)$	$\widehat{\rho}_y(h)$
0	1.00	1.00	1.00
± 1	$-.47$	$-.55$	$-.45$
± 2	.00	.17	$-.12$
± 3	.00	$-.02$	.14
± 4	.00	.15	.01
± 5	.00	$-.46$	$-.01$

~~**Example 1.24 ACF of a Speech Signal**~~

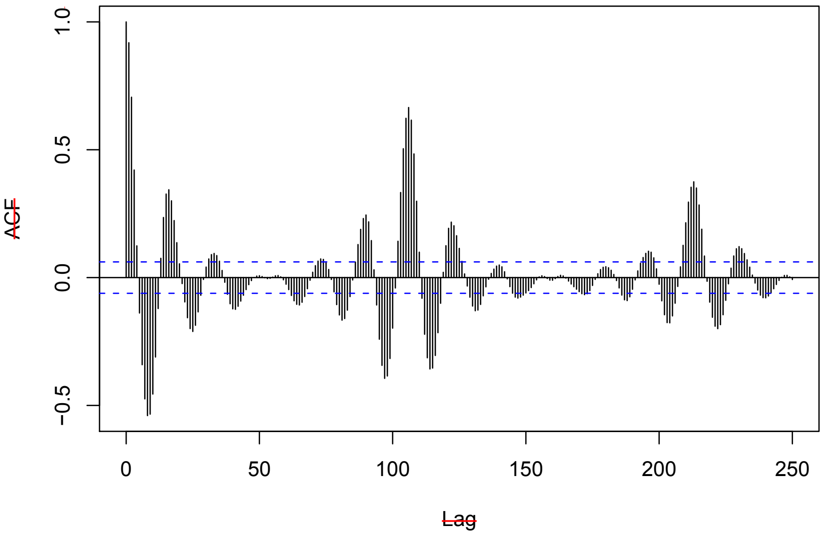
~~Computing the sample ACF as in the previous example can be thought of as matching the time series h units in the future, say, x_{t+h} against itself, x_t. Figure 1.13 shows the ACF of the speech series of Figure 1.3. The original series appears to contain a sequence of repeating short signals. The 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 put the data into `speech` as a time series object (if it is not there already from Example 1.3) and compute the sample ACF in R, use~~

```
acf(speech, 250)
```

Example 1.25 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 1.14 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 $24, 36, 48, \ldots$ 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. 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



~~**Fig. 1.13.** ACF of the speech series.~~

six months. The sign of the ACF 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 [see (1.36) and (1.39)].

To reproduce Figure 1.14 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")
```

1.7 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 $\boldsymbol{x}_t = (x_{t1}, x_{t2}, \dots, x_{tp})'$, which contains as its components p univariate time series. We denote the $p \times 1$ column vector of the observed series as $\boldsymbol{x}_t$. The row vector $\boldsymbol{x}_t'$ is its transpose. For the stationary case, the $p \times 1$ mean vector

$$\boldsymbol{\mu} = E(\boldsymbol{x}_t) \tag{1.41}$$

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

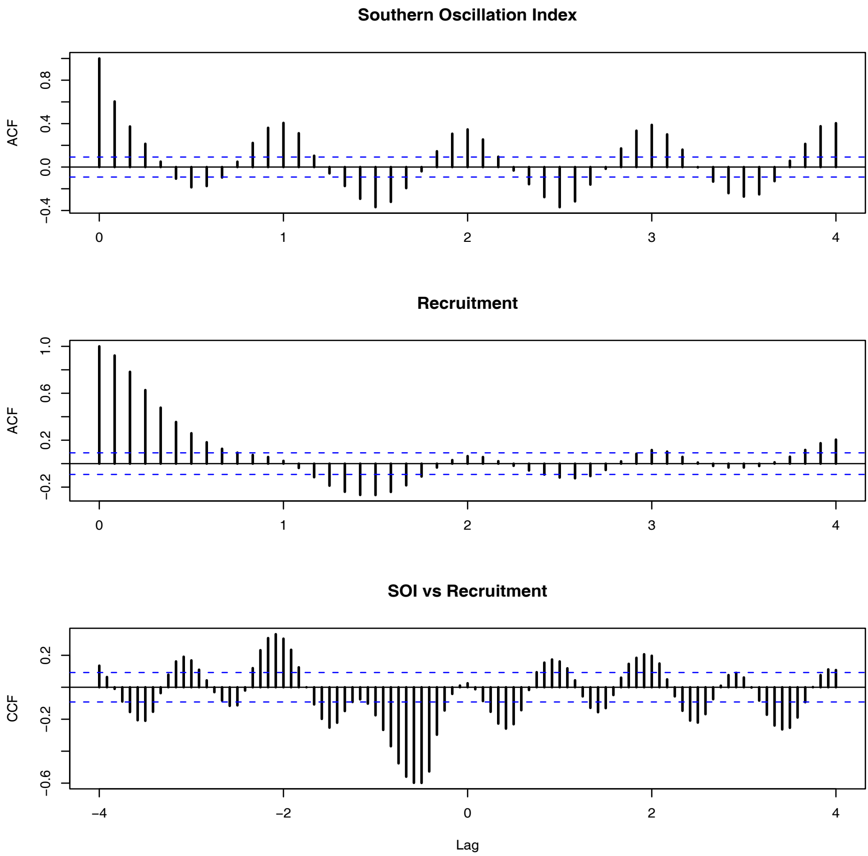


Fig. 1.14. Sample ACFs of the SOI series (top) and of the Recruitment series (middle), and the sample CCF of the two series (bottom); negative lags indicate SOI leads Recruitment. The lag axes are in terms of seasons (12 months).

$$\Gamma(h) = E[(\boldsymbol{x}_{t+h} - \boldsymbol{\mu})(\boldsymbol{x}_t - \boldsymbol{\mu})'] \tag{1.42}$$

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

$$\gamma_{ij}(h) = E[(x_{t+h,i} - \mu_i)(x_{tj} - \mu_j)] \tag{1.43}$$

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

$$\Gamma(-h) = \Gamma'(h). \tag{1.44}$$

Now, the sample autocovariance matrix of the vector series $\boldsymbol{x}_t$ is the $p \times p$ matrix of sample cross-covariances, defined as

$$\widehat{\Gamma}(h) = n^{-1} \sum_{t=1}^{n-h} (\boldsymbol{x}_{t+h} - \bar{\boldsymbol{x}})(\boldsymbol{x}_t - \bar{\boldsymbol{x}})', \tag{1.45}$$