Notes: Intro to Time Series and Forecasting – Ch2 Stationary Processes

Yingbo Li

01/29/2019

Table of Contents

Best linear predictor

- Goal: find a function of X_n that gives the "best" predictor of X_{n+h}.
  - We mean "best" by achieving minimum mean squared error
  - Under joint normality assumption of X_n and X_{n+h}, the best estimator is

$$m(X_n) = E(X_{n+h} \mid X_n) = \mu + \rho(h)(X_n - \mu)$$

- Best linear predictor

$$\ell(X_n) = aX_n + b$$

  - For Gaussian processes, $\ell(X_n)$ and $m(X_n)$ are the same.
  - The best linear predictor only depends on the mean and ACF of the series $\{X_n\}$

Properties of ACVF $\gamma(\cdot)$ and ACF $\rho(\cdot)$

- $\gamma(0) \geq 0$
- $|\gamma(h)| \leq \gamma(0)$ for all h
- $\gamma(h)$ is a even function, i.e., $\gamma(h) = \gamma(-h)$ for all h
- A function $\kappa : \mathbb{N} \to \mathbb{R}$ is nonnegative definite if

$$\sum_{i,j=1}^{n} a_i \kappa(i-j) a_j \geq 0$$

  for all $n \in \mathbb{N}^+$ and vectors $\mathbf{a} = (a_1, \ldots, a_n)' \in \mathbb{R}^n$
- Theorem: a real-value function defined on the integers is the autocovariance function of a stationary time series if and only if it is even and nonnegative definite
- ACF $\rho(\cdot)$ has all above properties of ACVF $\gamma(\cdot)$
  - Plus one more: $\rho(0) = 1$

MA(q) process, q-dependent, and q-correlated

- A time series $\{X_t\}$ is
  - q-dependent: if X_s and X_t are independent for all $|t-s| > q$.
  - q-correlated: if $\rho(h) = 0$ for all $|h| > q$.
- Moving-average process of order q: $\{X_t\}$ is a MA(q) process if

$$X_t = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q}$$

  where $\{Z_t\} \sim \mathrm{WN}(0, \sigma^2)$
- A MA(q) process is q-correlated
- Theorem: a stationary q-correlated time series with mean 0 can be represented as a MA(q) process

Linear processes: definitions

- A time series $\{X_t\}$ is a linear process if

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

where $\{Z_t\} \sim \mathrm{WN}(0, \sigma^2)$, and the constants $\{\psi_j\}$ satisfy

$$\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$$

- Equivalent representation using backward shift operator B

$$X_t = \psi(B) Z_t, \quad \psi(B) = \sum_{j=-\infty}^{\infty} \psi_j B^j$$

- Special case: moving average MA(∞)

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

Linear processes: properties

- In the linear process $X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}$ definition, the condition $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$ ensures
  - The infinite sum X_t converges with probability 1
  - $\sum_{j=-\infty}^{\infty} \psi_j^2 < \infty$, and hence X_t converges in mean square, i.e., X_t is the mean square limit of the partial sum $\sum_{j=-n}^{n} \psi_j Z_{t-j}$

Apply a linear filter to a stationary time series, then the output series is also stationary

- Theorem: let $\{Y_t\}$ be a stationary time series with mean 0 and ACVF γ_Y. If $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$, then the time series

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j Y_{t-j} = \psi(B) Y_t$$

is stationary with mean 0 and ACVF

$$\gamma_X(h) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \psi_j \psi_k \gamma_Y(h+k-j)$$

- Special case of the above result: If $\{X_t\}$ is a linear process, then its ACVF is

$$\gamma_X(h) = \sum_{j=-\infty}^{\infty} \psi_j \psi_{j+h} \sigma^2$$

Combine multiple linear filters

- Linear filters with absoluately summable coefficients

$$\alpha(B) = \sum_{j=-\infty}^{\infty} \alpha_j B^j, \quad \beta(B) = \sum_{j=-\infty}^{\infty} \beta_j B^j$$

can be applied successively to a stationary series $\{Y_t\}$ to generate a new stationary series

$$W_t = \sum_{j=-\infty}^{\infty} \psi_j Y_{t-j}, \quad \psi_j = \sum_{k-\infty}^{\infty} \alpha_k \beta_{j-k} = \sum_{k-\infty}^{\infty} \beta_k \alpha_{j-k}$$

or equivalently,

$$W_t = \psi(B) Y_t, \quad \psi(B) = \alpha(B)\beta(B) = \beta(B)\alpha(B)$$

AR(1) process $X_t - \phi X_{t-1} = Z_t$, in linear process formats

- If $|\phi| < 1$, then

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

  - Since X_t only depends on $\{Z_s, s \leq t\}$, we say $\{X_t\}$ is causal or future-independent

