Spring 2013 Statistics 153 (Time Series) : Lecture Four

Aditya Guntuboyina

29 January 2015

1 Dealing with Both Trend and Seasonality

In the last class, we studied time series models for datasets having pure trend and pure seasonality. There are three basic methods: parametric function fitting, smoothing and differencing. Differencing is the easiest of the the three techniques and works well in practice.

We shall now focus attention on datasets having both trend and seasonality. We fit additive trend and seasonality models: $X_t = m_t + s_t + W_t$ to the data. The function m_t is assumed to capture the trend (it has no seasonal component to it) and s_t is assumed to capture the seasonality (it has no trend component).

Again we follow the same three approaches (easiest being differencing).

1.1 Fit Parametric Functions

Fit a model of the form $X_t = m_t + s_t + W_t$ to the data. Use a linear or quadratic function for m_t and a sinusoid for s_t.

1.2 Smoothing

This is slightly involved.

Suppose we want to estimate m_t for December, 2011. The smoothing method suggests taking a local average of X values near December, 2011. How near? If we, for instance, take an average from Oct, 2011 to Mar, 2012, then we might get a value which is affected by the fact that the average is over the winter period. To overcome this fact, the sensible thing to do is average values either from June, 2011 to May, 2012 or from July, 2011 to June 2012. Both of these are viable options and the seasonality effect is now taken care of because all months (and hence all seasons) are represented equally in this average. There is still the minor question of which of these two averages to use. One can just take a further average of these two averages. This corresponds to taking a weighted average with a weight of 0.5/12 to June, 2011 and June, 2012 and a weight of 1/12 to the rest of the months. In other words, we estimate m_t by

$$\hat{m}_t := \frac{0.5X_{t-d/2} + X_{t-(d/2)+1} + \cdots + X_{t+(d/2)-1} + 0.5X_{t+d/2}}{d}.$$

The observations $Y_t = X_t - \hat{m}_t$ then correspond to de-trended data. One can then average $Y_1, Y_{1+d}, Y_{1+2d}, \dots$ to estimate s_1 and similarly for the other seasonality parameters.

1.3 Differencing

The data $X_t - \hat{m}_t - \hat{s}_t$ do not have any trend or seasonality. One then fits a stationary model to this dataset. Another way to obtain such data with no trend or seasonality is to do differencing.

Because the trend function satisfies $s_t = s_{t-d}$, the difference $X_t - X_{t-d}$ equals $m_t - m_{t-d} + W_t - W_{t-d}$ and hence this differenced data does not have any seasonal component. The trend $m_t - m_{t-d}$ can then be eliminated using the methods previously studied, in particular, by differencing.

2 Stationary Time Series

The concept of a stationary time series is the most important thing in this course.

Stationary time series are typically used for the residuals after trend and seasonality have been removed.

Stationarity allows a systematic study of time series forecasting.

In order to avoid confusion, we shall, from now on, make a distinction between the observed time series data $x_1, \dots, x_n$ and the sequence of random variables $X_1, \dots, X_n$. The observed data $x_1, \dots, x_n$ is a realization of $X_1, \dots, X_n$. It is convenient of think of the random variables $\{X_t\}$ as forming a doubly infinite sequence:

$$\dots, X_{-2}, X_{-1}, X_0, X_1, X_2, \dots.$$

The notion of stationarity will apply to the doubly infinite sequence of random variables $\{X_t\}$. Strictly speaking, stationarity does not apply to the data $x_1, \dots, x_n$. However, one frequently abuses terminology and uses stationary data to mean that the random variables of which the observed data are a realization have a stationary distribution.

Stationary essentially means that the dependence is invariant over time.

Definition 2.1 (Strict or Strong Stationarity)**.** *A doubly infinite sequence of random variables $\{X_t\}$ is **strictly stationary** if the joint distribution of $(X_{t_1}, X_{t_2}, \dots, X_{t_k})$ is the same as the joint distribution of $(X_{t_1+h}, X_{t_2+h}, \dots, X_{t_k+h})$ for every choice of times $t_1, \dots, t_k$ and h.*

Roughly speaking, stationarity means that the joint distribution of the random variables remains constant over time. For example, under stationarity, the joint distribution of today's and tomorrow's random variables is the same as the joint distribution of the variables from *any* two successive days (past or future).

Note how stationarity makes the problem of forecasting or prediction feasible. From the data, we can study how a particular day's observation depends on the those of the previous days and because under stationarity, such a dependence is assumed to be constant over time, one can hope to use it to predict future observations from the current data.

Many of the things that we shall be doing with stationarity actually go through even with the following notion that is weaker than strong stationarity.

