Time Series Analysis Classnotes

Dustin Leatherman

December 14, 2019

Contents

Time Series Analysis and Its Applications - 4th Edition

1 Chapter 1 - Characteristics of Time Series

1.1 Definitions

- **Filtered Series**: A linear combination of values in a time series.
- **Autoregression**: A time series where the current value x_t is dependent on a

function of previous values $x_{t-1}, x_{t-2}, \ldots$, etc. The order of Autoregression is dependent on the number of previous values.

- **Random Walk (with Drift)**: An AR(1) model with some constant δ called *drift*. When $\delta = 0$, this is called a Random Walk.
  - $x_t = \delta + x_{t-1} + w_t$
- **Signal-to-noise Ratio (SNR)**: $SNR = \frac{A}{\sigma_w}$
  - A: Amplitude of the Waveform
  - σ_w: Additive noise term
  - Note: A sinusoidal wave form can be written as $Acos(2\pi\omega t + \phi)$
- **Weak Stationarity**: A time series where the mean is constant. In this case, $h = |s - t|$ where h is

the separation between points x_s and x_t is important.

- **Note**: Many modeling practices attempt to reduce or transform a time series to white noise to then model it. This is known as *pre-whitening* and is typically done prior to performing Cross-Correlation Analysis (CCA).

1.2 Mean

1.2.1 Population

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

1. Moving Average $\mu_{vt} = E(v_t) = \frac{1}{3}[E(w_{t-1}) + E(w_t) + E(w_{t+1})] = 0$
2. Random Walk with Drift $\mu_{xt} = E(x_t) = \delta t + \sum_{j=1}^{t} E(w_j) = \delta t$

1.2.2 Sample

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

1.3 Autocovariance

- the second moment product for all s and t.
- Measures linear dependence between two points on the same series observed at different times.

Population: $\gamma_x(s,t) = cov(x_s, x_t) = E[(x_s - \mu_s)(x_t - \mu_t)]$
Sample: $\hat{\gamma}(h) = n^{-1}\sum_{t=1}^{n-h}(x_{t+h} - \bar{x})(x_t - \bar{x})$ where $\hat{\gamma}(-h) = \hat{\gamma}(h) \forall h \in [0, n-1]$

- This estimator guarantees a non-negative result.

1.3.1 Covariance of Linear Combos

Let U and V be linear combinations with finite variance of the randome variables X_j and Y_k.

$$
\begin{aligned}
U =& \sum_{j=1}^{m} a_j X_j \\
V =& \sum_{k=1}^{r} b_k Y - k
\end{aligned}
\tag{2}
$$

Then,

- $cov(U,V) = \sum_{j=1}^{m}\sum_{k=1}^{r} a_j b_k cov(X_j, Y_k)$
- $cov(U,U) = var(U)$

1.3.2 Moving Average

$\gamma_v(s,t) = cov(v_s, v_t) = cov(\frac{1}{3}(w_{s-1} + w_s + w_{s+1}), \frac{1}{3}(w_{t-1} + w_t + w_{t+1}))$

$$\gamma_v(s,t) = \begin{cases} \frac{3}{9}\sigma_w^2 & s = t \\ \frac{2}{9}\sigma_w^2 & |s-t| = 1 \\ \frac{1}{9}\sigma_w^2 & |s-t| = 2 \\ 0 & |s-t| > 2 \end{cases} \tag{3}$$

1.3.3 Random Walk

$\gamma_x(s,t) = cov(x_s, x_t) = cov(\sum_{j=1}^{s} w_j, \sum_{k=1}^{t} w_k) = min(s,t)\sigma_w^2$

- covariance of walk is dependent on time opposed to lag, unlike Linear combos and Moving Average.

