Chapter 3

Autocovariance and Autocorrelation

If the $\{X_n\}$ process is weakly stationary, the covariance of X_n and X_{n+k} depends only on the lag k. This leads to the following definition of the "autocovariance" of the process:

$$\gamma(k) = \mathrm{cov}(X_{n+k},\, X_n) \tag{3.1}$$

Questions:

1. What is the variance of the process in terms of $\gamma(k)$?
2. Show $\gamma(0) \geq |\gamma(k)|$ for all k.
3. Show $\gamma(k)$ is an even function i.e. $\gamma(-k) = \gamma(k)$.

The autocorrelation function, $\rho(k)$, is defined by

$$\rho(k) = \frac{\gamma(k)}{\gamma(0)} \tag{3.2}$$

This is simply the correlation between X_n and X_{n+k}. Another interpretation of $\rho(k)$ is the optimal weight for scaling X_n into a prediction of X_{n+k} i.e. the weight, a say, that minimizes $E(X_{n+k} - aX_n)^2$.

There are additional constraints on $\gamma(k)$ beyond symmetry and maximum value at zero lag. To illustrate, assume X_n and X_{n+1} are perfectly correlated, X_{n+1} and X_{n+2} perfectly correlated. It is clearly nonsensical to set the correlation between X_n and X_{n+2} to zero.

To obtain the additional constraints define the $p \times 1$ vector

$$X' = [X_{n+1}, X_{n+2}, \dots, X_{n+p}]$$

and a set of constants

$$a' = [a_1, a_2, \dots a_p]$$

Consider now the following linear combination of the elements of X:

$$Z = \sum a_i X_i = a'X$$

The variance of Z is $a'\Sigma_{XX}a$ where Σ_{XX} is the $p \times p$ covariance matrix of the random vector X. (I'll discuss this in class.) If $\{X_n\}$ is a stationary random process Σ_{XX} has the following special form:

$$\Sigma_{XX} = \sigma^2 \begin{bmatrix} 1 & \rho(1) & \rho(2) & \dots & \rho(p-1) \\ \rho(1) & 1 & \rho(1) & \dots & \rho(p-2) \\ \rho(2) & \rho(1) & 1 & \ddots & \\ \vdots & & & \ddots & \rho(1) \\ \rho(p-1) & & & \rho(1) & 1 \end{bmatrix}$$

The variance of Z cannot be negative and so $a'\Sigma_{XX}a \geq 0$ for all nonzero a. We say Σ_{XX} is a "positive semi-definite". This implies

$$\sum_{r=1}^{n} \sum_{s=1}^{n} a_r a_s \rho(r-s) \geq 0$$

We will return to this condition later in the course after we have discussed power spectra.

3.1 Autocovariances of Some Special Processes

1. **The Purely Random Process**: This process has uncorrelated, zero-mean, random variables with constant variance. It is usually denoted by $\{\epsilon_n\}$ and its mean, variance and autocovariance are given by

$$\begin{aligned} E(\epsilon_n) &= 0 \\ \mathrm{var}(\epsilon_n) &= \sigma^2 \\ \rho(k) &= 0 \qquad k \neq 0 \end{aligned}$$

