10/20/16

Motivation:

- Analyze, model and **predict** data that is observed in a sequential order
- Data is no longer independent, and so standard inferential procedures don't work anymore/are invalid
- Decompose dependent data into independent components
- We care less about finding relationships between a response variable and covariates. We typically want to forecast a response using just its past values.

Regression_Example.R and *ConsIndex.txt*

Definitions

An **observed time series** $\{x_t : \ t \in T\}$ is a collection of observations of a variable of interest over time.

A **time series** is a stochastic process indexed by time. Specifically, we have a sequence of random variables $\{X_t : \ t \in T\}$, where T is an index of time points.

- if T is a discrete set, i.e. $T = \{1, 2, 3, \ldots\}$, then $\{X_t\}$ is a **discrete time series**.
- if T is a continuous interval, i.e. $T = \{t > 0\}$, then $\{X_t\}$ is a **continuous time series**.

A **time series model** is the specification of the joint distribution of the random variables $\{X_t : t \in N\}$: $P(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_n \leq x_n)$ for $-\infty < x_1, x_2, \ldots, x_n < \infty$ and $n \in N$. But, in general, we can't hope to estimate all of the parameters in such a model with the data we've observed.

But, most of the information about a distribution is contained in the first two moments:

- First Moments: $E[X_t]$, $t = 1, 2, \ldots$ -> *means*
- Second Moments: $E[X_t X_{t+h}]$, $t = 1, 2, \ldots$ and $h = 0, 1, 2, \ldots$ -> *variances/covariances*

Main take-away: we don't need the whole joint distribution. Our modeling will be based on **second-order properties**.

$\{x_t\}$ —observed from—> $\{X_t\}$

Zero Mean Models

IID Noise

If $\{X_1, X_2, \ldots, X_k\}$ are iid random variables with $E[X_t] = 0, \ t = 1, 2, \ldots, k$, then $P(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_n \leq x_n) = independence = \prod_{t=1}^{k} P(X_t \leq x_t) = identically\ distributed = \prod_{t=1}^{k} F(x_t)$. In this special case, the joint distribution is defined by one marginal distribution with zero mean.

White noise

A white noise process is a sequence of **uncorrelated** (not necessarily independent!) random variables $\{X_t\}$ each with mean 0, and finite variance σ^2.

We denote this by $\{X_t\} \sim WN(0, \sigma^2)$.

- $E[X_t] = 0$
- $Var(X_t) = \sigma^2$ finite

- $Cov(X_i, X_j) = 0$ for $i \neq j$

(note: IID noise is a subset of White noise)

Classical Time Series Decomposition

$X_t = m_t + s_t + \epsilon_t$

- m_t: trend term (average change in X_t over time)
- s_t: seasonal term (regular periodic fluctuations)
- ϵ_t: error (unexplained variation in X_t's)

Lecture 1.pptx

Example

Consider average seasonal temperature over many years where we wish to fit a model of the form $X_t = m_t + s_t + \epsilon_t$.

Here, we assume m_t is a polynomial in t, and s_t can be represented with indicator/dummy variables:

- $W_1 = \begin{cases} 1 & \text{if spring} \\ 0 & \text{otherwise} \end{cases}$
- $W_2 = \begin{cases} 1 & \text{if fall} \\ 0 & \text{otherwise} \end{cases}$
- $W_3 = \begin{cases} 1 & \text{if winter} \\ 0 & \text{otherwise} \end{cases}$

$X_t = \sum_{i=0}^{p} \beta_i t^i + \sum_{j=1}^{3} \alpha_j W_j + \epsilon_t$, $\epsilon_t \sim N(0, \sigma^2)$ (iid)

- We typically estimate α's and β's using OLD, which implies that we are making OLS assumptions (which still may not be valid).
- If the assumptions are invalid, then we use the **Box-Jenkins** class of models (i.e. AR, MA, ARMA, SARIMA)

AirPassengers Analysis.R

10/25/16

Recap

- Time series –> $\{X_t : t \in N\}$ <– a time series model puts constraints on the first and second moments of these random variables.
- Observed time series –> $\{x_t : t \in N\}$.

