Stationary Processes (Chapter 2)

2.1 Basic Properties

$\{X_t\}$ stationary time series

Mean: $\mu = \mathrm{E}\, X_t$ for all t.

ACVF: $\gamma(h) = \mathrm{cov}(X_{t+h}, X_t)$, h=0, ±1, . . .

ACF: $\rho(h) = \dfrac{\gamma(h)}{\gamma(0)}$

1

DEFINITION. $\{X_t\}$ is a **Gaussian time series** if all of its joint distributions are multivariate normal, i.e. $(X_1, \ldots, X_n)$ is multivariate normal for all n.

Problem: Predict X_{n+h} from X_n . (Assume $\{X_t\}$ is a stationary Gaussian time series.)

Soln: $X_{n+h} \mid X_n$ has a normal distribution with

conditional mean:

$$\mathrm{E}(X_{n+h} \mid X_n) = \mathrm{E}\, X_{n+h} + \frac{\mathrm{Cov}(X_{n+h}, X_n)}{\mathrm{Var}(X_n)}(X_n - \mathrm{E}\, X_n),$$
$$= \mu + \rho(h)(X_n - \mu)$$

2

conditional variance:

$$\mathrm{Var}(X_{n+h} \mid X_n) = \gamma(0)(1 - \rho^2(h)).$$

The "best" mean square predictor of X_{n+h} in terms of X_n is then

$$\mu + \rho(h)(X_n - \mu)$$

Remark: For Gaussian time series,

best MS predictor = best linear predictor

3

Basic Properties of $\gamma(h)=\mathrm{Cov}(X_{t+h}, X_t)$:

(i) $\gamma(0) \geq 0$,

(ii) $|\gamma(h)| \leq \gamma(0)$,

(iii) $\gamma(h) = \gamma(-h)$, ($\gamma(.)$ is an even function)

Thm 2.1.1: $\gamma(.)$ is the ACVF of a stationary TS iff

(i) $\gamma(.)$ is an even function

(ii) $\gamma(.)$ is **non-negative definite** (nnd)

$$\sum_{i,j=1}^{n} a_i \gamma(i-j)\, a_j \geq 0 \quad \text{for all } n.$$

4

Ex. Is $\gamma(h) := \cos(\omega h)$ an ACVF?

Ans. Yes. Verification of nnd property is difficult.

Easier to check that $\cos(\omega h)$ is the ACVF of

$$X_t = A \cos(\omega t) + B \sin(\omega t)$$

where A & B are uncorrelated (0,1) rv's.

Ex 2.1.1. The function

$$\gamma(h) := \begin{cases} 1, & \text{if } h = 0, \\ \rho, & \text{if } h = 1 \text{ or } -1, \\ 0, & \text{otherwise.} \end{cases}$$

is an ACVF iff $|\rho| \leq .5$

5

Problem 2.4. Which of the following functions are ACVFs?

(a) $\gamma(h) = (-1)^{|h|}$

(b) $\gamma(h) = 1 + \cos(\pi h/2) + \cos(\pi h/4)$

(c) $\gamma(h) = \begin{cases} 1, & \text{if } h = 0, \\ .4, & \text{if } h = 1 \text{ or } -1, \\ 0, & \text{otherwise.} \end{cases}$

6

DEFINITION. $\{X_t\}$ is a **strictly stationary time series**

if $\quad (X_1, \ldots, X_n) \stackrel{d}{=} (X_{1+h}, \ldots, X_{n+h})$

for all integers h and $n > 0$.

Properties of strictly stationary time series:

(a) The X_i's are identically distributed.

(b) $(X_t, X_{t+h}) \stackrel{d}{=} (X_1, X_{1+h})$ for all integers t and h.

(c) $\{X_t\}$ is weakly stationary if $E X_t^2 < \infty$.

(d) weak stationarity $\not\Rightarrow$ strict stationarity.
(true for Gaussian time series)

(e) IID sequences are strictly stationary.