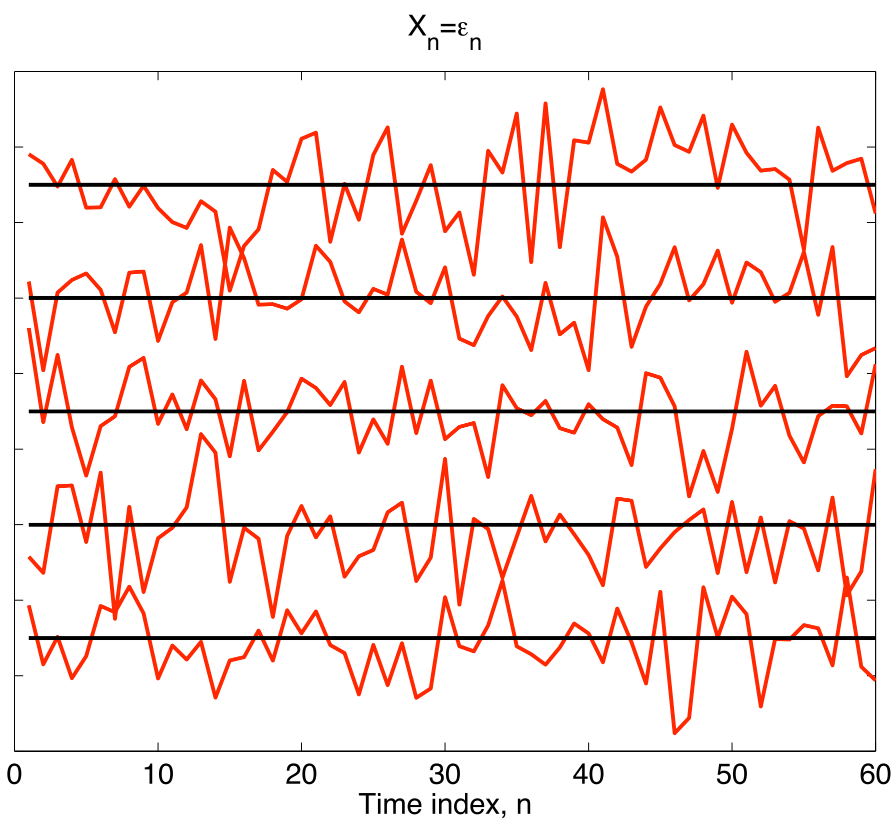

Figure 3.1: Five realizations from a purely random (Gaussian) process.

2. **The AR(1) Process**: This is the skater example with a purely random process for the forcing. The mean is zero, the variance is $\sigma^2/(1-a^2)$ and the autocorrelation is $a^{|k|}$.

3. **Sums and Differences**

   Question: What is the variance and autocorrelation of $X_n + X_{n-1}$ and $X_n - X_{n-1}$ if $\{X_n\}$ is a stationary random process with variance σ^2 and autocorrelation $\rho(k)$? Interpret for the special case of an AR(1) process as $a \to 1$.

4. **The Moving Average Process, MA(q)**: This is just a weighted average of a purely random process:

$$X_n = b_0\epsilon_n + b_1\epsilon_{n-1} + \ldots + b_q\epsilon_{n-q}$$

   The mean, variance and autocorrelation of the MA(p) process are given by

$$\begin{aligned} E(X_n) &= 0 \\ \mathrm{var}(X_n) &= \sigma^2(b_0^2 + b_1^2 + \ldots b_q^2) \end{aligned}$$

   and

$$\rho(k) = \frac{b_0 b_k + b_1 b_{k+1} + \ldots b_{q-k} b_q}{b_0^2 + b_1^2 + \ldots b_q^2} \qquad 1 \le k \le q$$

   and zero otherwise (see page 26, 104 of Shumway and Stoffer).

   Question: What is the variance and autocorrelation of the MA(q) process when the weights are equal? Plot the autocorrelation function.

3.2 Estimation of Autocovariance

Assume we have $2N+1$ observations[1] of the stationary random process $\{X_n\}$:

$$X_{-N},\ X_{-N+1},\ldots X_{-1},\ X_0,\ X_1,\ldots X_{N-1},\ X_N$$

The natural estimate of autocovariance is based on replacing *ensemble averaging by time averaging*:

$$\hat{\gamma}_{XX}(k) = \frac{1}{2N+1}\sum_{n=-N}^{N-k}(X_{n+k}-\overline{X})(X_n-\overline{X}) \qquad k \geq 0 \tag{3.3}$$

To simplify the following discussion of the properties of this estimator, assume the process has zero mean and use the following simplified estimator:

$$\hat{\gamma}(k) = \frac{1}{2N+1}\sum_{n=-N}^{N} X_{n+k}X_n \tag{3.4}$$

Question: Show $E\,\hat{\gamma}(k) = \gamma(k)$ i.e. the estimator is unbiased.

The covariance between the estimator for lag k_1 and k_2 is

$$\mathrm{cov}(\hat{\gamma}(k_1),\hat{\gamma}(k_2)) = E(\hat{\gamma}(k_1)\hat{\gamma}(k_2)) - \gamma(k_1)\gamma(k_2) \tag{3.5}$$

The first term on the right hand-side of (3.5) is the expected value of

$$\hat{\gamma}(k_1)\hat{\gamma}(k_2) = \frac{1}{(2N+1)^2}\sum_{u=-N}^{N}\sum_{v=-N}^{N} X_{u+k_1}X_uX_{v+k_2}X_v$$

[1] We assume an odd number of observations because it makes the math easier in the frequency domain.