Stationarity

Strict stationarity

A time series $\{X_t\}$ is said to be **strictly stationary** if the joint distribution of $X_{t_1}, X_{t_2}, \ldots, X_{t_n}$ is the same as that of $X_{t_1+h}, X_{t_2+h}, \ldots, X_{t_n+h}$ for all $n, h, t_1, t_2, \ldots, t_n \in N$.

i.e., a strictly stationary time series preserves **all** statistical properties over time shift.

Problems:

- We often can't specify the joint distribution of these random variables and so this assumption is usually impossible to check.
- Also, this assumption tends to be too strict and is not often met.

This motivates the need for a weaker version of stationarity.

But first…

Let $\{X_t\}$ be a time series.

- The **mean function** of $\{X_t\}$ is $\mu_X(t) = E(X_t)$,
- The **covariance function** of $\{X_t\}$ is $\gamma_X(r, s) = Cov(X_r, X_s) = E(X_r X_s) - \mu_X(r)\mu_X(s)$.

Weak stationarity

A time series $\{X_t\}$ is **weakly stationary** if $E(X_t^2) < \infty$ and:

- $\{X_t\}$ is $\mu_X(t) = E(X_t)$ is independent of t,
- $\gamma_X(t, t+h) = Cov(X_t, X_{t+h})$ is independent of t for all h.
  - covariance depends on h but not t

Remarks:

- Strict stationarity ==> weak stationarity
- From now on, "stationarity" means *weak* stationarity
- For a stationary time series $\{X_t\}$:
  - $E(X_t) = \mu_X$
  - $Cov(X_t, X_{t+h}) = \gamma_X(t, t+h) = \gamma_X(0, h) = \gamma_X(h)$

Definitions

Let $\{X_t\}$ be a stationary time series.

- The **autocovariance function** (ACVF) of $\{X_t\}$ at **lag** h is $\gamma_X(h)$.
- The **autocorrelation function** (ACF) of $\{X_t\}$ at **lag** h is $\rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)} = Corr(X_t, X_{t+h})$.

- $\gamma_X(h) = \gamma_X(-h)$.

(Reminder: $Corr(X,Y) = \frac{Cov(X,Y)}{SD(X)SD(Y)} = \frac{Cov(X_t,X_{t+h})}{\sqrt{Var(X_t)Var(X_{t+h})}} = \frac{Cov(X_t,X_{t+h})}{\sqrt{Cov(X_t,X_t)Cov(X_{t+h},X_{t+h})}} = \frac{Cov(X_t,X_{t+h})}{\sqrt{\gamma_X(0)\gamma_X(0)}} = \frac{\gamma_X(h)}{\gamma_X(0)}$)

Examples

First Order Autoregression: AR(1)

Assume $\{X_t\}$ is a stationary time series satisfying the equations

$$X_t = \Phi X_{t-1} + Z_t$$

for $t \in Z$, $|\Phi| < 1$ and $Z_t \sim WN(0, \sigma^2)$. Also assume Z_t and X_s are uncorrelated for all $s < t$. Calculate the ACVF and ACF of $\{X_t\}$.

- $E(X_t) = \Phi E(X_{t-1}) + E(Z_t) \to E(X_t) = \Phi E(X_{t-1}) \to E(X_t) = 0$ since $\{X_t\}$ is stationary.
- $\gamma_X(h) = Cov(X_t, X_{t-h}) = E(X_t X_{t-h}) = E(\Phi X_{t-1}X_{t-h} + Z_t X_{t-h}) = \Phi E(X_{t-1}X_{t-h}) + E(Z_t X_{t-h}) = \Phi E(X_{t-1}X_{t-h}) = \Phi\gamma_X(h-1) = \Phi^h \gamma_X(0)$ (assume $h > 0$).

By stationarity, $\gamma_X(h) = \gamma_X(-h)$ so $\gamma_X(h) = \Phi^{|h|}\gamma_X(0)$.

- $\gamma_X(0) = Cov(X_t, X_t) = E(X_t^2) = E(\Phi^2 X_{t-1}^2 + 2\Phi X_{t-1}Z_t + Z_t^2) = \Phi^2 E(X_{t-1}^2) + 2\Phi E(X_{t-1}Z_t) + E(Z_t^2) = \Phi^2\gamma_X(0) + \sigma^2 \Rightarrow \gamma_X(0) = \frac{\sigma^2}{1-\Phi^2}$

$\therefore \gamma_X(h) = \frac{\Phi^{|h|}\sigma^2}{1-\Phi^2}$ for $h \in Z$

$\therefore \rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)} = \Phi^{|h|}$ for $h \in Z$

ACF signature for AR(1) is exponential decay.

