

Time Series Analysis

Andrew Siegel

April 3, 2017

Outline

Time Series Analysis

Introduction

Ensembles

Stationarity

Regression

Estimators

AR(p) Processes

MA(q) Processes

Outline

Time Series Analysis

Introduction

Ensembles

Stationarity

Regression

Estimators

AR(p) Processes

MA(q) Processes

Time Series

Introduction

- ▶ A *time series* is a set of n random variables x_1, x_2, \dots, x_n taken at times t_1, t_2, \dots, t_n that are described by an n -dimensional joint PDF:

$$F(c_1, c_2, \dots, c_n) = P\{x_1 \leq c_1, x_2 \leq c_2, \dots, x_n \leq c_n\}$$

- ▶ It is typically more practical and useful to consider the n marginal CDFs and PDFs:

$$F_t(x) = P\{x_t \leq x\} \tag{1}$$

$$f_t(x) = \frac{\partial F_t(x)}{\partial x} \tag{2}$$

Time Series

Mean Function

- ▶ The *mean function* of the random variable x_t is:

$$\mu_t = E(x_t) \equiv \int_{-\infty}^{\infty} x f_t(x) dx$$

- ▶ Example: The mean function of a random walk with drift is:

$$x_t = \delta t + \sum_{j=1}^t w_j$$

where δ is the drift (a constant) and w_j are white noise terms. Since $E(w_j) = 0$, then the mean is:

$$\mu_{xt} = E(x_t) = \delta t$$

- ▶ So in the most general case, the mean function itself is a function of time.

Introduction

Mean Function

- ▶ The mean function of smoothed white noise w_t is

$$\mu_{v_t} = E(v_t) = \frac{1}{3}[E(w_{t-1}) + E(w_t) + E(w_{t+1})] = 0$$

- ▶ The mean function of a signal plus noise is, e.g.

$$\mu_{x_t} = E(x_t) = E[2\cos(2\pi t/50 + .6\pi) + w_t] \quad (3)$$

$$= 2\cos(2\pi t/50 + .6\pi) + E(w_t) \quad (4)$$

$$= \cos(2\pi t/50 + .6\pi) \quad (5)$$

Introduction

Autocovariance

- ▶ The autocovariance function describes the linear dependence between two time points t and s in the same timeseries:

$$\gamma_x(s, t) = \text{cov}(x_s, x_t) = E[(x_s - \mu_s)(x_t - \mu_t)]$$

- ▶ Note that $\gamma_x(t, t) = E[(x_t - \mu_t)^2] = \text{var}(x_t)$. Also note that:

$$|\gamma(s, t)| \leq \sqrt{\gamma(s, s)\gamma(t, t)}$$

- ▶ Considering these bounds, we can define the autocorrelation function (ACF) as the normalized autocovariance:

$$\rho(s, t) = \frac{\gamma(s, t)}{\sqrt{\gamma(s, s)\gamma(t, t)}}$$

- ▶ For examples, see Example 1.16 (autocovariance of white noise) and 1.17 (autocovariance of a moving average) in Shumway and Stoffer

Introduction

Autocovariance

- Consider applying a three-point moving average to the white noise series w_t

$$\begin{aligned}\gamma(s, t) &= \text{cov}(v_s, v_t) \\ &= \text{cov}\left\{\frac{1}{3}(w_{s-1} + w_s + w_{s+1}), \frac{1}{3}(w_{t-1} + w_t + w_{t+1})\right\}.\end{aligned}$$

- When $s = t$ we have

$$\begin{aligned}\gamma_v(t, t) &= \frac{1}{9} \text{cov}\{(w_{t-1} + w_t + w_{t+1}), (w_{t-1} + w_t + w_{t+1})\} \\ &= \frac{1}{9} [\text{cov}(w_{t-1}, w_{t-1}) + \dots \\ &= \frac{3}{9} \sigma_w^2\end{aligned}$$

Introduction

Autocovariance, cont.