How do we deal with these messy, quadruple products? Turns out that if the process is normal (i.e. any collection of $\{X_n\}$ has a multivariate normal distribution - see page 550 of Shumway and Stoffer - I'll discuss in class) we can take advantage of the following result:

If Z_1 through Z_4 have a multivariate normal distribution with zero mean:

$$E(Z_1Z_2Z_3Z_4) = E(Z_1Z_2)E(Z_3Z_4)+E(Z_1Z_3)E(Z_2Z_4)+E(Z_1Z_4)E(Z_2Z_3)$$

Using this result the quadruple product breaks down into simpler terms that can be described with autocovariance functions. It is then possible (see page 515 of Shumway and Stoffer, page 326 of Priestley)) to obtain the following approximation for large sample sizes:

$$\mathrm{cov}(\hat{\gamma}(k_1), \hat{\gamma}(k_2)) \approx$$

$$\frac{1}{2N+1} \sum_{s=-\infty}^{\infty} \gamma(s)\gamma(s+k_1-k_2) + \gamma(k_1+s)\gamma(k_2-s) \quad (3.6)$$

Setting $k_1 = k_2$ we get the variance of the sample autocovariance:

$$\mathrm{var}(\hat{\gamma}(k)) \approx \frac{1}{2N+1} \sum_{s=-\infty}^{\infty} \gamma^2(s) + \gamma(k+s)\gamma(k-s) \quad (3.7)$$

These are extremely useful formulae that allows us to estimate variance and covariances of the sample autocovariance functions. As we will see on the next page, they will help us avoid over-interpreting sample autocovariance functions.

Don't Over Interpret Wiggles in the Tails

The (asymptotic) covariance of the sample autocovariance is

$$\operatorname{cov}(\hat{\gamma}(k{+}m), \hat{\gamma}(k)) \approx \frac{1}{2N+1}\sum_{s=-\infty}^{\infty} \gamma(s)\gamma(s+m) + \gamma(k+m+s)\gamma(k-s)$$

Assume $\gamma(k) = 0$ for $k > K$, i.e. the system has limited memory. As we move into the tails ($k \to \infty, m > 0$) it is straightforward to show that the second product drops out (I'll draw some pictures in class) to give

$$\operatorname{cov}(\hat{\gamma}(k+m), \hat{\gamma}(k)) \to \frac{1}{2N+1}\sum_{s=-\infty}^{\infty} \gamma(s+m)\gamma(s)$$

Thus neighboring sample covariances will, in general, be correlated, even at lags where the true autocovariance is zero i.e. $k > K$. In other words we can expect the structure in $\hat{\gamma}(k)$ to "damp out" less quickly than $\gamma(k)$.

If we put $m = 0$ in this formula we obtain the (asymptotic) variance of the sample autocovariance in the tails. Note it too does not go to zero. (What does it tend to?)

Bottom line: Watch out for those wiggly tails in the autocovariance- they may not be real!

The following Matlab code generates a realization of length $N = 500$ from an AR(1) process with an autoregressive coefficient $a = 0.6$. The sample and theoretical autocovariances are then plot over each other. Use the above results to interpret this figure.

```
a=0.6; N=500; Nlags=50;

x=randn(1);
X=[x];
for n=1:N
  x=a*x+randn(1)*sqrt(1-a^2);
  X=[X; x'];
end

[Acf,lags]=xcov(X,Nlags,'biased');
plot(lags,Acf,'k-',lags,a.^abs(lags),'r-')
xlabel('Lags');

legend('Sample AutoCovariance','True AutoCovariance')
grid
```

