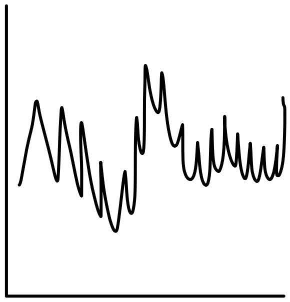
MAT 3379

Time Series Analysis

Study Guide

Winter 2024



$$X_t = Y_t + m_t + S_t$$

Examples

- White noise: $\{Z_t\}$ sequence of independent random variables with mean 0 and variance 1
- Random Walk: $\{Z_t\}$ sequence of iid random variables with mean 0 and variance σ_Z^2. $X_t = \sum_{i=1}^{t} Z_t$, $t=1,2,\dots$
- Model with trend: Sometimes a trend is present in time series. $X_t = 1+2t+Z_t$, $t=1,2,\dots$
  Trend is $m_t = 1+2t$
- Economics Trend: $X_t = P_t e^{rt}$
  P_t is real price, r is interest rate

Time Series Structure

$$X_t = m_t + Y_t + S_t$$

- m_t is a trend
- S_t is a seasonal part
- Y_t is a stationary part

Eliminate Trend

- Differencing: Compute $\nabla X_t = X_t - X_{t-1}$, $t=2,\dots,n$
- Polynomial Fitting: Assume $m_t = a+bt$
  Estimate a and b by minimizing
  $\sum_{t=1}^{n}(X_t - a - bt)^2$, $\hat{m}_t = \hat{a} + \hat{b}t$.
  Detrended time series: $\hat{Y} = X_t - \hat{m}_t$
- Exponential Smoothing: $\alpha \in (0,1)$
  Trend: $\hat{m}_1 = X_1$, $\hat{m}_t = \alpha X_t + (1-\alpha)\hat{m}_{t-1}$, $t=2,\dots,n$
  De-trended time series: $\hat{Y}_t = X_t - \hat{m}_t$
- Moving Average Smoothing: $q \in \mathbb{Z}^+$
  $\hat{m}_t = (2q+1)^{-1}\sum_{j=-q}^{q} X_{t+j}$ $\quad q+1 \le t \le n-q$
  Detrended time series: $\hat{Y}_t = X_t - \hat{m}_t$

Mean Function

- $\mu_X(t) = E[X_t]$

Covariance Function

- $\gamma_X(t,s) = Cov(X_t, X_s) = E[X_t X_s] - E[X_t]E[X_s]$

Note: $\gamma_X(t,t) = Var(X_t)$

Properties of Covariance

- For $a \in \mathbb{R}$, $Cov(X,a) = 0$
- For $a,b \in \mathbb{R}$, $Cov(X, aU+bV) = a\,Cov(X,U) + b\,Cov(X,V)$
- $Cov(X,Y)^2 \le Var(X)Var(Y)$

Stationary Time Series

- $\mu_X(t)$ does not depend on t
- $\gamma_X(t,s)$ depends only on $h = t-s$
- Covariance function is non-negative definite

Some useful Properties

- $Cov(A,B) = E[AB] - E[A]E[B]$
- If A, B are independent, then $Cov(A,B) = 0$
- $E[aA + bB] = aE[A] + bE[B]$
- $Cov(A,A) = Var(A)$
- $Var(A+bB) = Var(A) + b^2 Var(B)$ if A, B are independent

MA(1) Model

$\{Z_t\}$ white noise
$\theta \in \mathbb{R}$ $(\theta \ne 0)$
$X_t = Z_t + \theta Z_{t-1}$

- $\mu_X(t) = 0$
- $\gamma_X(t,t+h) = \begin{cases} (1+\theta^2)\sigma_Z^2 \,; & h=0 \\ \theta\sigma_Z^2 \,; & h=1 \\ 0 \,; & h \ge 2 \end{cases}$

Autocorrelation

- $\rho_X(h) = \dfrac{\gamma_X(h)}{\gamma_X(0)} = \dfrac{Cov(X_h, X_0)}{Var(X_0)}$

Partial Autocorrelation

- Correlation between X_t and X_{t+h} after conditioning out the "in-between" variables $X_{t+1}, \dots, X_{t+h-1}$

PACF between X_1 and X_2 when conditioning out X_3:

$$\rho_{12.3} = \frac{Corr(X_1,X_2) - Corr(X_1,X_3)Corr(X_2,X_3)}{\sqrt{1-Corr^2(X_2,X_3)}\,\sqrt{1-Corr^2(X_1,X_3)}}$$