- ▶ When $s = t + 1$

$$\begin{aligned}\gamma_v(t+1, t) &= \frac{1}{9} \text{cov}\{(w_t + w_{t+1} + w_{t+2}), (w_{t-1} + w_t + w_{t+1})\} \\ &= \frac{1}{9} [\text{cov}(w_t, w_t) + \text{cov}(w_{t+1}, w_{t+1})] \\ &= \frac{2}{9} \sigma_w^2\end{aligned}$$

- ▶ Similarly, when $|s - t| = 2$, $\gamma_v(t+2, t) = \frac{1}{9} \sigma_w^2$, so

$$\gamma_v(h) = \begin{cases} \frac{3}{9} \sigma_w^2 & h = 0, \\ \frac{2}{9} \sigma_w^2 & h = \pm 1, \\ \frac{1}{9} \sigma_w^2 & h = \pm 2, \\ 0 & |h| > 2 \end{cases}$$

Outline

Time Series Analysis

Introduction

Ensembles

Stationarity

Regression

Estimators

AR(p) Processes

MA(q) Processes

Time Series

Ensembles

- ▶ Question: how to we estimate the mean function μ_t and form the sample sum

$$\hat{\gamma}(h) = \frac{1}{n} \sum_{t=1}^{n-h} (x_{t+h} - \bar{x})(x_t - \bar{x})$$

in practice?

- ▶ If you have many realizations of a timeseries, you can estimate an ensemble average for each x_t by averaging over *realizations*.
- ▶ Suppose you have $j = 1, \dots, m$ realizations of a timeseries, and each realization has $t = 1, \dots, n$ sample points.
- ▶ Then for each sample point, the estimated ensemble average is:

$$\mu_t \approx \bar{x}_t = \sum_{j=1}^m x_{tj}$$

- ▶ Often ensembles are not obtainable in practice and we have to use other approaches.

Outline

Time Series Analysis

Introduction

Ensembles

Stationarity

Regression

Estimators

AR(p) Processes

MA(q) Processes

Time Series

Estimation of Correlation

- ▶ When analyzing real data we typically do not have access to large ensembles and certainly do not know the PDFs
- ▶ thus we have to estimate statistical quantities from sample data
- ▶ For a stationary time series the mean function is constant and we estimate with the sample mean:

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

Time Series

Error in Sample Mean for Correlated Samples

- ▶ the standard error is the square root of $var(\bar{x})$

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

- ▶ Note that this reduces to familiar $\frac{\sigma_x^2}{n}$ in the case of independent samples.

Time Series

Estimation of Correlation

- ▶ The *sample autocovariance* function is defined as

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

with $\hat{\gamma}(-h) = \hat{\gamma}(h)$ for $h = 0, 1, \dots, n-1$

- ▶ The *sample autocorrelation* function is defined as

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

Time Series

Stationarity

- ▶ If we are justified in making some assumptions about the timeseries, we can compute statistics (e.g. mean and autocovariance) with only a single realization.
- ▶ We introduce the concept of a *strictly stationary timeseries*: with strict stationarity, the probabilistic behavior of the timeseries $\{x_1, x_2, \dots, x_n\}$ is identical to that of the timeshifted set $\{x_{1+h}, x_{2+h}, \dots, x_{n+h}\}$.
- ▶ This assumption is difficult to justify in practice
- ▶ We instead introduce the concept of a *weakly stationary timeseries*: with weak stationarity, the mean value μ_t is constant for all t ; and the autocovariance $\gamma(s, t) = \gamma(s + h, t + h)$ for any lag h .
- ▶ We use these requirements in practice, and when we refer to a *stationary timeseries*, we typically mean weak stationarity.

Time Series

Stationarity

- ▶ Thus, the sample mean of a stationary timeseries is just

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

- ▶ The variance of the mean is derived in equation (1.29) of Shumway.
- ▶ From this definition, we can easily show that if the samples are uncorrelated, the variance of the mean decreases proportionally to $1/n$ and the covariance decreases as $1/\sqrt{n}$.
- ▶ However, adjacent values may be correlated, even if they come from the same PDF; in this case, the variance of the mean must be calculated as function of lag, h , and the autocovariance $\gamma_x(h)$.
- ▶ In many practical cases a good estimate of $\gamma_x(h)$ is difficult to obtain.