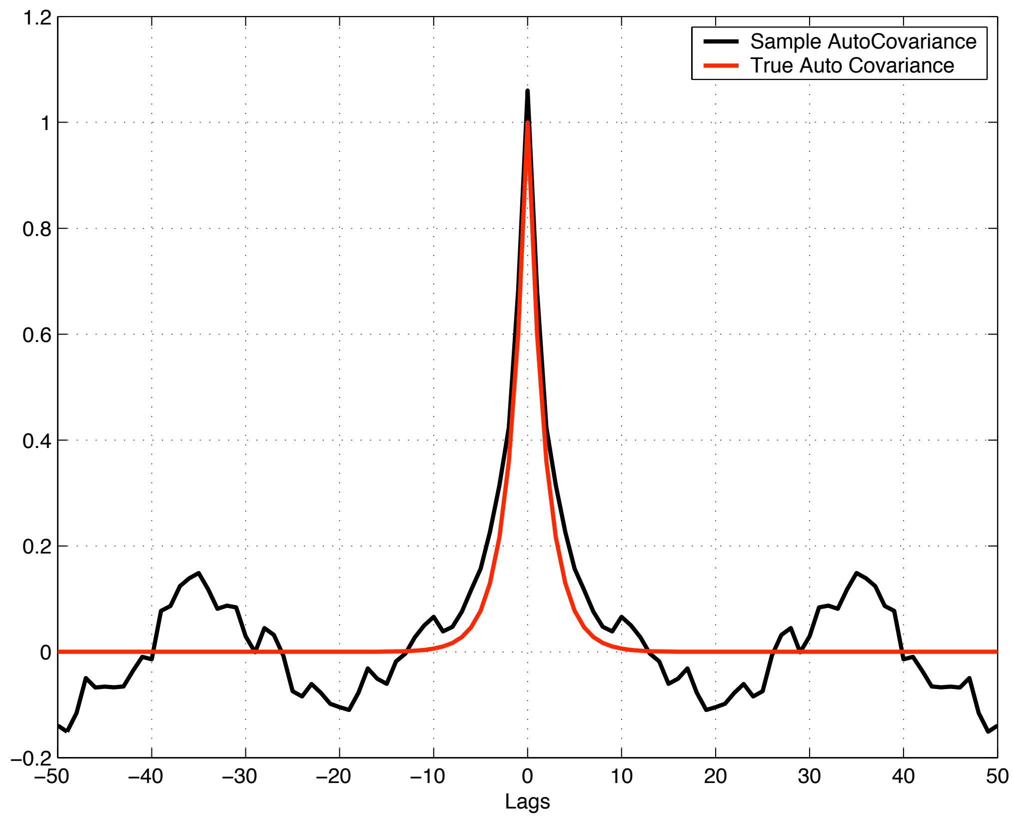

Figure 3.2: Sample and theoretical autocovariance functions for an AR(1) process with $a = 0.6$. The realization is of length $N = 500$. The important point to note is that the wiggles in the tails are due entirely to sampling variability - they are not real in the sense they do not reflect wiggles in the tails of the true autocovariance function, $\gamma(k)$.

3.3 Estimation of the Autocorrelation Function

The estimator for the theoretical autocorrelation, $\rho(k)$, is

$$\hat{\rho}(k) = \frac{\hat{\gamma}(k)}{\hat{\gamma}(0)}$$

where $\hat{\gamma}(k)$ is the sample autocovariance function discussed above.

This estimator is the ratio of two random quantities, $\hat{\gamma}(k)$ and $\hat{\gamma}(0)$, and so it should be no surprise that its sampling distribution is more complicated than that of $\hat{\gamma}(k)$.

If the process is Gaussian process and the sample size is large, it can be shown (page 519 of Shumway and Stoffer)

$$\begin{aligned}&\mathrm{cov}\left[\hat{\rho}(\mathrm{k+m}), \hat{\rho}(\mathrm{k})\right] \approx \\ &\quad \frac{1}{2N+1}\sum_{s=-\infty}^{\infty} \rho(s)\rho(s+m)+\rho(s+k+m)\rho(s-k)+2\rho(k)\rho(k+m)\rho^2(s) \\ &\qquad\qquad - 2\rho(k)\rho(s)\rho(s-k-m) - 2\rho(k+m)\rho(s)\rho(s-k)\end{aligned}$$

This is a messy formula (don't bother copying onto you exam cheat sheets) but it yields useful information, particularly when we make some simplifying assumptions.