1.3.4 Cross-covariance

Covariance between two time series x and y
Population: $\gamma_{xy}(s,t) = cov(x_s, y_t) = E[(x_s - \mu_{xs})(y_t - \mu_{yt})]$
Sample: $\hat{\gamma_{xy}}(h) = n^{-1}\sum_{t=1}^{n-h}(x_{t+h} - \bar{x})(y_t - \bar{y})$

1.4 Autocorrelation (ACF)

Measures the linear predictability of a time eseries at time t (x_t) using only the value x_s.
Population: $\rho(s,t) = \frac{\gamma(s,t)}{\sqrt{\gamma(s,s)\gamma(t,t)}}$
Sample: $\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}$

- For large sample sizes, the sample ACF is $\sim N(0, \frac{1}{n})$

1.4.1 Cross-correlation

Correlation between two different time series x and y

Population: $\rho_{xy}(s,t) = \frac{\gamma_{xy}(s,t)}{\sqrt{\gamma_x(s,s)\gamma_y(t,t)}}$

Sample: $\hat{\rho_{xy}}(h) = \frac{\hat{\gamma_{xy}}(h)}{\sqrt{\hat{\gamma_x}(0)\hat{\gamma_y}(0)}}$

- For large samples, $\hat{\rho_{xy}} \sim N(0, \frac{1}{n})$

1.5 Stationary Time Series

A measure of regularity over the course of a time series.

1.5.1 Strict Stationary

A time series for which the probabilistic behavior of every collection of values $(x_{t1}, x_{t2}, ..., x_{tk})$ is identical to that of the time shifted set $(x_{t1+h}, ..., x_{tk+h})$.
i.e. $Pr(x_{t1} \leq c1, ..., x_{tk} \leq c_k) = Pr(x_{t1+h} \leq c1, ..., x_{tk+h} \leq c_k)$

Mean: $\mu_t = \mu_s$ for all s and t indicating that μ_t is *constant*.

Autocovariance: $\gamma(s,t) = \gamma(s+h, t+h)$

- The process depends only on time *difference* between s and t rather than the actual times.

This definition is too restrictive and unrealistic for most applcations.

1.5.2 Weakly Stationary

A time series for which

1. μ_t is constant and does not depend on time t
2. $\gamma(s,t)$ depends on s and t only through their difference $|s-t|$

If a time series is normal, then it implies it is strict stationary.

1. Autocorrelation Function (ACF) $\rho(h) = \frac{\gamma(t+h,t)}{\sqrt{\gamma(t+h,t+h)\gamma(t,t)}} = \frac{\gamma(h)}{\gamma(0)}$
   - Moving Averages **are** Stationary
   - Random Walks are **not** Stationary since the mean depends on time

1.5.3 Trend Stationarity

When the Mean function is dependent on time but the Autocovariance function is not, the model can be considered as having a stationary behavior around a linear trend. a.k.a trend stationarity.

1.5.4 Autocovariance Function Properties

1. $\gamma(h)$ is non-negative definite meaning that that variance and linear combinations of such will never be negative.

   $0 \leq var(a_1 x_1 + ... + 1_n x_n) = \sum_{j=1}^{n} \sum_{k=1}^{n} a_j a_k \gamma(j-k)$

2. $\gamma(h=0) = E[(x_t - \mu)^2]$ is the variance of the time series and thus Cauchy-Swarz inequality implies $|\gamma(h)| \leq \gamma(0)$

3. $\gamma(h) = \gamma(-h)$ for all h. i.e. symmetrical

1.5.5 Joint Stationarity

Both time series are stationary and the Cross-Covariance Function is a function only of lag h.

$\gamma_{xy}(h) = cov(x_{t+h}, y_t) = E[(x_{t+h} - \mu_x)(y_t - \mu_y)]$

Cross-correlation Function (CCF) of a jointly stationary time series x_t and y_t is defined as $\rho_{xy}(h) = \frac{\gamma_{xy}(h)}{\sqrt{\gamma_x(0)\gamma_y(0)}}$

Generally $cov(x_2, y_1) \neq cov(x_1, y_2)$ and $\rho_{xy}(h) \neq \rho_{xy}(-h)$; however, $\rho_{xy}(h) = \rho_{yx}(-h)$.