- If $|\phi| > 1$, then

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

  - This is because $X_t = -\phi^{-1} Z_{t+1} + \phi^{-1} X_{t+1}$
  - Since X_t depends on $\{Z_s, s \geq t\}$, we say $\{X_t\}$ is noncausal

- If $\phi = \pm 1$, then there is no stationary linear process solution

ARMA$(1,1)$ process: definitions

- The time series $\{X_t\}$ is a ARMA$(1,1)$ process if it is **stationary** and satisfies

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

where $\{Z_t\} \sim \mathrm{WN}(0, \sigma^2)$ and $\phi + \theta \neq 0$

- Equivalent represention using the backward shift operator

$$\phi(B) X_t = \theta(B) Z_t, \quad \text{where } \phi(B) = 1 - \phi B,\ \theta(B) = 1 + \theta B,$$

ARMA$(1,1)$ process in linear process format

- If $\phi \neq \pm 1$, by letting $\chi(z) = 1/\phi(z)$, we can write an ARMA$(1,1)$ as

$$X_t = \chi(B)\theta(B)Z_t = \psi(B)Z_t, \quad \text{where } \psi(B) = \sum_{j=-\infty}^{\infty} \psi_j B^j$$

  - If $|\phi| < 1$, then $\chi(z) = \sum_{j=0}^{\infty} \phi^j z^j$, and

$$\psi_j = \begin{cases} 0, & \text{if } j \leq -1, \\ 1, & \text{if } j = 0, \\ (\phi + \theta)\phi^{j-1}, & \text{if } j \geq 1 \end{cases} \qquad \text{Causal}$$

  - If $|\phi| > 1$, then $\chi(z) = -\sum_{j=-\infty}^{-1} \phi^j z^j$, and

$$\psi_j = \begin{cases} -(\theta + \phi)\phi^{j-1}, & \text{if } j \leq -1, \\ -\theta\phi^{-1}, & \text{if } j = 0, \\ 0, & \text{if } j \geq 1 \end{cases} \qquad \text{Noncausal}$$

- If $\phi = \pm 1$, then there is no such stationary ARMA$(1,1)$ process

Invertibility

- Invertibility is the dual concept of causaility
  - Causal: X_t can be expressed by $\{Z_s, s \leq t\}$
  - Invertible: Z_t can be expressed by $\{X_s, s \leq t\}$
- For an ARMA$(1, 1)$ process,
  - If $|\theta| < 1$, then it is invertible
  - If $|\theta| > 1$, then it is noninvertible

Estimation of the series mean $\mu = E(X_t)$

- The sample mean $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ is an unbised estimator of μ
  - Mean squared error

$$E(\bar{X}_n - \mu)^2 = \frac{1}{n}\sum_{h=-n}^{n}\left(1 - \frac{|h|}{n}\right)\gamma(h)$$

- Theorem: If $\{X_t\}$ is a stationary time series with mean 0 and ACVF $\gamma(\cdot)$, then as $n \to \infty$,

$$V(\bar{X}_t) = E(\bar{X}_n - \mu)^2 \longrightarrow 0, \quad \text{if } \gamma(n) \to 0,$$

$$nE(\bar{X}_n - \mu)^2 \longrightarrow \sum_{|h|<\infty} \gamma(h), \quad \text{if } \sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty$$

Confidence bounds of μ

- If $\{X_t\}$ is Gaussian, then

$$\sqrt{n}(\bar{X}_n - \mu) \sim \mathrm{N}\left(0, \sum_{|h|<n} \left(1 - \frac{|h|}{n}\right) \gamma(h)\right)$$

- For many common time series, such as linear and ARMA models, when n is large, $\bar{X}_n$ is approximately normal:

$$\bar{X}_n \sim \mathrm{N}\left(\mu, \frac{v}{n}\right), \quad v = \sum_{|h|<\infty} \gamma(h)$$

  - An approximate 95% confidence interval for μ is

$$\left(\bar{X}_n - 1.96 v^{1/2}/\sqrt{n},\ \bar{X}_n + 1.96 v^{1/2}/\sqrt{n}\right)$$