Sample Mean, Sample Autocovariance, Sample Autocorrelation

- Sample Mean: $\hat{\mu} = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$
- Sample Variance: $\hat{\sigma}_X^2 = \hat{\gamma}_X(0) = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$
- Sample Autocovariance:
  $\hat{\gamma}_X(h) = \frac{1}{n-1}\sum_{t=1}^{n-h}(X_t - \bar{X})(X_{t+h} - \bar{X})$
- Sample Autocorrelation: $\hat{\rho}_X(h) = \dfrac{\hat{\gamma}_X(h)}{\hat{\gamma}_X(0)}$
- Sample PACF at lag 2:
  $$\hat{\alpha}(2) = \frac{\hat{\rho}_X(2) - \hat{\rho}_X(1)}{1 - \hat{\rho}_X^2(1)}$$

Linear Processes

Let $\{Z_t\}$ be a white noise
Let $\{\psi_j\}$ be a sequence of constants

$$X_t = \sum_{j=0}^{\infty}\psi_j Z_{t-j}, \quad t=1,2,\dots$$

The model is called:
- linear model
- (infinite order) moving average
- (causal) moving average
- (one-sided) moving average

The linear process is well-defined if $\sum_{j=0}^{\infty}|\psi_j| < \infty$

The linear process is also stationary with
- $E[X_t] = 0$
- $\gamma_X(h) = \sigma_Z^2 \sum_{j=0}^{\infty}\psi_j \psi_{j+h}$

Difference and Backward operators

- Difference operator ∇ defined as
  $\nabla X_t = X_t - X_{t-1}$
- Backward Operator B defined as
  $BX_t = X_{t-1}$

ARMA models

Let $\{Z_t\}$ be a white noise. Then ARMA(p,q) mode is a stationary solution to

$$\phi(B)X_t = \theta(B)Z_t \quad (*)$$

where

$$\phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \dots - \phi_p z^p$$
$$\theta(z) = 1 + \theta_1 z + \theta_2 z^2 + \dots + \theta_q z^q$$

- $\phi(z)$ autoregressive polynomial
- $\theta(z)$ moving average polynomial
- $(*)$ does not guarantee a stationary solution or a causal solution

Stationarity and Causality

The existence of causal and stationary solution means that ARMA model can be written as

$$X_t = \sum_{j=0}^{\infty}\psi_j Z_{t-j}$$

How to check if there is a stationary and causal solution?
- Take AR polynomial $\phi(z)$
- Find roots of ϕ: $\phi(z) = 0$
- If all roots are such that $|z| \ne 1$, then the model is stationary
- If all roots are such that $|z| > 1$, then the model is causal

Linear Representations

Given an ARMA model. If it is stationary and causal, then we can write it as

$$X_t = \sum_{j=1}^{\infty}\psi_j Z_{t-j}$$

The linear representation of the AR(1) model is

$$X_t = \sum_{j=1}^{\infty}\phi^j Z_{t-j} \quad |\phi| < 1$$

The linear representation of the ARMA(1,1) model is

$$X_t = \sum_{j=0}^{\infty}\psi_j Z_{t-j}$$

where $\psi_0 = 1$ and $\psi_j = \phi^{j-1}(\theta + \phi)$

Recursive approach to ACF

To find a recursive definition for the covariance of an AR(p) model, first multiply every term by X_{t-h} in the model. Then apply expected value to all terms and use that to find the recursive definition and initial cases for the covariance.

For AR(1), the ACF is

$$\gamma_X(0) = \sigma_Z^2 \frac{1}{1-\phi^2}, \quad \gamma_X(h) = \phi\,\gamma_X(h-1), \; h \ge 1$$

ACF and PACF of ARMA(p,q) models

For AR(p) models, the autocorrelation function is never 0 for all lags h. The partial autocorrelation function is 0 for all $h > p$.

For MA(q) models, the autocorrelation function is 0 for all $h > q$. The partial autocorrelation function is never 0.

Yule-Walker Formula for Forecasting

Consider a stationary sequence with $\mu_X = E[X_t]$ and autocovariance $\gamma_X(h)$. We denote $P_n X_{n+k}$, the predicted value of X_{n+k} given that we have n observations $X_1, \dots, X_n$.

We use linear predictors:

$$P_n X_{n+k} = a_0 + \sum_{i=1}^{n} a_i X_{n+1-i}$$

We minimize the mean square error

$$E\left[(X_{n+k} - P_n X_{n+k})^2\right]$$

Goal: estimate $a_0, a_1, \dots, a_n$

We obtain $a_0 = \mu\left(1 - \sum_{i=1}^{n} a_i\right)$

If we assume $\mu = 0$, then $a_0 = 0$.

