

ANDREW TULLOCH

TIME SERIES AND MONTE
CARLO INFERENCE

TRINITY COLLEGE
THE UNIVERSITY OF CAMBRIDGE

Contents

1	<i>Time Series Analysis</i>	5
1.1	<i>Introduction</i>	5
1.2	<i>Stationary Processes</i>	6
1.3	<i>State Space Models</i>	8
1.4	<i>Stationary Processes</i>	8
1.4.1	<i>Linear Processes</i>	8
1.4.2	<i>Forecasting Stationary Time Series</i>	9
1.4.3	<i>Innovation Algorithm</i>	11
1.5	<i>ARMA Processes</i>	12
1.5.1	<i>ACF and PACF of an ARMA(p, q) Process</i>	13
1.5.2	<i>Forecasting ARMA Processes</i>	14
1.6	<i>Estimation of ARMA Processes</i>	16
1.6.1	<i>Yule-Walker Equations</i>	16
1.6.2	<i>Estimation for Moving Average Processes Using the Innovations Algorithm</i>	18
1.6.3	<i>Maximum Likelihood Estimation</i>	19
1.6.4	<i>Order Selection</i>	20
1.7	<i>Spectral Analysis</i>	21
1.7.1	<i>The Spectral Density of an ARMA Process</i>	22
1.7.2	<i>The Periodogram</i>	23

2 *Bibliography* 25

1

Time Series Analysis

1.1 Introduction

References:

(i) Brockwell and Davis [2009]

(ii) Brockwell and Davis [2002]

Definition 1.1 (Time Series). A set of observations (X_t) , each being recorded at a predictable time $t \in T_0$.

In a continuous time series, T_0 is continuous. In a discrete time series, T_0 is discrete.

Definition 1.2 (Time Series Model). Specification of joint distribution (or only means and covariances) of a sequence of random variables of which X_t is a realization.

Remark 1.3. A complete probability model specifies the joint distribution of all the random variables $X_t, t \in T$.

This often requires too many estimators, so we only specify the first and second order moments.

Example 1.4. When X_t is multivariate IID -

$$\mathbb{P}(X_1 = x_1, \dots, X_n = x_n) = \prod_{i=1}^n F(x_i) \quad (1.1)$$

Example 1.5. First order moving average model

Example 1.6. Trend and seasonal component.

1.2 Stationary Processes

Intuitively, a stationary time series is one where the joint distribution is invariant to time shifts.

Definition 1.7 (Mean, Covariance function). Define the mean function $\mu_X(t) = \mathbb{E}(X_t)$.

Define the covariance function $\gamma_X(t, s) = \text{Cov}(X_t, X_s) = \mathbb{E}((X_t - \mu_X(t))(X_s - \mu_X(s)))$.

Definition 1.8 (Weak Stationarity). A time series X_t is stationary if

- (i) $\mathbb{E}(|X_t|^2) < \infty$ for all $t \in \mathbb{Z}$
- (ii) $\mathbb{E}(X_t) = c$ for all $t \in \mathbb{Z}$
- (iii) $\gamma_X(t, s) = \gamma_X(t + h, s + h)$ for all $t, s, h \in \mathbb{Z}$

Definition 1.9 (Strict Stationarity). A time series X_t is said to be strict stationary if the joint distributions of X_{t_1, \dots, t_k} and X_{t_1+h, \dots, t_k+h} are identical for all k and for all $t_1, \dots, t_k, h \in \mathbb{Z}$.

Definition 1.10 (Autocovariance function). For a stationary time series X_t , define the autocovariance function

$$\gamma_X(t) = \text{Cov}(X_{t+h}, X_t). \quad (1.2)$$

and the autocorrelation function

$$\rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)}. \quad (1.3)$$

Lemma 1.11 (Properties of the autocovariance function).

$$\gamma(0) \geq 0 \quad (1.4)$$

$$|\gamma(h)| \leq \gamma(0) \quad (1.5)$$

$$\gamma(h) = \gamma(-h) \quad (1.6)$$

for all h .

Note that these all hold for the autocorrelation function ρ , with the additional condition that $\rho(0) = 1$.

Theorem 1.12. A real-valued function defined on the integers is the autocovariance function of a stationary time series if and only if it is even and nonnegative definite.

Example 1.13. Consider a white noise, with X_t a time series with X_t uncorrelated with mean zero and variance σ^2 .