Definition 2.2 (Second-Order or Weakly or Wide-Sense Stationarity)**.** *A doubly infinite sequence of random variables $\{X_t\}$ is **weak stationary** if*

1. *The mean of the random variable X_t, denoted by $\mathbb{E}X_t$, is the same for all times t.*
2. *The covariance between X_{t_1} and X_{t_2} is the same as the covariance between X_{t_1+h} and X_{t_2+h} for every choice of times t_1, t_2 and h.*

Weak stationarity means that the *second order properties* (means and covariances) of the random variables remain constant over time. Unlike strong stationarity, the joint distribution of the random variables may well change over time.

Another way of phrasing the condition

$$\mathrm{cov}(X_{t_1}, X_{t_2}) = \mathrm{cov}(X_{t_1+h}, X_{t_2+h}) \qquad \text{for all } t_1, t_2 \text{ and } h$$

is that the covariance between two random variables only depends on the *time lag* between them. In other words, the covariance between X_{t_1} and X_{t_2} only depends on the time lag $|t_1 - t_2|$ between them. Thus, if we define $\gamma(h)$ to be the covariance between the random variables corresponding to any two times with a lag of h, we have

$$\mathrm{cov}(X_{t_1}, X_{t_2}) = \gamma(|t_1 - t_2|) \qquad \text{for all times } t_1 \text{ and } t_2$$

as a consequence of weak stationarity.

The function $\gamma(h)$ is called the **Autocovariance Function** of the stationary sequence $\{X_t\}$. The notion of autocovariance function only applies to a stationary sequence of random variables but not to data. The Autocovariance function is abbreviated to *acvf*.

The variance of each random variable X_t is given by $\gamma(0)$. By stationarity, all random variables have the same variance.

Let $\rho(h)$ denote the correlation between two random variables in the stationary sequence $\{X_t\}$ that are separated by a time lag of h. Because $\gamma(h)$ denotes the corresponding covariance, it follows that

$$\rho(h) := \frac{\gamma(h)}{\gamma(0)}.$$

The function $\rho(h)$ is called the **Autocorrelation Function** of the stationary sequence $\{X_t\}$. Once again, this is a notion that only applies to random variables but not to data. On the contrary, the Sample Autocorrelation Function that we looked at before only applies to data. The Autocorrelation function is abbreviated to *acf*.

Because $\rho(h)$ is a correlation, it follows that $|\rho(h)| \leq 1$. Also $\rho(0)$ equals 1.

Definition 2.3 (Gaussian Process). *The sequence $\{X_t\}$ is said to be **gaussian** if the joint distribution of $(X_{t_1}, \dots, X_{t_k})$ is multivariate normal for every choice of times $t_1, \dots, t_k$.*

Important Note: $(X_{t_1}, \dots, X_{t_k})$ is multivariate normal means that every **linear combination** of $(X_{t_1}, \dots, X_{t_k})$ is univariate normal. In particular, it is much stronger than saying that each of $X_{t_1}, \dots, X_{t_k}$ has a univariate normal distribution. If $X_{t_1}, \dots, X_{t_k}$ are **independent** normal random variables, then their joint distribution is an example of a multivariate normal distribution.

The following statement is a consequence of the fact that multivariate normal distributions are determined by their means and covariances:

$$\text{weak stationarity} + \text{Gaussian Process} \implies \text{Strong Stationarity}.$$

In the rest of this course, stationarity would always stand for weak stationarity. Unless explicitly mentioned, do not assume that a stationary series is strongly stationary.

2.1 Examples

2.1.1 White Noise

The random variables $\ldots, X_{-2}, X_{-1}, X_0, X_1, X_2, \ldots$ are said to be *white noise* if they have mean zero and the following covariance:

$$\begin{aligned} \operatorname{cov}(X_{t_1}, X_{t_2}) &= \sigma^2 \qquad \text{if } t_1 = t_2 \\ &= 0 \qquad \text{if } t_1 \neq t_2. \end{aligned} \tag{1}$$

In other words, a the random variables in a white noise series are uncorrelated, have mean zero and a constant variance.

This is clearly a stationary series. What is its acf?

The white noise series is only a very special example of stationarity. Stationarity allows for considerable dependence between successive random variables in the series. The only requirement is that the dependence should be constant over time.

2.1.2 Moving Average Process of Order 1

Given a white noise series Z_t with variance σ^2 and a number θ, set

$$X_t = Z_t + \theta Z_{t-1}.$$

This is called a *moving average* of order 1. The series is stationary with mean zero and acvf:

$$\begin{aligned} \gamma_X(h) &= \sigma^2(1+\theta^2) \qquad \text{if } h = 0 \\ &= \theta\sigma^2 \qquad \text{if } h = 1 \\ &= 0 \qquad \text{otherwise.} \end{aligned} \tag{2}$$

As a consequence, X_{t_1} and X_{t_2} are uncorrelated whenever t_1 and t_2 are two or more time points apart. This time series has *short memory.*

The autocorrelation function, acf, for $\{X_t\}$ is given by

$$\rho_X(h) = \frac{\theta}{1+\theta^2}$$

for $h = 1$ and 0 for $h > 1$. What is the maximum value that $\rho_X(1)$ can take?