7

2.2 Linear Time Series:

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}, \quad \{Z_t\} \sim \mathrm{WN}(0,\sigma^2),$$

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

Write

$$X_t = (\cdots + \psi_{-1}B^{-1} + \psi_0 B^0 + \psi_1 B^1 + \cdots)\, Z_t$$
$$= \psi(B)\, Z_t,$$

where $\psi(B)$ is the operator (filter)

$$\psi(B) = (\cdots + \psi_{-1}B^{-1} + \psi_0 B^0 + \psi_1 B^1 + \cdots).$$

8

Properties of linear processes $X_t = \psi(B)\, Z_t$. (see p.52)

- weakly stationary (strictly stationary if $\{Z_t\} \sim$ IID)
- $E\, X_t = 0$.
- $\gamma(h) = \sum_{j=-\infty}^{\infty} \psi_j\, \psi_{j+h}\, \sigma^2$ for $h \geq 0$.

Ex 2.2.1 (An AR(1) process). Let $\{X_t\}$ be a stationary solution of the difference equations

$$X_t = \phi\, X_{t-1} + Z_t \ , \quad \{Z_t\} \sim \mathrm{WN}(0,\sigma^2).$$

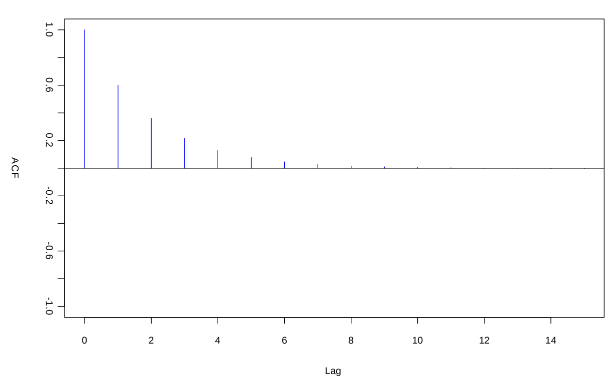
If $|\phi|<1$, the unique stationary solution is given by

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

ACVF of an AR(1) process:

$$\gamma(h) = \sum_{j=0}^{\infty} \phi^j\, \phi^{j+h}\, \sigma^2 = \frac{\phi^h \sigma^2}{1-\phi^2}, \quad h \geq 0 \,.$$

$$\rho(h) = \phi^{|h|} \,.$$



2.3 Introduction to ARMA processes

Definition: $\{X_t\}$ is an ARMA(1,1) process if it is stationary and satisfies (for every t),

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

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

Note: Using the backward shift operator we can rewrite the ARMA equations as

$$(1-\phi B) X_t = (1+\theta B) Z_t$$

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

where $\phi(B) = (1-\phi B)$ and $\theta(B) = (1+\theta B)$ are the AR and MA polynomials respectively.

11

Case $|\phi| < 1$: Setting

$$\chi(B) = 1 + \phi B + \phi^2 B^2 + \phi^3 B^3 + \ldots$$

we see that

$$\chi(B)\,\phi(B) = (1 + \phi B + \phi^2 B^2 + \phi^3 B^3 + \ldots)(1-\phi B) = 1$$

and hence

$$\chi(B)\,\phi(B) X_t = \chi(B)\,(1+\theta B) Z_t$$

or

$$X_t = \chi(B)(1+\theta B) Z_t$$

$$= \psi(B) Z_t = Z_t + \psi_1 Z_{t-1} + \psi_2 Z_{t-2} + \ldots,$$

where $\psi(B) = 1 + \psi_1 B + \psi_2 B^2 + \psi_3 B^3 + \ldots$

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

12

Case $|\phi| < 1$:

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

is a ***causal function*** of $\{Z_t\}$.

Case $|\phi| > 1$: Using a similar argument as before, one can show that

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

is a ***noncausal function*** of $\{Z_t\}$.

Case $|\phi| = 1$: No stationary solution exists.

13

2.4 Properties of sample mean and ACF

Sample mean:

$$\bar{X}_n = n^{-1} \sum_{t=1}^{n} X_t$$

Properties of $\bar{X}_n$:

$$E\,\bar{X}_n = \mu \quad \text{(unbiased)}$$

$$\mathrm{Var}(\bar{X}_n) = n^{-2} \sum_{j=1}^{n} \sum_{j=1}^{n} \mathrm{Cov}(X_i, X_j)$$

$$= n^{-1} \sum_{|h|<n} (1-|h|/n)\gamma(h)$$

14

Hence,

$$\mathrm{Var}(\overline{X}_n) \longrightarrow 0 \quad \text{if } \gamma(h) \longrightarrow 0.$$

$$n\mathrm{Var}(\overline{X}_n) \longrightarrow \sum_{h=-\infty}^{\infty} \gamma(h) \quad \text{if } \sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty.$$