  - To estimate v, we can use

$$\hat{v} = \sum_{|h|<\sqrt{n}} \left(1 - \frac{|h|}{\sqrt{n}}\right) \hat{\gamma}(h)$$

Estimation of ACVF $\gamma(\cdot)$ and ACF $\rho(\cdot)$

- Use sample ACVF $\hat{\gamma}(\cdot)$ and sample ACF $\hat{\rho}(\cdot)$

$$\hat{\gamma}(h) = \frac{1}{n} \sum_{t=1}^{n-|h|} (X_{t+|h|} - \bar{X}_n)(X_t - \bar{X}_n), \quad \hat{\rho}(\cdot) = \hat{\gamma}(h)/\hat{\gamma}(0)$$

  - Even if the factor $1/n$ is replaced by $1/(n-h)$, they are still biased
  - They are nearly unbiased for large n

- When h is slightly smaller than n, the estimators $\hat{\gamma}(\cdot), \hat{\rho}(\cdot)$ are unreliable since there are only few pairs of (X_{t+h}, X_t).
  - A useful guide for them to be reliable (by Jenkins):

$$n \geq 50, \quad h \leq n/4$$

Asymptotic distribution of $\hat{\rho}(\cdot)$

- For linear models, esp ARMA models, when n is large, $\hat{\boldsymbol{\rho}}_k = (\hat{\rho}(1), \ldots, \hat{\rho}(k))'$ is approximately normal

$$\hat{\boldsymbol{\rho}}_k \sim \mathrm{N}(\boldsymbol{\rho}, n^{-1}W)$$

- By Bartlett's formula, W is the covariance matrix with entries

$$w_{ij} = \sum_{k=1}^{\infty} [\rho(k+i) + \rho(k-i) - 2\rho(i)\rho(k)] \times [\rho(k+j) + \rho(k-j) - 2\rho(j)\rho(k)]$$

- Special cases
  - Marginally, for any $j \geq 1$,

$$\hat{\rho}(j) \sim \mathrm{N}(\rho(j), n^{-1}w_{jj})$$

  - iid noise

$$w_{ij} = \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{otherwise} \end{cases} \iff \hat{\rho}(k) \sim \mathrm{N}(0, 1/n),\ k = 1, \ldots, n$$

Best linear predictor: definition

- For a stationary time series $\{X_t\}$ with known mean μ and ACVF γ, our goal is to find the linear combination of $1, X_n, X_{n-1}, \ldots, X_1$ that forecasts X_{n+h} with minimum mean squared error
- Best linear predictor:

$$P_n X_{n+h} = a_0 + a_1 X_n + \cdots + a_n X_1 = a_0 + \sum_{i=1}^{n} a_i X_{n+1-i}$$

  - We need to find the coefficients $a_0, a_1, \ldots, a_n$ that minimize

$$E(X_{n+h} - a_0 - a_1 X_n - \cdots - a_n X_1)^2$$

  - We can take partial derivatives and solve a system of equations

$$E\left[X_{n+h} - a_0 - \sum_{i=1}^{n} a_i X_{n+1-i}\right] = 0,$$

$$E\left[\left(X_{n+h} - a_0 - \sum_{i=1}^{n} a_i X_{n+1-i}\right) X_{n+1-j}\right] = 0, \ j = 1, \ldots, n$$

Best linear predictor: the solution

- Plugging the solution $a_0 = \mu\left(1 - \sum_{i=1}^{n} a_i\right)$ in, the linear pedictor becomes

$$P_n X_{n+h} = \mu + \sum_{i=1}^{n} a_i (X_{n+1-i} - \mu)$$

- The solution of coefficients

$$\mathbf{a}_n = (a_1, \ldots, a_n)' = \mathbf{\Gamma}_n^{-1} \boldsymbol{\gamma}_n(h)$$

  - $\mathbf{\Gamma}_n = [\gamma(i-j)]_{i,j=1}^{n}$ and $\boldsymbol{\gamma}_n(h) = (\gamma(h), \gamma(h+1), \ldots, \gamma(h+n-1))'$

Best linear predictor $\hat{X}_{t+h} = P_n X_{n+h}$: properties

- Unbiasness

$$E(\hat{X}_{t+h} - X_{t+h}) = 0$$

- Mean squared error (MSE)

$$\begin{aligned} E(X_{t+h} - \hat{X}_{t+h})^2 &= E(X_{t+h}^2) - E(\hat{X}_{t+h}^2) \\ &= \gamma(0) - \mathbf{a}_n' \boldsymbol{\gamma}_n(h) \end{aligned}$$