We would then obtain

$$\underline{a_n} = \Gamma_n^{-1}\gamma(n;k)$$

where

$$\Gamma_n = [\gamma_X(i-j)]_{i,j=1}^{n}$$
$$\gamma(n;k) = (\gamma_X(k), \dots, \gamma_X(k+n-1))^T$$
$$\underline{a_n} = (a_1, \dots, a_n)$$

This equation is called the Yule-Walker formula for prediction.

Mean Squared Error:

$$MSPE_n(k) = \gamma_X(0) - \underline{a_n}^T\gamma(n;k)$$

Durbin-Levinson Algorithm

$$P_n X_{n+1} = \phi_{n1}X_n + \dots + \phi_{nn}X_1$$

$a_1 = \phi_{n1}, \dots, a_n = \phi_{nn}$

The coefficients $\phi_{n1}, \dots, \phi_{nn}$ are given by

$$\phi_{nn} = \left[\gamma(n) - \sum_{j=1}^{n-1}\phi_{n-1,j}\gamma(n-j)\right]V_{n-1}^{-1}$$

$$\begin{bmatrix}\phi_{n,1}\\ \vdots \\ \phi_{n,n-1}\end{bmatrix} = \begin{bmatrix}\phi_{n-1,1}\\ \vdots \\ \phi_{n-1,n-1}\end{bmatrix} - \phi_{nn}\begin{bmatrix}\phi_{n-1,n-1}\\ \vdots \\ \phi_{n-1,1}\end{bmatrix}$$

$$V_n = V_{n-1}\left[1 - \phi_{nn}^2\right]; \quad V_0 = \gamma(0)$$

$$\phi_{11} = \rho(1)$$

Partial Autocorrelation

$$\alpha(0) = 1; \quad \alpha(k) = \phi_{kk}\,; \; k \ge 1$$

Prediction Intervals

A prediction interval for X_{n+k} is

$$P_n X_{n+k} \pm Z_{\frac{\alpha}{2}}\sqrt{MSPE_n(k)}$$

For linear process, we have $MSPE_n(k) = \sigma_Z^2 \sum_{j=1}^{k-1}\psi_j^2$

for AR(1) with $k=1$, we have
CI: $\phi X_n \pm Z_{\frac{\alpha}{2}}\sigma_Z$

for AR(1) with $k=2$, we have
CI: $\phi^2 X_n \pm Z_{\frac{\alpha}{2}}\sigma_Z\sqrt{\frac{1+\phi^{2k}}{1-\phi^2}}$

Estimation of the Mean: iid case

We estimate the population mean μ using the sample mean

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad E[\bar{X}] = \mu$$

Using independence, we get

$$Var[\bar{X}] = \frac{\gamma_X(0)}{n}$$

Central Limit Theorem: Assume $X_1, \dots, X_n$ are i.i.d. with mean μ and variance $\gamma_X(0)$. Then

$$\frac{\bar{X}-\mu}{\sqrt{\frac{\gamma_X(0)}{n}}} \longrightarrow N(0,1)$$

The 95% confidence interval for the mean μ is

$$\left(\bar{X} - 1.96\sqrt{\frac{\gamma_X(0)}{n}},\ \bar{X} + 1.96\sqrt{\frac{\gamma_X(0)}{n}}\right)$$

Estimation of the mean: Time Series Case

We still have that $E[\bar{X}] = \mu$.

However, $Var(\bar{X}) = \gamma_X(0) + 2\sum_{h=1}^{\infty}\gamma_X(h)$ whenever the time series is short-range dependent $\sum_{h=-\infty}^{\infty}|\gamma_X(h)| < \infty$

Central Limit Theorem: Assume that $X_1, \dots, X_n$ is a stationary time series with mean μ, variance $\gamma_X(0)$, and the covariance function $\gamma_X(h)$. Then

$$\frac{\bar{X}-\mu}{\frac{\nu}{\sqrt{n}}} \longrightarrow N(0,1)$$

where $\nu^2 = \gamma_X(0) + 2\sum_{h=1}^{\infty}\gamma_X(h)$

The 95% confidence interval for the mean μ is

$$\left(\bar{X} - 1.96\frac{\nu}{\sqrt{n}},\ \bar{X} + 1.96\frac{\nu}{\sqrt{n}}\right)$$