Question: if $\{X_n\}$ is a purely random process show that for large sample sizes

$$\mathrm{var}\hat{\rho}(k) \approx \frac{1}{2N+1}$$

It can be shown more generally that, for large sample size, the sampling distribution of $\hat{\rho}(k)$ is approximately normal with mean $\rho(k)$ (the estimator is unbiased) and variance by the above equation with $m = 0$. If we evaluate the variance of $\hat{\rho}(k)$ by substituting the sample autocorrelation or a fitted model, we can calculate confidence intervals for $\rho(k)$ in the usual way.

We can also calculate approximate 5% significance levels for zero correlation using $\pm 1.96/\sqrt{2N+1}$. (Be careful of the problems of multiple testing however.)

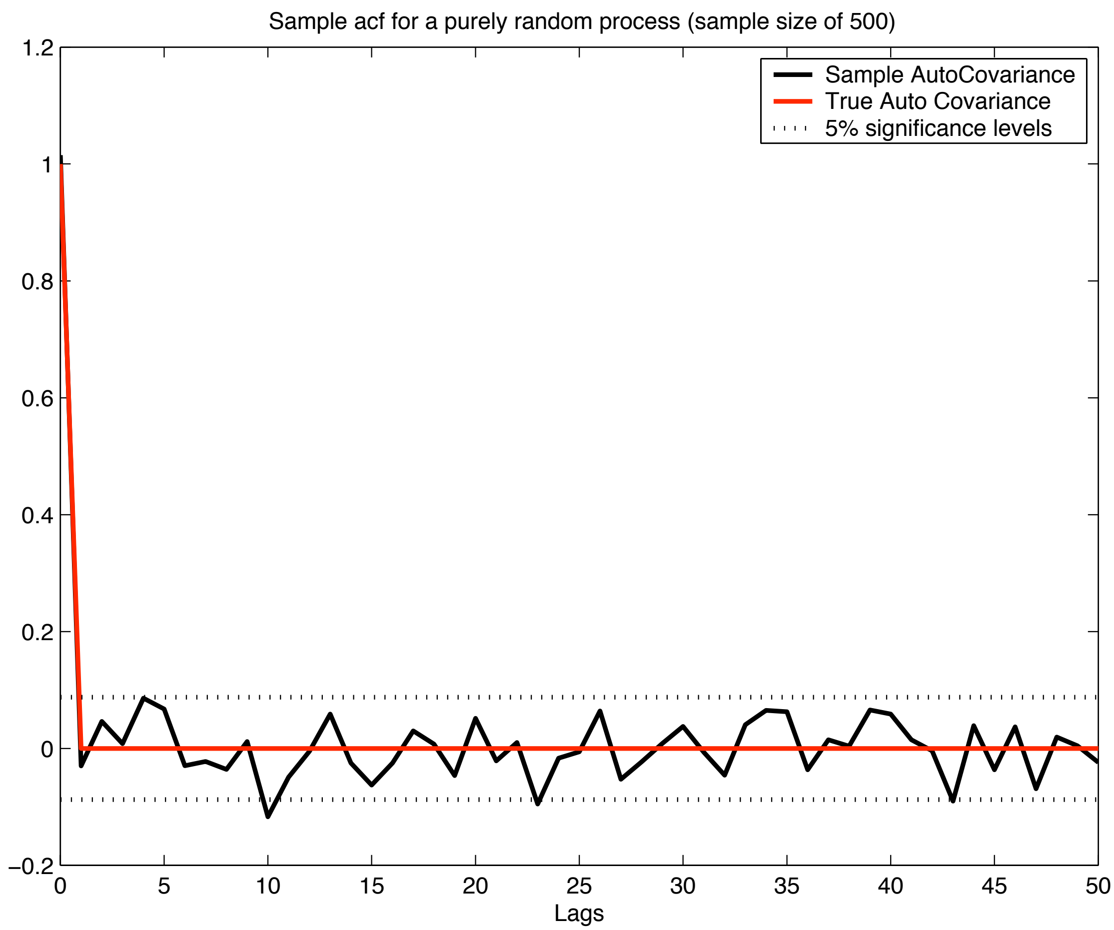

Figure 3.3: Sample auto correlation functions for a purely random process. The realization is of length 500. The dotted lines show the 5% significance levels of zero correlation.