- Orthogonality

$$E\left[(\hat{X}_{t+h} - X_{t+h}) X_j\right] = 0, \quad j = 1, \ldots, n$$

  – In general, orthogonality means

$$E\left[(\mathrm{Error}) \times (\mathrm{PredictorVariable})\right] = 0$$

Example: one-step prediction of an AR(1) series

- We predict X_{n+1} from $X_1, \ldots, X_n$

$$\hat{X}_{n+1} = \mu + a_1(X_n - \mu) + \cdots a_n(X_1 - \mu)$$

- The coefficients $\mathbf{a}_n = (a_1, \ldots, a_n)'$ satisfies

$$\begin{bmatrix} 1 & \phi & \phi^2 & \cdots & \phi^{n-1} \\ \phi & 1 & \phi & \cdots & \phi^{n-2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \phi^{n-1} & \phi^{n-2} & \phi^{n-3} & \cdots & 1 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} \phi_1 \\ \phi^2 \\ \vdots \\ \phi^n \end{bmatrix}$$

- By guessing, we find a solution $(a_1, a_2, \ldots, a_n) = (\phi, 0, \ldots, 0)$, i.e.,

$$\hat{X}_{n+1} = \mu + \phi(X_n - \mu)$$

  - Does not depend on $X_{n-1}, \ldots, X_1$
  - MSE $E(X_{t+1} - \hat{X}_{t+1})^2 = \sigma^2$

WOLG, we can assume $\mu = 0$ while predicting

- A stationary time series $\{X_t\}$ has mean μ
- To predict its future values, we can first create another time series

$$Y_t = X_t - \mu$$

and predict $\hat{Y}_{n+h} = P_n(\hat{Y}_{n+h})$ by

$$\hat{Y}_{n+h} = a_1 Y_n + \cdots + a_n Y_1$$

- Since ACVF $\gamma_Y(h) = \gamma_X(h)$, the coefficients $a_1, \ldots, a_n$ are the same for $\{X_t\}$ and $\{Y_t\}$
- The best linear predictor for $\hat{X}_{n+h} = P_n(\hat{X}_{n+h})$ is

$$\hat{X}_{n+h} - \mu = a_1(X_n - \mu) + \cdots + a_n(X_1 - \mu)$$

Prediction operator $P(\cdot \mid \mathbf{W})$

- X and $W_1, \ldots, W_n$ are random variables with finte 2nd moments
  - Note: $W_1, \ldots, W_n$ does not need to be stationary
- Best linear predictor:

$$\hat{X} = P(X \mid \mathbf{W}) = E(X) + a_1 \left[W_n - E(W_n)\right] + \cdots + a_n \left[W_1 - E(W_1)\right]$$

- Coefficients $\mathbf{a} = (a_1, \ldots, a_n)'$ satisfies

$$\boldsymbol{\Gamma} \mathbf{a} = \boldsymbol{\gamma}$$

where $\boldsymbol{\Gamma} = \left[Cov(W_{n+1-i}, W_{n+1-j})\right]_{i,j=1}^{n}$ and
$\boldsymbol{\gamma} = \left[Cov(X, W_n), \ldots, Cov(X, W_1)\right]'$

Properties of $\hat{X} = P(X \mid \mathbf{W})$

- Unbiased $E(\hat{X} - X) = 0$
- Orthogonal $E[(\hat{X} - X)W_i] = 0$ for $i = 1, \ldots n$
- MSE

$$E(\hat{X} - X)^2 = Var(X) - (a_1, \ldots, a_n) \begin{bmatrix} Cov(X, W_n) \\ \vdots \\ Cov(X, W_1) \end{bmatrix}$$

- Linear

$$P(\alpha_1 X_1 + \alpha_2 X_2 + \beta \mid \mathbf{W}) = \alpha_1 P(X_1 \mid \mathbf{W}) + \alpha_2 P(X_2 \mid \mathbf{W}) + \beta$$

- Extreme cases
  - Perfect prediction

$$P(\sum_{i=1}^{n} \alpha_i W_j + \beta \mid \mathbf{W}) = \sum_{i=1}^{n} \alpha_i W_j + \beta$$

  - Uncorrelated: if $Cov(X, W_i) = 0$ for all $i = 1, \ldots, n$, then

$$P(X \mid \mathbf{W}) = E(X)$$

Examples: predictions of AR(p) series

- A time series $\{X_t\}$ is an autoregression of order p, i.e., AR(p), if it is **stationary** and satisfies