First Order Moving Average: MA(1)

Consider process $X_t = Z_t + \theta Z_{t-1}$ where $t \in N$ and $\{Z_t\} \sim WN(0, \sigma^2)$ and $\theta \in R$. Show $\{X_t\}$ is stationary and derive its ACF.

- $\mu_X = E(X_t) = E(Z_t) + \theta E(Z_{t-1}) = 0$ for all t.
- $\gamma_X(h) = Cov(X_t, X_{t+h}) = Cov(Z_t + \theta Z_{t-1}, Z_{t+h} + \theta Z_{t+h-1}) = Cov(Z_t, Z_{t+h}) + \theta Cov(Z_t, Z_{t+h-1}) + \theta Cov(Z_{t-1}, Z_{t+h}) + \theta^2 Cov(Z_{t-1}, Z_{t+h-1})$

$$\gamma_X(h) = \begin{cases} \sigma^2(1+\theta^2) & \text{if } h = 0 \\ \theta\sigma^2 & \text{if } h = \pm 1 \\ o & \text{oterwise} \end{cases} \qquad \leftarrow \text{independent of t.}$$

(Reminder: $Cov(X+Y, W+Z) = Cov(X,W) + Cov(X,Z) + Cov(Y,W) + Cov(Y,Z)$)

$\therefore \{X_t\}$ is stationary.

- $\gamma_X(0) = \sigma^2(1+\theta^2)$

and $\rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)} = \{1$ if $h = 0$, $\frac{\theta}{1+\theta^2}$ if $h \pm 1$, 0 otherwise.

ACF signature of MA(1) is a spike for $h = 0, 1$ and then nothing for $h > 1$.

We've seen that the ACF can provide information regarding which model may be appropriate for an observed time series. To do this in practice, we need a sample estimate of the ACF.

Definitions

Let $x_1, x_2, \ldots, x_n$ be our observed time series.

- the **sample mean** is $\widehat{\mu_x} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$,
- the **sample autocovariance** is $\widehat{\gamma_x}(h) = \frac{1}{n}\sum_{t=1}^{n-h}(x_{t+h} - \bar{x})(x_t - \bar{x})$,
- the **sample autocorrelation** is $\widehat{\rho_x}(h) = \frac{\widehat{\gamma_x}(h)}{\widehat{\gamma_x}(0)}$

Note:

- θ, a Greek letter, denotes a paramters (unknown number),
- $\widehat{\theta}$, is a sample estimate of θ (known number),
- $\widetilde{\theta}$, is an estimator, a random variable.

The sample ACF can be used to investigate the "uncorrelatedness" in a time series. For example, we might use this to evaluate the uncorrelated assumption in residuals.

(Reminder: independence $\Rightarrow$ uncorrelated; uncorrelated $\not\Rightarrow$ independence)

For stationary time series, $\widetilde{\rho}(h) \dot{\sim} N(0, \frac{1}{n})$ (n = number of data points).

Consequently, an approximate 95% confidence interval for $\rho_x(h)$ is $\pm\frac{1.96}{\sqrt{n}}$.

If $\widetilde{\rho}(h)$ falls outside these limits, for any h, we judge this to be significant.

SACF Examples.R

10/27/16

Recap

- Autocovariance function (ACVF): $\gamma_X(h) = Cov(X_t, X_{t-h})$ for all $h \in Z$
- Autocorrelation function (ACF): $\rho_X(h) = Corr(X_t, X_{t-h}) = \frac{\gamma_X(h)}{\gamma_X(0)}$
  - Properties of ACVF:
    * $\gamma_X(0) = Var(X_t)$
    * $\gamma_X(-h) = \gamma_X(h)$
    * $|\rho_X(h)| \leq 1$

Why is stationarity important?

In order to build a model that forecasts with any accuracy, we require an assumptions that something doesn't vary with time. After accounting for deterministic trend and/or seasonality, we hope that the remaining randomness can be described as stationary.

