8 Covariance Stationary Time Series

So far in the course we have looked at what we have been calling time series data sets. We need to make a series of assumptions about our data set in order to accomplish the aims of our analysis. We will do this in analogy with making inferences about a population (confidence intervals and tests of hypothesis for means and variances, and so on) from a sample in introductory statistics courses.

For example, in introductory statistics we get one sample $X_1, \ldots, X_n$ of size n from a population. We assume the population has mean μ and variance σ^2. If we assume the population has a normal distribution, we get $100(1-\alpha)$% confidence intervals for μ and σ^2 as

$$\bar{X} \pm t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}, \qquad \left(\frac{(n-1)s^2}{\chi^2_{\alpha/2,n-1}}, \frac{(n-1)s^2}{\chi^2_{1-\alpha/2,n-1}}\right),$$

where $\bar{X}$ and s^2 are the mean and variance of the sample and $t_{\beta,v}$ and $\chi^2_{\beta,v}$ are the values of t and χ^2 random variables with v degrees of freedom having area β to their right, that is, the $100(1-\beta)$ percentiles of the t and χ^2 distributions.

These are called $100(1-\alpha)$% confidence intervals because if we took random samples many, many times and got the intervals each time, then $100(1-\alpha)$% of the samples would give intervals containing the true value of the parameter being estimated.

In time series analysis, a data set is not a random sample from a population, so how do we end up getting confidence intervals and tests of hypotheses for parameters?

8.1 The Ensemble of Realizations

The first step in our logic that will allow us to do inferential statistics with time series is as follows. We imagine that our data set is just part of a *realization* that lasts from far into the past until far into the future. Further, we imagine that the realization we are observing is just one of many, many realizations that we could have observed. Here are two examples of how this thought process makes sense:

1. If we are observing an EEG record, it would make sense to imagine that the realization we are looking at today is similar to, but not identical to, a record we could observe at some future day. We can imagine that there is some random mechanism generating realizations. Another way to think of this is that a person's EEG is very long and we are looking at *independent* pieces of it.

2. If we must measure our time series variable with error, then immediately it makes sense to think of our realization as just one of many we could see.

8.2 Definitions

1. The set of all possible realizations that could be observed is called the ensemble of realizations. The set of possible values at a particular time point t is denoted by $X(t)$, and a time series is denoted by $\{X(t), t \in T\}$. A time series data set is one part of one realization and is denoted by $x(1), \ldots, x(n)$.

2. We denote by

$$\mu(t) = \mathsf{E}(X(t)), \qquad K(s,t) = \mathsf{Cov}(X(s), X(t)),$$

the mean and the covariance functions of X.

3. If $\mu(t)$ is the same for each t and $K(s,t)$ only depends on how far apart s and t are (that is $K(s,t) = R(|t-s|)$ for a function R called the autocovariance function of X), then we say that X is covariance stationary with mean μ and autocovariance function R and we have:

   (a) The autocorrelation function of X is given by

$$\rho(v) = \mathsf{Corr}(X(t), X(t+v)),$$

   and the spectral density function $f(\omega), \omega \in [0,1]$ is given by

$$f(\omega) = \sum_{v=-\infty}^{\infty} R(v) e^{-2\pi i v \omega}, \quad \omega \in [0,1].$$

4. A white noise time series $\{\epsilon(t), t \in Z\}$ is a time series satisfying $\mathsf{E}(\epsilon(t)) = 0$, $R(v) = \delta_v \sigma^2$, where the Kroneker delta function δ_v is 1 if $v = 0$ and zero otherwise. We denote a white noise time series by $X \sim WN(\sigma^2)$.

5. A time series model is a mathematical formula expressing how the realizations of the series are formed. For example, a moving average model of order one with coefficient β and noise variance σ^2 (which we denote by $X \sim MA(1, \beta, \sigma^2)$) means

$$X(t) = \epsilon(t) + \beta\epsilon(t-1), \qquad t \in Z,$$

where $\epsilon \sim WN(\sigma^2)$.

8.3 Ensemble Mean Interpretation of ρ and f

We spent much of the first part of this course studying the sample correlogram $\hat{\rho}$ and sample spectral density function $\hat{f}$ (which at the frequencies $0, 1/n, 2/n, \ldots$ is the periodogram). The "true" or "population" correlogram ρ and spectral density function f can under most circumstances be thought of as the average value of $\hat{\rho}$ and $\hat{f}$ where the average is taken over all realizations in the ensemble of realizations.

Thus these are the quantities we would really like to know about. For EEG data, for example, $\hat{\rho}$ and $\hat{f}$ are calculated only for the realization we have observed, but what we really would like to know about would be the average of these quantities over many realizations.

8.4 Rules for Expectation, Variance, and Covariance

It is important to distinguish the mean $\mu = \mathsf{E}(X)$, $\sigma^2 = \mathsf{Var}(X)$, and $\rho = \mathsf{Corr}(X, Y)$ for random variables from $\bar{X}$, s^2, and r for data. One way to think of this is that μ, σ^2, and ρ are defined for populations (such as the $X(t)$'s) while $\bar{X}$, $\hat{R}(0)$, and $\hat{\rho}(v)$ are defined for a data set.