Then

$$\gamma_X(h) = \sigma^2 \mathbb{I}(h = 0) \quad (1.7)$$

$$\rho_X(h) = \mathbb{I}(h = 0) \quad (1.8)$$

Example 1.14 (First order moving average MA(1)).

$$X_t = Z_t + \theta Z_{t-1} \quad (1.9)$$

with $Z_t \sim \text{WN}(0, \sigma^2)$. Then

$$\gamma_X(h) = \begin{cases} \sigma^2(1 + \theta^2) & h = 0 \\ \sigma^2\theta & |h| = 1 \\ 0 & \text{otherwise} \end{cases} \quad (1.10)$$

$$\rho_X(h) = \begin{cases} 1 & h = 0 \\ \frac{\theta}{1+\theta} & |h| = 1 \\ 0 & \text{otherwise} \end{cases} \quad (1.11)$$

Definition 1.15 (Sample Autocovariance). The sample autocovariance function of $\{x_1, \dots, x_n\}$ is defined by

$$\hat{\gamma}(h) = \frac{1}{n} \sum_{j=1}^{n-h} (x_{j+h} - \bar{x})(x_j - \bar{x}), 0 \leq h < n \quad (1.12)$$

and $\hat{\gamma}(h) = \hat{\gamma}(-h)$, $-n < h \leq 0$.

Note that the divisor is n rather than $n - h$ since this ensures that the sample autocovariance matrix

$$\hat{\Gamma}_n = (\hat{\gamma}(i - j))_{i,j} \quad (1.13)$$

is positive semidefinite.

1.3 State Space Models

Definition 1.16. The observation equation is

$$Y_t = G_t X_t + W_t. \quad (1.14)$$

The state equation is

$$X_{t+1} = F_t X_t + V_t \quad (1.15)$$

$\{Y_t\}$ has a state-space representation if there exists a state-space model for $\{Y_t\}$ as specified by the previous equations.

Theorem 1.17 (De Finitte). *If $\{X_1, V_1, V_2, \dots\}$ are independent, then $\{X_t\}$ has the Markov property - that is, $X_{t+1}|X_t, X_{t-1}, \dots = X_{t+1}|X_t$.*

¹

¹ All of Section 8.1 in Introduction to Time Series and Forecasting

In the stable case, there is a unique stationary solution, given by

$$X_t = \sum_{j=0}^{\infty} F^j V_{t-j-1} \quad (1.16)$$

Definition 1.18. The state equation is said to be “stable” if the matrix F has all its eigenvalues in the interior of the unit circle .

1.4 Stationary Processes

1.4.1 Linear Processes

Definition 1.19 (Wold Decomposition). If X_t is a nondeterministic stationary time series, then

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j} + V_t \quad (1.17)$$

where

- (i) $\psi_0 = 1$ and $\sum_{j=0}^{\infty} \psi_j^2 < \infty$,
- (ii) $Z_t \sim WN(0, \sigma^2)$,
- (iii) $\text{Cov}(Z_s, V_t) = 0$ for all s, t ,
- (iv) $Z_t = \tilde{P}_t Z_t$ for all t ,

(v) $V_t = \tilde{P}_s V_t$ for all s, t ,

(vi) V_t is deterministic.

The sequences Z_t, ψ_j, V_t are unique and can be written explicitly as

$$Z_t = X_t - \tilde{P}_{t-1} X_t \quad (1.18)$$

$$\psi_j = \frac{\mathbb{E}(X_t Z_{t-j})}{\mathbb{E}(Z_t)^2} \quad (1.19)$$

$$V_t = X_t - \sum_{j=0}^{\infty} \psi_j Z_{t-j}. \quad (1.20)$$

Definition 1.20. A times series $\{X_t\}$ is a **linear process** if it has the representation

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j} \quad (1.21)$$

where $Z_t \sim WN(0, \sigma^2)$ and $\{\psi_j\}$ is a sequence of constants with $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$.