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + Z_t$$

  where $\{Z_t\} \sim \mathrm{WN}(0, \sigma^2)$, and $Cov(X_s, Z_t) = 0$ for all $s < t$

- When $n > p$, the one-step prediction of an AR(p) series is

$$P_n X_{n+1} = \phi_1 X_n + \phi_2 X_{n-1} + \cdots + \phi_p X_{n+1-p}$$

  with MSE $E\left(X_{n+1} - P_n X_{n+1}\right)^2 = E(Z_{n+1})^2 = \sigma^2$

- h-step prediction of an AR(1) series (proof by recursions)

$$P_n X_{n+h} = \phi^h X_n, \quad \mathrm{MSE} = \sigma^2 \frac{1 - \phi^{2h}}{1 - \phi^2}$$

Recursive methods for one-step prediction

- The best linear predictor solution $\mathbf{a} = \mathbf{\Gamma}^{-1}\boldsymbol{\gamma}$ needs matrix inversion
- Alternatively, we can use recursion to simplify one-step prediction of $P_n X_{n+1}$, based on $P_j X_{j+1}$ for $j = 1, \ldots, n-1$
- We will introduce
  - Durbin-Levinson algorithms: good for AR(p)
  - Innovation algorithm: good for MA(q); innovations are uncorrelated

Durbin-Levinson algorithm

- Assume $\{X_t\}$ is mean zero, stationary, with ACVF $\gamma(h)$

$$\hat{X}_{n+1} = \phi_{n,1}X_n + \cdots \phi_{n,n}X_1, \quad \text{with MSE } v_n = E(\hat{X}_{n+1} - X_{n+1})^2$$

1. Start with $\hat{X}_1 = 0$ and $v_0 = \gamma(0)$

For $n = 1, 2, \ldots$, compute step 2-4 successively

2. Compute $\phi_{n,n}$ (partial autocorrelation function (PACF) at lag n)

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

3. Compute $\phi_{n,1}, \ldots, \phi_{n,n-1}$

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

4. Compute v_n

$$v_n = v_{n-1}(1 - \phi_{n,n}^2)$$

Innovation algorithm

- Assume $\{X_t\}$ is any mean zero (not necessarily stationary) time series with covariance $\kappa(i,j) = Cov(X_i, X_j)$
- Predict $\hat{X}_{n+1} = P_n X_{n+1}$ based on innovations, or one-step prediction errors $X_j - \hat{X}_j$, $j = 1, \ldots, n$

$$\hat{X}_{n+1} = \theta_{n,1}(X_n - \hat{X}_n) + \cdots + \theta_{n,n}(X_1 - \hat{X}_1) \quad \text{with MSE } v_n$$

1. Start with $\hat{X}_1 = 0$ and $v_0 = \kappa(1,1)$

For $n = 1, 2, \ldots$, compute step 2-3 successively

2. For $k = 0, 1, \ldots, n-1$, compute coefficients

$$\theta_{n,n-k} = \left[\kappa(n+1, k+1) - \sum_{j=0}^{k-1} \theta_{k,k-j}\theta_{n,n-j}v_j\right] / v_k$$

3. Compute the MSE

$$v_n = \kappa(n+1, n+1) - \sum_{j=0}^{n-1} \theta_{n,n-j}^2 v_j$$

h-step predictors using innovations

- For any $k \geq 1$, orthoganlity ensures

$$E\left[\left(X_{n+k} - P_{n+k-1}X_{n+k}\right)X_j\right] = 0, \quad j = 1, \ldots, n$$

  Thus, we have

$$P_n(X_{n+k} - P_{n+k-1}X_{n+k}) = 0$$

- The h-step prediction:

$$\begin{aligned} P_n X_{n+h} &= P_n P_{n+h-1} X_{n+h} \\ &= P_n \left[\sum_{j=1}^{n+h-1} \theta_{n+h-1,j} \left(X_{n+h-j} - \hat{X}_{n+h-j} \right) \right] \\ &= \sum_{j=h}^{n+h-1} \theta_{n+h-1,j} \left(X_{n+h-j} - \hat{X}_{n+h-j} \right) \end{aligned}$$

References

- Brockwell, Peter J. and Davis, Richard A. (2016), *Introduction to Time Series and Forecasting, Third Edition*. New York: Springer