In the Box-Jenkins class of models, we can use AR (autoregressive), MA (moving average), and ARMA models to model stationary time series.

First, notation:

Backshift operator: B, where $BX_t = X_{t-1}$ i.e. $B^2 X_t = X_{t-2}$.

Generally, $B^n X_t = X_{t-n}$ and $B^0 = I$

MA(q) Process

A process/time series $\{X_t\}$ is called a **moving average process of order q** if

$$X_t = \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \ldots + \theta_q \epsilon_{t-q}$$

where $\{\epsilon_t\} \sim WN(0, \sigma^2)$ and $\theta_1, \theta_2, \ldots, \theta_q$ are constants.

Remarks:

- MA(q) processes are stationary (*exercise*: prove this!)
- An MA(q) process is **q-correlated** (i.e., $\rho_X(h) = Corr(X_t, X_{t-h}) = 0$ for $h > q$ and not necessarily 0 for $h \leq q$)

Thus, **the ACF signature of an MA(q) process is non-zero spiked for $h = 0, 1, 2, \ldots, q$ and then no spikes for ever after**.

- An MA(q) process can be denoted as: $X_t = \epsilon_t + \theta_1 B^1 \epsilon_t + \ldots + \theta_q B^q \epsilon_t = (1 + \sum_{s=1}^{q} \theta_s B^s)\epsilon_t = \theta^q(B)\epsilon_t$

where $\theta^q(z) = 1 + \sum_{s=1}^{q} \theta_s z^s$ is the **generating function**.

- An MA(q) is **invertible** if the complex roots of $\theta^q(z)$ lie outside the unit circle. i.e. For all z such that $\theta^q(z) = 0$, then $|z| > 1$.

Example

$X_t = \epsilon_t + 0.2\epsilon_{t-1} + 0.7\epsilon_{t-2}$

$\theta(z) = 1 + 0.2z + 0.7z$

The roots of $\theta(z)$ are $z = \frac{-0.2 \pm \sqrt{0.2^2 - 4(0.7)(1)}}{2(0.7)} = \frac{-0.2 \pm \sqrt{2.76}i}{1.4} \Rightarrow z = -0.14 \pm 1.19i$

$|z| = \sqrt{(-0.14)^2 + (1.19)^2} = 1.198 > 1$

So $\{X_t\}$ is invertible.

Reminders:

- The zeros of a quadratic of the form $ax^2 + bx + c$ are $x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$
- $c = a + ib \Rightarrow |c| = \sqrt{a^2 + b^2}$

AR(p) Process

The process $\{X_t\}$ is called an **autoregressive process of order p** if

$$X_t = \Phi_1 X_{t-1} + \Phi_2 X_{t-2} + \ldots + \Phi_p X_{t-p} + \epsilon_t$$

where $\epsilon_t \sim WN(0, \sigma^2)$ and $\Phi_1, \Phi_2, \ldots, \Phi_p$ are constants.

- An AR(p) process can be denoted as:

$X_t - \Phi_1 X_{t-1} - \Phi X_{t-2} - \ldots - \Phi_p X_{t-p} = \epsilon_t$

$\Leftrightarrow X_t - \Phi_1 B^1 X_t - \Phi_2 B^2 X_t - \ldots - \Phi_p B^p X_t = \epsilon_t$

$\Leftrightarrow (1 - \sum_{r=1}^{p} \Phi_r B^r) X_t = \epsilon_t$

$\Leftrightarrow \Phi^p(B) X_t = \epsilon_t$

where $\Phi^p(z) = 1 - \sum_{r=1}^{p} \Phi_r z^r$ is the **generating function**.

- An AR(p) process is **stationary** is the complex roots of $\Phi^p(z)$ lie outside the unit circle. i.e. For all z such that $\Phi(z) = 0$, we require $|z| > 1$.

Example

$X_t = \Phi X_{t-1} + \epsilon_t \Rightarrow (1 - \Phi B) X_t = \epsilon_t$

$\Phi(z) = 1 - \Phi z \Rightarrow \Phi(z) = 0$ if $z = \frac{1}{\Phi}$

For stationarity, we need $|z| > 1 \Rightarrow |\frac{1}{\Phi}| > 1 \Rightarrow |\Phi| > 1$.