Mathematically, the quantities μ, $R(v)$, and $\rho(v)$ are calculated via integration from assumed joint probability distributions for the "populations" $\{X(t), t \in Z\}$, for example, if we assume normal distributions for the $X(t)$'s, we have

$$\mu = \int x f(x) dx,$$

where f is the pdf of the normal distribution. In this course, we will not need to do such integration but rather use a set of simple rules for calculating parameters from models. You can think of $X(t)$ either as a population, or more mathematically, as a random variable.

Note that quantities such as μ, $R(v)$, and $\rho(v)$ are called *moments* of the random variables.

1. The variance of a random variable is defined to be

   $$\mathsf{Var}(X) = \mathsf{E}\left((X - \mathsf{E}(X))^2\right) = \mathsf{E}(X^2) - \left(\mathsf{E}(X)^2\right),$$

   which means

   $$\mathsf{Var}(X) = \mathsf{E}(X^2)$$

if $\mathsf{E}(X) = 0$.

2. The covariance of two random variables is defined to be

$$\text{Cov}(X, Y) = \mathsf{E}\left((X - \mathsf{E}(X))(Y - \mathsf{E}(Y))\right),$$

which means

$$\text{Cov}(X, Y) = \mathsf{E}(XY)$$

if the means of X and Y are zero.

3. The correlation of random variables X and Y is given by

$$\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}.$$

4. The mean and variance of a constant times a random variable are given by

$$\mathsf{E}(aX) = a\mathsf{E}(X), \qquad \text{Var}(aX) = a^2\text{Var}(X).$$

5. The mean of the sum of random variables is the sum of the means:

$$\mathsf{E}(X + Y) = \mathsf{E}(X) + \mathsf{E}(Y).$$

6. The variance of the sum of random variables is not generally the sum of the variances:

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y).$$

If $\text{Cov}(X, Y) = 0$, the variance of a sum is the sum of the variances.

8.5 Application of Moment Rules to Time Series Models

We illustrate the moment rules on three simple time series models. In the next topic, we study some more general rules that will greatly simplify determining moments for a set of complex models.

White Noise

The white noise model says that our time series is just a set of zero mean (that is, E$(X(t)) = 0$ for all t), uncorrelated (that is, Corr$(X(t), X(t+v)) = 0$ unless $v = 0$) random variables with constant variance $R(0) = \sigma^2$.

Thus we have for all $\omega \in [0, 1]$,

$$f(\omega) = \sum_{v=-\infty}^{\infty} R(v) e^{-2\pi i v \omega} = R(0)e^0 = R(0) = \sigma^2,$$

which shows why we call such a time series model white noise; it is often used to model "noise," and its "spectrum" is constant for all frequencies in analogy with white light.

Random Walk

A time series is said to follow a random walk model if

$$X(t) = X(t-1) + \epsilon(t), \quad t \geq 1,$$

where the starting value $X(0)$ has mean zero, variance σ_X^2, $\epsilon \sim WN(\sigma^2)$, and $X(0)$ is uncorrelated with all of the ϵ's (we will see why we can't assume the process started in the inifinite past, that is, why t must start at some time such as time 0).