Yule-Walker Estimator

For AR(p) models, we have

$$\Gamma_p = [\gamma_X(i-j)]_{i,j=1}^{p}$$

$$\underline{\Phi}_p = \begin{pmatrix}\phi_1 \\ \vdots \\ \phi_p\end{pmatrix} \qquad \gamma_{X,p} = \begin{pmatrix}\gamma_X(1) \\ \vdots \\ \gamma_X(p)\end{pmatrix}$$

We then get

$$\underline{\Phi}_p = \Gamma_p^{-1}\gamma_{X,p}$$

$$\sigma_Z^2 = \gamma_X(0) - \underline{\Phi}^T\gamma_{X,p}$$

The Yule-Walker estimators are

$$\hat{\underline{\Phi}}_p = \hat{\Gamma}_p^{-1}\hat{\gamma}_{X,p}$$

$$\hat{\sigma}_Z^2 = \hat{\gamma}_X(0) - \hat{\underline{\Phi}}_p\hat{\gamma}_{X,p}$$

$\hat{\Gamma}_p$ and $\hat{\gamma}_{X,p}$ obtained by replacing γ_X with $\hat{\gamma}_X$

Confidence Intervals for AR(p)

For a big sample size, the Yule-Walker estimators are approximately normal:

$$\hat{\underline{\Phi}}_p \sim N_p\left(\underline{\Phi}_p, \frac{1}{n}\sigma_Z^2\Gamma_p^{-1}\right)$$

For $p=1$:

$$\hat{\phi} \sim N\left(\phi, \frac{1}{n}\sigma_Z^2\gamma_X^{-1}(0)\right)$$

Practical CI for ϕ of AR(1):

$$\hat{\phi} \pm z_{\frac{\alpha}{2}}\frac{1}{\sqrt{n}}\hat{\sigma}_Z\sqrt{\hat{\gamma}^{-1}(0)}$$

$$\hat{\phi} = \frac{\hat{\gamma}_X(1)}{\hat{\gamma}_X(0)}, \quad \hat{\sigma}_Z^2 = \hat{\gamma}_X(0) - \hat{\phi}\hat{\gamma}_X(1)$$

Practical CI for ϕ_1, ϕ_2 of AR(2):

$$\hat{\phi}_1 \pm z_{\frac{\alpha}{2}}\frac{1}{\sqrt{n}}\sqrt{1-\hat{\phi}_2^2}$$

$$\hat{\phi}_2 \pm z_{\frac{\alpha}{2}}\frac{1}{\sqrt{n}}\sqrt{1-\hat{\phi}_2^2}$$

Diagnostic Tests

After fitting the model to a time series, we can compute residuals.

Say we fitted the AR(1) model by estimating ϕ by $\hat{\phi}$ and σ_Z^2 by $\hat{\sigma}_Z^2$.

The residuals are: $\hat{Z}_t = X_t - \hat{\phi}X_{t-1}$

We need to check if the residuals have similar properties to white noise.

To do this, we need to check that the residuals do not show significant autocorrelation.

The sample correlations are typically not zero. They follow a $N(0, \frac{1}{n})$ distribution.

The 95% confidence interval for sample correlations is $\pm\frac{1.96}{\sqrt{n}}$.

If any sample correlations fall into this interval we treat them as 0.

Otherwise, they're significant and we cannot assume independence

Ljung-Box Test

Let h be a positive integer and define

$$Q_h = n\sum_{j=1}^{h}\frac{\hat{\gamma}_X(j)}{\hat{\gamma}_X(0)}$$

Under the i.i.d. assumption, $Q_h \sim \chi^2(h)$.

A large value of Q_h suggests that the sample autocorrelation of the data are too large for the data to be a sample from an i.i.d. sample.

We reject the i.i.d. assumption if $Q_h > \chi^2_{1-\alpha}(h)$

Complete Time Series Analysis

- Load and plot data
- Remove Trend
- Plot ACF and PACF and identify model
- Fit model
- Run diagnostic tests
- Make Prediction
- Add trend back

Maximum Likelihood Estimation

iid case

Assume $X_1, \dots, X_n \overset{iid}{\sim} f(x;\theta)$

Goal: estimate θ

Since the random variables are independent, their joint density is

$$f(x_1, \dots, x_n) = \prod_{i=1}^{n} f(x_i)$$

The likelihood function is defined as

$$L(\theta) = L(\theta; X_1, \dots, X_n) = \prod_{i=1}^{n} f(X_i;\theta)$$

The log-likelihood function is defined as $\ell(\theta) = \log L(\theta)$

$$\hat{\theta}_{MLE} = \arg\max_{\theta} L(\theta) = \arg\max_{\theta}\ell(\theta)$$

$\hat{\theta}_{MLE}$ is called the maximum likelihood estimator of θ.