Outline

Time Series Analysis

Introduction

Ensembles

Stationarity

Regression

Estimators

AR(p) Processes

MA(q) Processes

Time Series

Multiple Regression

- ▶ A principle goal in time series analysis is to explain one time series (the *output variable*) as a function of other time series (*input variables*).
- ▶ This process can be used to forecast future values of the output time series, or simply to deduce causality and thereby make decisions on how to manipulate a system.
- ▶ A simple but deceptively powerful technique of doing such data modeling is referred to as *multiple linear regression*.
- ▶ Linear in this case does not imply a linear relationship between inputs and outputs but rather a linear combination of arbitrarily complex (often non-linear) functions of the inputs.

Time Series

Multiple Regression, cont

Specifically, A multiple regression model is described by:

$$x_1 = \beta_1 z_{11} + \beta_2 z_{12} + \cdots + \beta_q z_{1q} + w_1$$

$$x_2 = \beta_1 z_{21} + \beta_2 z_{22} + \cdots + \beta_q z_{2q} + w_2$$

...

$$x_n = \beta_1 z_{n1} + \beta_2 z_{n2} + \cdots + \beta_q z_{nq} + w_n$$

where

- ▶ $x_t, t = 1 \dots n$ are the *response* variables;
- ▶ $z_{tj}, t = 1 \dots n, j = 1 \dots q$ are the *inputs*;
- ▶ $\beta_j, j = 1 \dots q$ are the regression coefficients;
- ▶ and $w_t, t = 1 \dots n$ are the *errors* (representing anything that departs from our model. We assume for now that the errors are iid with mean zero and variance σ_w^2

Multiple Regression

Matrix Form

- ▶ One way to consider the model: There are n observations, and for each observation, there are q observables. The n white noise terms represent our uncertainty about the value for each observation.
- ▶ Using matrix notation, the model can be simply expressed as:

$$\mathbf{x} = \mathbf{Z}\boldsymbol{\beta} + \mathbf{w}$$

where \mathbf{x} and \mathbf{w} are $n \times 1$ matrices (column vectors), $\boldsymbol{\beta}$ is an $q \times 1$ matrix (column vector), and \mathbf{Z} is an $n \times q$ matrix.

Multiple regression

Example

- ▶ Fit a linear time model to the global temperature time series given in Shumway

$$x_t = \beta_1 + \beta_2 t + w_t, \quad t = 1880, \dots, 2009$$

- ▶ In this case $q = 2$, $z_{t1} = 1$, and $z_{t2} = t$.
- ▶ The values of β are found by solving an optimization problem (e.g. least squares).
- ▶ Note that the β are themselves random variables since they are obtained from a sample process – i.e. we do not have the entire population statistics of x_t and the errors w_t are unknown.
- ▶ Thus what we seek are *estimators* $\hat{\beta}$ with desirable properties.
- ▶ We consider a less trivial example once we study how **$\hat{\beta}$** is obtained.

Outline

Time Series Analysis

Introduction

Ensembles

Stationarity

Regression

Estimators

AR(p) Processes

MA(q) Processes

Time Series

Multiple Regression, cont

- ▶ The model represents unknown population values for β .
- ▶ Since we do not know the white noise terms, \mathbf{w} , we cannot deterministically solve for β .
- ▶ Hence we will derive estimates, denoted as $\hat{\beta}$, by minimizing some criterion using sample statistics.
- ▶ A classic criterion is to find least-squares fit that minimizes S :

$$S = \left(\mathbf{x} - \mathbf{Z}\hat{\beta} \right)^T \left(\mathbf{x} - \mathbf{Z}\hat{\beta} \right)$$

- ▶ Note that a different sample from the population would yield a similar but slightly different $\hat{\beta}$. $\hat{\beta}$ is then itself a random variable.

Time Series

Basic Matrix Properties

- ▶ We use column vectors as lowercase \mathbf{x} and matrices as uppercase \mathbf{A} . The following properties are used in our derivations:
- ▶ $\frac{\partial}{\partial \mathbf{x}} (\mathbf{x}^T \mathbf{y}) = \mathbf{y}$
- ▶ $\frac{\partial}{\partial \mathbf{x}} (\mathbf{x}^T \mathbf{A} \mathbf{x}) = 2\mathbf{A}\mathbf{x}$ if \mathbf{A} is symmetric
- ▶ $(\mathbf{A}\mathbf{B})^T = \mathbf{B}^T \mathbf{A}^T$