1.5.6 Linear Process

Linear combination of white noise variates w_t, given by $x_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j w_{t-j}, \sum_{j=-\infty}^{\infty} |\psi_j| < \infty$

1. Autocovariance for $h \geq 0$ $\gamma_x(h) = \sigma_w^2 \sum_{j=-\infty}^{\infty} \psi_{j+h} \psi_j$

   models that do not depend on the future are considered **causal**. In causal linear processes, $\psi_j = 0$ for $j < 0$

1.5.7 Gaussian (Normal) Process

A process is said to be Gaussian if the n-dimensional vectors $x = (x_{t1}, x_{t2}, ..., x_{tn})^T$ for every collection of distinct time points $t_1, t_2, ..., t_n$ and every positive integer n have a multivariate normal distribution.

- A Gaussian Process is Strictly Stationary. Gaussian Time series form the basis of modeling many time series.

- **Wold Decomposition**: A stationary non-deterministic time series is a causal linear process with $\Sigma\psi_j^2 < \infty$

1.6 Vector Time Series

$\underset{(p\times 1)}{x_t} = (x_{t1}, ..., x_{tp})^T$

1.6.1 Mean

1. Population $\vec{\mu} = E(x_t)$

2. Sample Vector $\bar{x} = n^{-1}\sum_{t=1}^{n} x_t$

1.6.2 Autocovariance Matrix

1. Population $\Gamma(h) = E[(x_{t+h} - \mu)(x_t - \mu)^T]$

   - $\Gamma(-h) = \Gamma^T(h)$ holds

2. Sample $\underset{(p\times p)}{\hat{\Gamma}(h)} = n^{-1}\sum_{t=1}^{n-h}(x_{t+h} - \bar{x})(x_t - \bar{x})^T$

   - $\hat{\Gamma}(-h) = \hat{\Gamma}^T(h)$ holds

1.7 Multidimensional Series

In cases where a series is indexed by more than time alone, a *multidimensional process* can be used. For example, a coordinate may be defined as (s_1, s_2). Thus, $\underset{(r\times 1)}{s} = (s_1, ..., s_r)^T$ where s_i is the coordinate of the ith index.

1.7.1 Mean

- **Population**: $\mu = E(x_s)$

- **Sample**: $\bar{x} = (S_1 S_2 ... S_r)^{-1}\Sigma_{s1}\Sigma_{s2}...\Sigma_{sr}x_{s1,s2,...,sr}$

1.7.2 Autocovariance

- **Population**: $\gamma(h) = E[(x_{s+h} - \mu)(s_x - \mu)]$ with multidimensional lag vector h, $h = (h_1, ..., h_r)^T$
- **Sample**: $\hat{\gamma}(h) = (S_1 S_2 ... S_r)^{-1} \Sigma_{s1} \Sigma_{s2} ... \Sigma_{sr} (x_{s+h} - \bar{x})(x_s - \bar{x})$

1.7.3 Autocorrelation

- **Sample**: $\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}$ with

$\hat{\gamma}$ defined above

1.7.4 Variogram

Sampling requirements for multidimensional processes are severe since there must be some uniformity across values. When observations are irregular in time space, modifications to the estimators must be made. One such modificaiton is the variogram.

$2V_x(h) = var(x_{s+h} - x_s)$

- **Sample Estimator**: $2\hat{V}_x(h) = \frac{1}{N(h)} \Sigma_s (s_{x+h} - x_s)^2$
  - $N(h)$: Number of points located within h

Issues

- negative estimators for the covariance function occur
- Indexing issues?

2 Chapter 2 - Time Series Regression and Exploratory Data Analysis

2.1 Exploratory Data Analysis

It is necessary for time series data to be stationary so lags are possible. It is tough to measure time series if the dependence structure is not regular. At bare minimum, the autocovariance and mean functions must be stationary for some period of time.