For a large class of time series models (including linear),

$$n^{1/2}(\overline{X}_n - \mu) \quad \text{is approx} \quad N(0, \sum_h \gamma(h)).$$

Approx 95% Confidence Interval for μ:

$$\overline{X}_n \pm 1.96\, \hat{v}^{1/2}/\sqrt{n}$$

where

$$\hat{v} = \sum_{|h|<\sqrt{n}} (1-|h|/n^{.5})\hat{\gamma}(h) \quad \text{(computed in ITSM)}$$

15

Ex 2.4.1 (An AR(1) process). Let $\{X_t\}$ be the AR(1) process with mean μ ,

$$X_t - \mu = \phi\,(X_{t-1} - \mu) + Z_t\ ,\quad \{Z_t\} \sim \mathrm{WN}(0,\sigma^2).$$

In this case, the asymptotic variance of $\overline{X}_n$ is

$$\sum_{h=-\infty}^{\infty} \gamma(h) = (1 + 2\sum_{h=1}^{\infty} \phi^h)\, \sigma^2/(1-\phi^2)$$

$$= \sigma^2/(1-\phi)^2$$

95% confidence bounds are given by:

$$\overline{X}_n \pm 1.96\, n^{-.5}\sigma/(1-\phi)$$

where σ and ϕ are replaced by estimates.

16

2.4.2 Sample ACF

Recall that

$$\hat{\gamma}(h) = n^{-1} \sum_{t=1}^{n-h} (X_{t+h} - \bar{X})(X_t - \bar{X})$$

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

Estimators are biased even if n^{-1} is replaced by $(n-h)^{-1}$.

17

Sample covariance matrix,

$$\hat{\Gamma}_k = \begin{bmatrix} \hat{\gamma}(0) & \hat{\gamma}(1) & \cdots & \hat{\gamma}(k-1) \\ \hat{\gamma}(1) & \hat{\gamma}(0) & \cdots & \hat{\gamma}(k-2) \\ \vdots & \vdots & \cdots & \vdots \\ \hat{\gamma}(k-1) & \hat{\gamma}(k-2) & \cdots & \hat{\gamma}(0) \end{bmatrix}$$

is positive definite for all $k \geq 1$.

18

For linear time series models,

$$\hat{\rho}(h) \quad \text{approx} \quad N(\rho(h)\,,\, n^{-1} w_{hh})$$

where

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

(Bartlett's Formula)

Ex 2.4.2 (IID noise). $w_{ij} = 1$ if $i = j$, 0 otherwise.

$\hat{\rho}(1), \ldots, \hat{\rho}(h)$ approx IID $N(0,n^{-1})$.

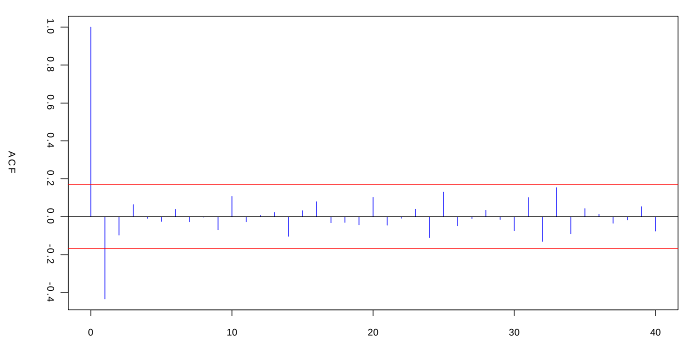
19

Ex. 2.4.3 (An MA(1) process).

$$X_t = Z_t + \theta Z_{t-1}, \quad \{Z_t\} \sim WN(0,\sigma^2).$$

$$w_{ii} = \begin{cases} 1-3\rho^2(1)+4\rho^4(1), & \text{if } i = 1, \\ 1+2\rho^2(1), & \text{if } i > 1. \end{cases}$$

Figure 2.1. Sample ACF of MA(1), $\theta = -.8$, $\rho(1) = -.49$.