Multiple regression

Statement of Normal Equations

- ▶ The least squared estimate $\hat{\beta}$ seeks to minimize

$$Q = \sum_{t=1}^n w_t^2 = \sum_{t=1}^n (x_t - \beta^T \mathbf{z}_t)^2$$

- ▶ Which in the original matrix notation yields the *normal equations*

$$\hat{\beta} = (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{x}$$

Multiple Regression

Derivation of Normal Equations

- ▶ We start with equation for square error:

$$\begin{aligned}Q &= \sum_{t=1}^n w_t^2 = \sum_{t=1}^n (x_t - \beta^T z_t)^2 \\&= (\mathbf{x} - \mathbf{Z}\hat{\beta})^T (\mathbf{x} - \mathbf{Z}\hat{\beta}) \\&= \mathbf{x}^T \mathbf{x} - \mathbf{x}^T \mathbf{Z}\hat{\beta} - \hat{\beta}^T \mathbf{Z}^T \mathbf{x} + \hat{\beta}^T \mathbf{Z}^T \mathbf{Z} \hat{\beta}\end{aligned}$$

- ▶ then we minimize with respect to $\hat{\beta}$

$$\frac{\partial Q}{\partial \hat{\beta}} = 0 = -\mathbf{Z}^T \mathbf{x} - \mathbf{Z}^T \mathbf{x} + 2\mathbf{Z}^T \mathbf{Z} \hat{\beta}$$

$$\mathbf{Z}^T \mathbf{Z} \hat{\beta} = \mathbf{Z}^T \mathbf{x}$$

$$\hat{\beta} = (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{x}$$

Multiple Regression

Normal Equations

- ▶ This result is referred to as the *normal equations*:

$$\hat{\beta} = (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{x}$$

- ▶ A few key points: It is extremely general, only linear in the coefficients $\hat{\beta}$. The inputs \mathbf{Z} can be arbitrary functions as in previous examples.
- ▶ Least squares is just one model for finding the coefficients β but is easy and has certain pleasing properties that are outlined on next slides.
- ▶ Note that the $\hat{\beta}$ themselves are random variables, estimators for β , since they are obtained with sample data.

- Consider a less trivial example, where M_t denotes cardiovascular mortality, T_t denotes temperature and P_t denotes the particulate levels.

$$M_t = \beta_1 + \beta_2 t + w_t$$

$$M_t = \beta_1 + \beta_2 t + \beta_3(T_t - \bar{T}) + w_t$$

$$M_t = \beta_1 + \beta_2 t + \beta_3(T_t - \bar{T})$$

$$+ \beta_4(T_t - \bar{T})^2 + w_t$$

$$M_t = \beta_1 + \beta_2 t + \beta_3(T_t - \bar{T})$$

$$+ \beta_4(T_t - \bar{T})^2 + \beta_5 P_t + w_t$$

- These represent a sequence of increasingly complex models for the mortality rate in terms of temperature and particulate levels. Note that respectively we have $z_{t,1:q} = (1, t), (1, t, T_t), (1, t, T_t, T_t^2), (1, t, T_t, T_t^2, P_t)$

$\hat{\beta}$ as Unbiased Estimator

- ▶ Recall the normal equations

$$\hat{\beta} = (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{x}$$

$$\mathbf{x} = \mathbf{Z}\beta + \mathbf{w}$$

- ▶ Substituting gives

$$\begin{aligned}\hat{\beta} &= (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T (\mathbf{Z}\beta + \mathbf{w}) \\ &= (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{Z}\beta + (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{w} \\ &= \beta + (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{w}\end{aligned}$$

- ▶ Taking the expectation of each side yields

$$E[\hat{\beta}] = E[\beta] + E[(\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{w}]$$

Of course $E[\beta]$ is just β since it is constant. The second term $E[(\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{w}]$ can be written as $(\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T E[\mathbf{w}] = \mathbf{0}$ if we assume *zero conditional mean of errors*.