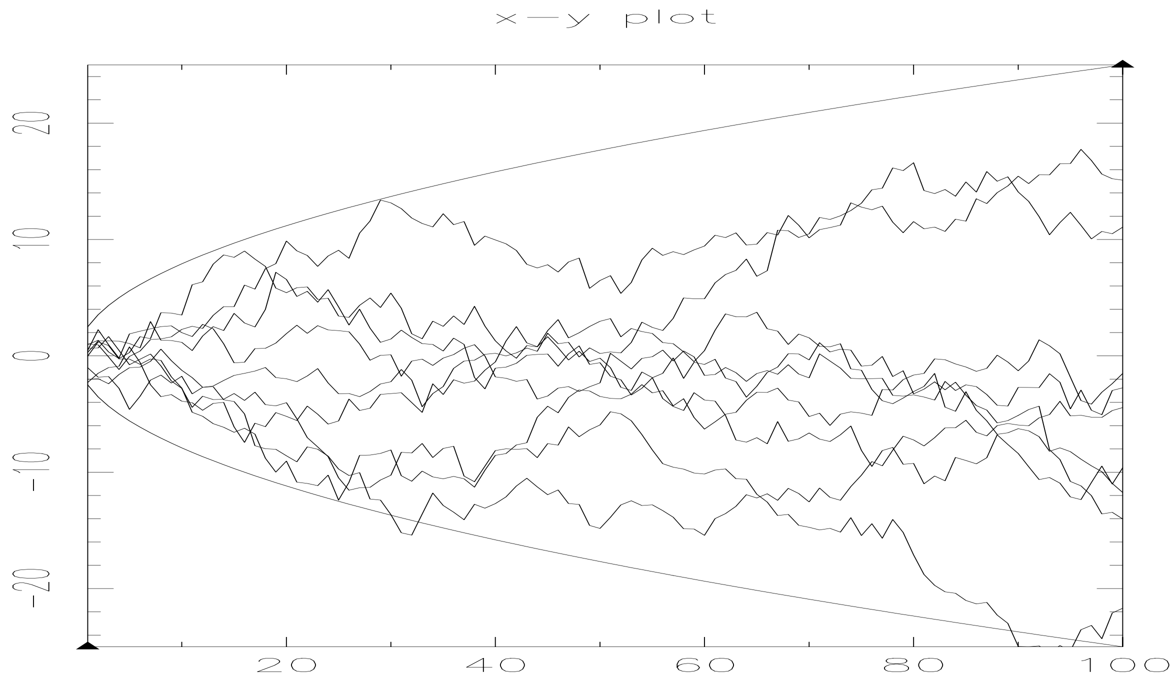
By successive substitution we can write

$$X(t) = X(0) + \sum_{j=1}^{t} \epsilon(j),$$

and so

$$\text{Var}(X(t)) = \sigma_X^2 + t\sigma^2,$$

which shows that a random walk is not covariance stationary since the variance at time t does depend on t. In fact this variance is growing linearly with t (see the figure below which shows 10 realizations of length 100 from a random walk where $X(0) = 0$ and the ϵ's are N(0,1); the values of $\pm 2.5\sqrt{\text{Var}(X(t)} = \pm 2.5\sqrt{t}$ are also plotted), which means that if the process started in the infinite past, the variance would have become infinite by time 0.


x–y plot
20
10
0
–10
–20
20
40
60
80
100

Moving Average Process

If

$$X(t) = \epsilon(t) + \beta\epsilon(t-1), \quad t \in Z,$$

where $\epsilon \sim WN(\sigma^2)$, then

$$\mathsf{E}(X(t)) = \mathsf{E}(\epsilon(t)) + \beta\mathsf{E}(\epsilon(t-1)) = 0 + \beta 0 = 0,$$

so the first requirement for covariance stationarity is met; namely the mean (which is zero for all t) is constant over time. Now we have to see if $\mathsf{Cov}(X(t), X(t+v))$ only depends on v no matter which t we use. We have

$$\mathsf{Cov}(X(t), X(t+v)) = \mathsf{E}(X(t)X(t+v)),$$

since $\mathsf{E}(X(t) = 0$ for all t. Now we have

$$\mathsf{E}(X(t)X(t+v)) = \mathsf{E}([\epsilon(t)+\beta\epsilon(t-1)][\epsilon(t+v)+\beta\epsilon(t+v-1)]),$$

which becomes the sum of four expectations (using the First, Outside, Inside, and Last or FOIL rule):

$$\begin{aligned}\mathsf{E}(\epsilon(t)\epsilon(t+v)) \quad &+ \quad \beta\mathsf{E}(\epsilon(t)\epsilon(t+v-1)) + \beta\mathsf{E}(\epsilon(t-1)\epsilon(t+v)) \\ &+ \quad \beta^2\mathsf{E}(\epsilon(t-1)\epsilon(t+v-1)).\end{aligned}$$

The expected value of the product of two ϵ's can only take on two values; if the arguments are the same, then the expectation is σ^2, while if they are different then, the expectation is zero.

Thus we have

$$\text{Cov}(X(t), X(t+v)) = \begin{cases} \sigma^2(1+\beta^2), & \text{if } v = 0 \\ \sigma^2\beta, & \text{if } v = \pm 1 \\ 0, & \text{if } |v| > 1 \end{cases}$$

which shows an MA(1) process is covariance stationary with this covariance as $R(v)$ which gives

$$\rho(v) = \frac{R(v)}{R(0)} = \begin{cases} 1, & \text{if } v = 0 \\ \beta/(1+\beta^2), & \text{if } v = \pm 1 \\ 0, & \text{if } |v| > 1. \end{cases}$$

Thus to get the spectral density function f, we have since all of the values of $R(v)$ are zero except for $v = 0$ and $v = \pm 1$:

$$\begin{aligned} f(\omega) &= R(0) + R(1)e^{-2\pi i\omega} + R(-1)e^{2\pi i\omega} \\ &= R(0) + 2R(1)\cos(2\pi\omega) \\ &= \sigma^2((1+\beta^2) + 2\beta\cos(2\pi\omega)), \end{aligned}$$

since $R(1) = R(-1)$ and $e^{i\theta} + e^{-i\theta} = 2\cos(\theta)$.

Thus the spectral density of an MA(1) is a constant plus another constant (which has the same sign as that of β) times a cosine that goes through half a cycle for $\omega \in [0, 0.5]$, which means f can only look two ways; if $\beta > 0$, it has an excess of low frequency, while if $\beta < 0$, it has an excess of high frequency. Finally, if $\beta = 0$, we have that X is white noise, that is,

$$MA(1, \beta = 0, \sigma^2) = WN.$$