20

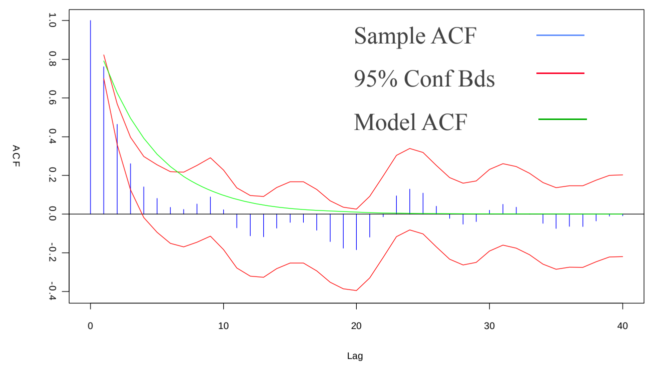
Ex 2.4.4 (An AR(1) process--Lake Huron Residuals).

$X_t = \phi X_{t-1} + Z_t, \quad \rho(h) = \phi^{|h|}$.

$w_{ii} = (1-\phi^{2i})(1+\phi^2)(1-\phi^2)^{-1}-2i\phi^{2i}$

Lake Huron Residuals $y_1, \ldots, y_{98}$.

Model : $Y_t = .791\ Y_{t-1} + Z_t$



21

2.5 Forecasting Stationary Time Series

Suppose $\{X_t\}$ is stationary with mean μ , ACVF $\gamma(.)$.

Linear Prediction Operator P_n :

$$P_n(X_{n+h}) = \text{“best” linear predictor of } X_{n+h} \text{ in terms of } 1, X_1, \ldots, X_n.$$

$$= a_0 + a_1 X_n + \cdots + a_n X_1$$

where $a_0, a_1, \ldots, a_n$, are chosen to minimize

$$S(\boldsymbol{a}) = E(X_{n+h} - a_0 - a_1 X_n - \ldots - a_n X_1)^2.$$

22

First show that $a_0 = \mu(1 - \sum_{i=1}^{n} a_i)$ and hence

$$P_n(X_{n+h}) = \mu + a_1 (X_n - \mu) + \cdots + a_n (X_1 - \mu)$$

where

$$\mathbf{a}_n = (a_1, \ldots, a_n)' = \Gamma_n^{-1} \gamma_n(h)$$

$$\gamma_n(h) = (\gamma(h), \gamma(h+1), \ldots, \gamma(h+n-1))'$$
$$= (\mathrm{Cov}(X_{n+h}, X_n), \ldots, \mathrm{Cov}(X_{n+h}, X_1))'$$

Main principle: $P_n(X_{n+h})$ is chosen so that

$$X_{n+h} - P_n(X_{n+h}) \perp 1, X_n, \ldots, X_1.$$

Prediction Operator P(. | W)

$$\mathbf{W} = (W_n, \ldots, W_1)'$$

$$P(Y|\mathbf{W}) = \mu_Y + \mathbf{a}'(\mathbf{W} - \mu_W) \qquad \Gamma\mathbf{a} = \gamma.$$

Properties of the Prediction Operator $P(\cdot|\mathbf{W})$:

Suppose that $EU^2 < \infty$, $EV^2 < \infty$, $\Gamma = \mathrm{cov}(\mathbf{W}, \mathbf{W})$, and $\beta, \alpha_1, \ldots, \alpha_n$ are constants.

1. $P(U|\mathbf{W}) = EU + \mathbf{a}'(\mathbf{W} - E\mathbf{W})$, where $\Gamma\mathbf{a} = \mathrm{cov}(U, \mathbf{W})$.
2. $E[(U - P(U|\mathbf{W}))\mathbf{W}] = \mathbf{0}$ and $E[U - P(U|\mathbf{W})] = 0$.
3. $E[(U - P(U|\mathbf{W}))^2] = \mathrm{var}(U) - \mathbf{a}'\mathrm{cov}(U, \mathbf{W})$.
4. $P(\alpha_1 U + \alpha_2 V + \beta|\mathbf{W}) = \alpha_1 P(U|\mathbf{W}) + \alpha_2 P(V|\mathbf{W}) + \beta$.
5. $P\left(\sum_{i=1}^{n} \alpha_i W_i + \beta|\mathbf{W}\right) = \sum_{i=1}^{n} \alpha_i W_i + \beta$.
6. $P(U|\mathbf{W}) = EU$ if $\mathrm{cov}(U, \mathbf{W}) = \mathbf{0}$.
7. $P(U|\mathbf{W}) = P(P(U|\mathbf{W}, \mathbf{V})|\mathbf{W})$ if $\mathbf{V}$ is a random vector such that the components of $E(\mathbf{V}\mathbf{V}')$ are all finite.