Outline

Time Series Analysis

Introduction

Ensembles

Stationarity

Regression

Estimators

AR(p) Processes

MA(q) Processes

Time Series

AR Models

- ▶ We will study several models for describing timeseries with lagged autocorrelation:
 - AR: autoregressive model
 - MA: moving average model
 - ARMA: autoregressive moving average model
 - ARIMA: autoregressive integrated moving average model
- ▶ Classical regression describes a static process; that is, current values are a function of other values at the *same* timestep (i.e., $x_t = f(z_{1t}, z_{2t}, \dots)$).
- ▶ AR is one model that allows *past* values to influence the current values.
- ▶ An AR model of order P is called an AR(P) process and is described as:

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + w_t$$

Time Series

AR Models

- ▶ We can simplify this expression by defining the *backshift* operator B such that:

$$B^k x_t = x_{t-k}$$

- ▶ The backshift operator allows us to express this as:

$$x_t (1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) = w_t$$

- ▶ Furthermore, we can define the AR operator ϕ as:

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$$

- ▶ Finally, the backshift operator lets us express this as:

$$\boxed{\phi(B)x_t = w_t}$$

Time Series

Inverting an AR(1) Process

- ▶ The AR(1) process is:

$$\begin{aligned}x_t &= \phi x_{t-1} + w_t \\&= \phi^2 x_{t-2} + \phi w_{t-1} + w_t \\&\dots \\&= \phi^k x_{t-k} + \sum_{j=0}^{k-1} \phi^j w_{t-j}\end{aligned}$$

- ▶ If $|\phi| < 1$ then $\phi^k \rightarrow 0$ as $k \rightarrow \infty$ so:

$$x_t = \sum_j^{\infty} \phi^j w_{t-j}$$

Time Series

Statistics of AR(1) Process

- ▶ Assuming w_t is Gaussian, $E[x_t] = 0$.
- ▶ The autocovariance of the process is:

$$\begin{aligned}\gamma(h) &= \text{cov}(x_t, x_{t+h}) \\ &= E \left[\left(\sum_{j=0}^{\infty} \phi^j w_{t+h-j} \right) \left(\sum_{k=0}^{\infty} \phi^k w_{t-k} \right) \right] \\ &= E \left[\left(w_{t+h} + \cdots + \phi^h w_t + \phi^h w_t + \phi^{h+1} w_{t-1} + \cdots \right) \right. \\ &\quad \left. (w_t + \phi w_{t-1} + \cdots) \right]\end{aligned}$$

- ▶ Note that

$$E[w_i w_j] = \begin{cases} 0 & \text{if } i \neq j \\ \sigma_m^2 & \text{if } i = j \end{cases}$$

► Therefore

$$\begin{aligned}\gamma(h) &= \phi^h \sigma_w^2 + \phi^{h+1} \phi \sigma_w^2 + \phi^{h+1} \phi^2 \sigma_w^2 + \dots \\ &= \sigma_w^2 \sum_{j=0}^{\infty} \phi^{h+j} \phi^j \\ &= \sigma_w^2 \phi^h \sum_{j=1}^{\infty} \phi^{2j} \\ &= \boxed{\sigma_w^2 \phi^h \frac{1}{1 - \phi^2}}\end{aligned}$$

► And likewise, $\rho(h) = \phi^h$.

Outline

Time Series Analysis

Introduction

Ensembles

Stationarity

Regression

Estimators

AR(p) Processes

MA(q) Processes

Time Series

MA(q) Processes

- ▶ A moving average model of order q is called MA(q) and is defined as:

$$x_t = w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2} + \cdots + \theta_q w_{t-q}$$

- ▶ In the AR model, x_t was a combination of previous x terms; but in the MA model, x_t is a combination of past w terms.
- ▶ We define the moving average operator $\theta(B)$ as

$$\theta(B) = (1 + \theta B + \theta_2 B^2 + \cdots + \theta_q B^q)$$

- ▶ Then the equation becomes

$$x_t = \theta(B)w_t$$

Time Series

MA(1) Processes

- For an MA(1) process, it can be shown that:

$$x_t = w_t + \theta_1 w_{t-1}$$

$$E(x_t) = 0$$

$$\gamma(h) = \begin{cases} (1 + \theta^2)\sigma_w^2 & \text{if } h = 0 \\ \theta\sigma_w^2 & \text{if } h = 1 \\ 0 & \text{if } h > 1 \end{cases}$$

$$\rho(h) = \begin{cases} \frac{\theta}{1+\theta^2} & \text{if } h = 1 \\ 0 & \text{if } h > 1 \end{cases}$$