Partial Autocorrelation Dunction(PACF)

For a stationary process, the ACF of lag h measures the correlation between X_t and X_{t+h}. This correlation could be dure to a direct connection between X_t and X_{t+h}, but it may also be influenced by observations at intermediate lags: $X_{t+1}, X_{t+2}, \ldots, X_{t+h-1}$.

The PACF of lag h measures the correlation between X_t and X_{t+h} once the influence of the intermediate lags has been removed/accounted/controled for.

We remove this effect using **linear predictors**:

$$\widehat{X_t} = Pred(X_t | X_{t_1}, X_{t+2}, \ldots, X_{t+h-1})$$

$$\widehat{X_{t+h}} = Pred(X_{t+h} | X_{t_1}, X_{t+2}, \ldots, X_{t+h-1})$$

where this prediction is commonly based on a linear regression.

Thus, for a stationary time series {X_t}, the **partial autocorrelation function of lag h** is: $\alpha_X(h) = \begin{cases} Corr(X_t, X_t) = 1, & \text{if } h = 0 \\ Corr(X_t, X_{t+1}) = \rho_X(1), & \text{if } h = 1 \\ Corr(X_t, X_{t+h}) = Corr(X_t - \widehat{X_t}, X_{t+h} - \widehat{X_{t+h}}), & \text{if } h > 1 \end{cases}$

(assume without loss of generality that $h \geq 0$)

Example

Derive the PACF of an AR(1) process $X_t = \Phi X_{t-1} + \epsilon_t$.

$\alpha_X(h) = \begin{cases} 1, & \text{if } h = 0 \\ \rho(1) = \Phi & \text{if } h = 1 \end{cases}$

If $h = 2$:

-$\alpha(2) = Corr[X_t = \widehat{X_t}, X_{t+2} - \widehat{X_{t+2}}] = Corr[X_t - f(X_{t+1}), X_{t+2} - \Phi X_{t+1}] = Corr[X_t - f(X_{t+1}), \epsilon_{t+2}] = Corr[X_t, \epsilon_{t+2}] - Corr[f(X_{t+1}), \epsilon_{t+2}] = 0 - 0 = 0$

We can see that $\alpha(h) = 0$ for any $h \geq 2$.

So **PACF for an AR(1) has non-zero spikes for $h = 0, 1$ and is zero for all $h \geq 2$.**

Remarks:

- If $\{X_t\} \sim AR(p)$, then the PACF satisfies $\alpha(h) = 0$ for all $h > p$ and $\alpha(h) \neq 0$ necessarily for $h \leq p$.
- Whereas an ACF can be used to determine the order of an MA process, a PACF can be used to determine the order of an AR process.

ARMA(p,q) Process

$\{X_t\}$ is an **autoregressive moving average process of orders p and q** if

$$X_t - \Phi_1 X_{t-1} - \Phi_2 X_{t-2} - \ldots - \Phi_p X_{t-p} = \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \ldots + \theta_q \epsilon_{t-q}$$

$$\Phi^p(B) X_t = \theta^q(B) \epsilon_t$$

where $\{\epsilon_t\} \sim WN(0, \sigma^2)$ and $\Phi^p(z)$ and $\theta^q(z)$ are the AR and MA generating functions, and **we require them to have distinct roots**.

Remark:

- ARMA(p, 0) = AR(p)
- ARMA(0, q) = MA(q)

Example: ARMA(1,2)

$\Phi^1(B)X_t = \theta^2(B)\epsilon_t \Rightarrow (1-\Phi B)X_t = (1+\theta_1 B+\theta_2 B^2)\epsilon_t \Rightarrow X_t - \Phi X_{t-1} = \epsilon_t + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2}$

We require $\begin{cases} \Phi^1(z) = 1 - \Phi Z \\ \theta^2(z) = 1 + \theta_1 z + \theta_2 z^2 \end{cases}$

	ACF	**PACF**
MA(q)	Spike for $h \leq q$ and negligibly small spikes for $h > q$	Exponential decay
AR(p)	Exponential decay	Spiked for $h \leq p$ and "nothing" for $h > p$
ARMA(p,q)	q spikes then decay	p spikes then decay