2.1.1 Trend Stationary Models

$x_t = \mu_t + y_t$

- x_t: Observations
- μ_t: Trend
- y_t: Stationary Process

Strong trends often obscure behavior of the stationary process so detrending is a good first step.

$$\begin{aligned} \hat{y_t} =& x_t - \hat{\mu_t} \\ =& x_t - (\beta_0 + \beta_1 t) \end{aligned} \tag{4}$$

Using $\hat{\mu_t} = \beta_0 + \beta_1 t$ detrends the data.

2.1.2 Differencing

$x_t - x_{t-1} = (\mu_t + y_t) - (\mu_{t-1} + y_{t-1}) = \beta_1 + y_t - y_{t-1}$
First Difference Notation: $\nabla x_t = x_t - x_{t-1}$

1. Backshift Used to specify a specific difference from a given point in a time series. When $k < 0$, it becomes a forward-shift operator.

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

   A given difference can be represented as: $\nabla^d x_t = (1 - B)^d x_t$

   (a) Example - Second Difference
   $\nabla^2 x_t = (1 - B)^2 x_t = (1 - 2B + B^2) x_t = x_t - 2x_{t-1} + x_{t-2}$

   (b) Example - Fractional Differencing
   $-0.5 < d < 0.5$
   $\nabla^{0.5} x_t = (1 - B)^{0.5} x_t$
   Typically used for environmental time series in hydrology.

2. Pros

   - No parameters estimated in differencing operation
   - Not viable when goal is to coerce data to stationarity

3. Cons

- does not yield an estimate of the stationary process y_t
- Detrending more viable if trend is fixed

4. Transformations Just as transformations can fix non-normality, so can they fix non-stationarity. The Box-Cox family transformations are useful.

$$y_t = \begin{cases} (x_t^\lambda - 1)/\lambda & \lambda \neq 0 \\ logX_t & \lambda = 0 \end{cases} \tag{5}$$

2.1.3 Trig Identities to Discover a Signal in Noise

$$\begin{aligned} x_t =& Acos(2\pi\omega t + \phi) + w_t \\ acos(2\pi\omega t + \phi) =& \beta_1 cos(2\pi\omega t) + \beta_2 sin(2\pi\omega t) \\ \beta_1 =& acos(\phi) \\ \beta_2 =& -asin(\phi) \\ \omega = 1/50 & \\ x_t =& \beta_1 cos(2\pi t/50) + \beta_2 sin(2\pi t/50) + w_t \end{aligned} \tag{6}$$

2.1.4 Smoothing

Let a Moving Average be defined as
$m_t = \sum_{j=-k}^{k} a_j x_{t-j}$
where
$a_j = a_{-j} \geq 0, \sum_{j=-k}^{k} a_j = 1$

1. Kernal smoothing

   Moving Average smoother that uses a weight function (kernel) to average observations.

$$\begin{aligned} m_t =& \sum_{i=1}^{n} w_i(t) x_i \\ w_i(t) =& K(\frac{t-i}{b}) / \sum_{j=1}^{n} K(\frac{t-j}{b}) \end{aligned} \tag{7}$$

   where $K(.)$ is a kernel function.

(a) Example - Original Kernel Function
$K(z) = \frac{1}{\sqrt{(2\pi)}} exp(-z^2/2)$

2. Lowess

KNN Regression followed by robust weighted regression to obtain smoothed values.

3. Splines

Given the following:

$$\begin{aligned} x_t =& m_t + w_t \\ m_t =& \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 \\ t =& 1, ..., n \end{aligned} \tag{8}$$

Let t be divied into k intervals called *knots*. In each interval, fit a polynomial regression model. The most common is a *cubic spline* where the Order is 3 (as m_t is defined).

(a) Smoothing Spline
The following is a compromise between the model fit (smoothness) and the data (no smoothness).
$\sum_{t=1}^{n}[x_t - m_t]^2 + \lambda \int (m_t'')^2 dt$
$\lambda > 0$ controls the degree of smoothness.