2.5.1 Durbin - Levinson Algorithm

Assuming the series $\{X_n\}$ has zero mean,

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

and the MSE is $v_n = E(X_{n+1} - P_n(X_{n+1}))^2$.

The Durbin-Levinson Algorithm recursively computes the coefficients $(\phi_{n1}, \dots, \phi_{nn})$ and v_n from $(\phi_{n-1,1}, \dots, \phi_{n-1,n-1})$ and v_{n-1} . (see p. 70).

25

2.5.1 Durbin - Levinson Algorithm

$$P_n X_{n+1} = \phi_n' \mathbf{X}_n = \phi_{n1} X_n + \cdots + \phi_{nn} X_1,$$

The Durbin–Levinson Algorithm:

The coefficients $\phi_{n1}, \dots, \phi_{nn}$ can be computed recursively from the equations

$$\phi_{nn} = \left[\gamma(n) - \sum_{j=1}^{n-1} \phi_{n-1,j}\gamma(n-j)\right] v_{n-1}^{-1}, \qquad (2.5.20)$$

$$\begin{bmatrix} \phi_{n1} \\ \vdots \\ \phi_{n,n-1} \end{bmatrix} = \begin{bmatrix} \phi_{n-1,1} \\ \vdots \\ \phi_{n-1,n-1} \end{bmatrix} - \phi_{nn} \begin{bmatrix} \phi_{n-1,n-1} \\ \vdots \\ \phi_{n-1,1} \end{bmatrix} \qquad (2.5.21)$$

and

$$v_n = v_{n-1}\left[1 - \phi_{nn}^2\right], \qquad (2.5.22)$$

where $\phi_{11} = \gamma(1)/\gamma(0)$ and $v_0 = \gamma(0)$.

26

2.5.2 The Innovations Algorithm

Write,

$\hat{X}_{n+1} = P_n(X_{n+1})$ (= 0 if n=0),

$v_n = E(X_{n+1} - P_n(X_{n+1}))^2$.

The innovations, $X_1 - \hat{X}_1, X_2 - \hat{X}_2, \ldots, X_n - \hat{X}_n$, are orthogonal and we can write

$$\hat{X}_{n+1} = \theta_{n1}(X_n - \hat{X}_n) + \cdots + \theta_{nn}(X_1 - \hat{X}_1)$$

The innovations algorithm is a recursive procedure for computing the coefficients $(\theta_{n1}, \ldots, \theta_{nn})$ and v_n from $(\theta_{n-1,1}, \ldots, \theta_{n-1,n-1})$ and v_{n-1} (see p. 73).

27

2.5.2 The Innovations Algorithm

$$\hat{X}_{n+1} = \begin{cases} 0, & \text{if } n = 0, \\ \sum_{j=1}^{n} \theta_{nj}\left(X_{n+1-j} - \hat{X}_{n+1-j}\right), & \text{if } n = 1, 2, \ldots, \end{cases} \qquad (2.5.28)$$

The Innovations Algorithm:

The coefficients $\theta_{n1}, \ldots, \theta_{nn}$ can be computed recursively from the equations

$$v_0 = \kappa(1, 1),$$

$$\theta_{n,n-k} = v_k^{-1}\left(\kappa(n+1, k+1) - \sum_{j=0}^{k-1} \theta_{k,k-j}\theta_{n,n-j}v_j\right), \quad 0 \le k < n,$$

and

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

(It is a trivial matter to solve first for v_0, then successively for $\theta_{11}, v_1; \theta_{22}, \theta_{21}, v_2; \theta_{33}, \theta_{32}, \theta_{31}, v_3; \ldots$.)

Summary

Durbin-Levinson:

$$\hat{X}_{n+1} = \phi_{n1}X_n + \ldots + \phi_{nn}X_1$$

(Useful for time series with AR structure.)

Innovations Algorithm:

$$\hat{X}_{n+1} = \theta_{n1}(X_n - \hat{X}_n) + \cdots + \theta_{nn}(X_1 - \hat{X}_1)$$

(Useful for time series with MA structure, i.e.,

time series with $\gamma(h) = 0$ for some $|h| > q$.)

29