MLE for Time Series Models

$X_1, \dots, X_n$ observations from a stationary time series.

$f(x_1, \dots, x_n)$ their joint density.

Assume time series is Gaussian and centered. This assumption needs to be checked.

Let $X_n = (X_1, \dots, X_n)^T$

$$\hat{X}_n = (\hat{X}_1, \dots, \hat{X}_n)^T$$

$$U_n = (U_1, \dots, U_n)^T$$

$U_i = X_i - \hat{X}_i$, $i = 1, \dots, n$ are the residuals, called some time innovations.

$\Gamma_n = E[X_n^T X_n] = [\gamma_X(i-j)]_{i,j=1}^{n}$ is the covariance matrix of $(X_1, \dots, X_n)$.

The likelihood is

$$L = \frac{1}{(2\pi)^{n/2}}\frac{1}{|\Gamma_n|^{1/2}}\exp\left(-\frac{1}{2}X_n^T\Gamma_n^{-1}X_n\right)$$

$|\Gamma_n|$ is determinant of Γ_n.

The covariance depends on some model parameters.

Estimates are obtained by maximizing L with respect to the parameters.

Usually, there are no explicit formulas and everything is done numerically.

It turns out that

$$X_n^T\Gamma_n^{-1}X_n = U_n^T D^{-1} U_n$$

where D is a diagonal matrix $diag(v_0, \dots, v_{n-1})$. Thus, we have

$$X_n^T\Gamma_n^{-1}X_n = \sum_{i=1}^{n}(X_i - \hat{X}_i)^2/v_{i-1}$$

Furthermore, we also have

$$\det(\Gamma_n) = v_0 \cdots v_{n-1}$$

Hence, the likelihood function takes a form

$$\frac{1}{(2\pi)^{n/2}}\frac{1}{\sqrt{v_0 \cdots v_{n-1}}}\exp\left(-\frac{1}{2}\sum_{i=1}^{n}\frac{(X_i-\hat{X}_i)^2}{v_{i-1}}\right)$$

The likelihood function depends on

$$\beta = (\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q, \sigma_Z^2)$$

The function can be maximized to obtain MLE of β.

The maximum likelihood estimator is asymptotically normal with mean β and variance $n^{-1}V(\beta)$, where $V(\beta)$ is a matrix.

MLE for AR(p) models

For AR(1) model, the likelihood is

$$L(\phi, \sigma_Z) = \frac{1}{(2\pi)^{n/2}}\frac{1}{\sigma_Z^n}\exp\left(-\frac{1}{2}\sum_{i=2}^{n}\frac{(X_i - \phi X_{i-1})^2}{\sigma_Z^2}\right)$$

We get that

$$\hat{\phi}_{MLE} = \frac{\sum_{i=2}^{n}X_{i-1}X_i}{\sum_{i=2}^{n}X_{i-1}^2}$$

$$\hat{\sigma}_{MLE}^2 = \frac{1}{n}\sum_{i=2}^{n}(X_i - \hat{\phi}_{MLE}X_{i-1})^2$$

We also have

$$V(\phi) = (1-\phi^2)$$

MLE of ϕ and its asymptotic variance agree with Yule-Walker

For general AR(p) models, the Yule-Walker estimator and MLE of $(\phi_1, \dots, \phi_p)$ agree and in both cases the asymptotic variance is

$$V(\underline{\Phi}_p) = \sigma_Z^2\Gamma_p^{-1}$$

However, the MLE and Yule-Walker estimators of σ_Z^2 do not need to agree.

An easy approach for AR(p) models is to look at the white noise Z_t.

The likelihood can be rewritten as

$$L = \frac{1}{(\sqrt{2\pi})^n\sigma_Z^n}\exp\left(-\frac{1}{2\sigma_Z^2}\sum_{i=1}^{n}Z_i^2\right)$$

Then $Z_t = X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p}$

This approach only works for AR(p) models, but not MA(q) or ARMA(p,q) models.

Order Selection

A classic order selection is based on AIC (Akaike Information Criteria method). We consider several ARMA(p,q) models that depend on parameters $\underline{\phi} = (\phi_1, \dots, \phi_p)$ and $\underline{\theta} = (\theta_1, \dots, \theta_q)$. We consider several choices of p and q.

For each model, we calculate

$$AIC = 2\log L(\underline{\phi}, \underline{\theta}, \sigma_Z^2) - 2(p+q+1)\frac{n}{n-p-q-2}$$

We choose the model that maximizes the AIC value.