A linear process is called a **moving average** or $MA(\infty)$ if $\psi_j = 0$ for all $j < 0$, so

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}. \quad (1.22)$$

Proposition 1.21. Let Y_t be a stationary time series with mean zero and covariance function γ_Y . If $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$, then the time series

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j Y_{t-j} = \psi(B)Y_t \quad (1.23)$$

is stationary with mean zero and autocovariance function

$$\gamma_X(h) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \psi_j \psi_k \gamma_Y(h+k-j). \quad (1.24)$$

In the special case where X_t is a linear process,

$$\gamma_X(h) = \sum_{j=-\infty}^{\infty} \psi_j \psi_{j+h} \sigma^2. \quad (1.25)$$

1.4.2 Forecasting Stationary Time Series

Our goal is to find the linear combination of $1, X_n, X_{n-1}, \dots, X_1$ that forecasts X_{n+h} with minimum mean squared error. The best linear

predictor in terms of $1, X_n, \dots, X_1$ will be denoted by $P_n X_{n+h}$ and clearly has the form

$$P_n X_{n+h} = a_0 + a_1 X_n + \dots + a_n X_1. \quad (1.26)$$

To find these equations, we solve the convex problem by setting derivatives to zero, and obtain the result given below.

Theorem 1.22 (Properties of h -step best linear predictor $P_n X_{n+h}$). (i)

$$P_n X_{n+h} = \mu + \sum_{i=1}^n a_i (X_{n+1-i} - \mu) \quad (1.27)$$

where $\mathbf{a}_n = (a_1, \dots, a_n)$ satisfies

$$\Gamma_n \mathbf{a}_n = \gamma_n(h) \quad (1.28)$$

$$\Gamma_n = [\gamma(i-j)]_{i,j=1}^n \quad (1.29)$$

$$\gamma_n(h) = (\gamma(h), \gamma(h+1), \dots, \gamma(h+n-1)) \quad (1.30)$$

(ii)

$$\mathbb{E}\left((X_{n+h} - P_n X_{n+h})^2\right) = \gamma(0) - \langle \mathbf{a}_n, \gamma_n(h) \rangle \quad (1.31)$$

(iii)

$$\mathbb{E}(X_{n+h} - P_n X_{n+h}) = 0 \quad (1.32)$$

(iv)

$$\mathbb{E}((X_{n+h} - P_n X_{n+h})X_j) = 0 \quad (1.33)$$

for $j = 1, \dots, n$.

Definition 1.23 (Prediction Operator $P(\cdot|\mathbf{W})$). Suppose that $\mathbb{E}(U^2) < \infty$, $\mathbb{E}(V^2) < \infty$, $\Gamma = \text{Cov}(\mathbf{W}, \mathbf{W})$, and $\beta, \alpha_1, \dots, \alpha_n$ are constants.

(i)

$$P(U|\mathbf{W}) = \mathbb{E}(U) = \mathbf{a}'(\mathbf{W} - \mathbb{E}(\mathbf{W})) \quad (1.34)$$

where $\Gamma \mathbf{a} = \text{Cov}(U, \mathbf{W})$.

(ii)

$$\mathbb{E}((U - P(U|\mathbf{W}))\mathbf{W}) = 0 \quad (1.35)$$

and

$$\mathbb{E}(U - P(U|\mathbf{W})) = 0 \quad (1.36)$$

(iii)

$$\mathbb{E}\left((U - P(U|\mathbf{W}))^2\right) = \mathbb{V}(U) - \mathbf{a}'\text{Cov}(U, \mathbf{W}) \quad (1.37)$$

(iv)

$$P\alpha_1 + \alpha_2 V + \beta|\mathbf{W} = \alpha_1 P(U|\mathbf{W}) + \alpha_2 P(V|\mathbf{W}) + \beta \quad (1.38)$$

(v)

$$P\left(\sum_{i=1}^n \alpha_i W_i + \beta|\mathbf{W}\right) = \sum_{i=1}^n \alpha_i W_i + \beta \quad (1.39)$$

(vi)

$$P(U|\mathbf{W}) = EU \quad (1.40)$$

if $\text{Cov}(U, \mathbf{W}) = 0$.

1.4.3 Innovation Algorithm

Theorem 1.24. Suppose X_t is a zero-mean series with $\mathbb{E}(|X_t|^2) < \infty$ for each t and $\mathbb{E}(X_i X_j) = \kappa(i, j)$. Let $\hat{X}_n = 0$ if $n = 1$, and $P_{n-1}X_n$ if $n = 2, 3, \dots$, and let $v_n = \mathbb{E}((X_{n+1} - P_n X_{n+1})^2)$.

Define the innovations, or one-step prediction errors, as $U_n = X_n - \hat{X}_n$.

Then we can write

$$\hat{X}_{n+1} = \begin{cases} 0 & n = 0 \\ \sum_{j=1}^n \theta_{nj}(X_{n+1-j} - \hat{X}_{n+1-j}) & \end{cases} \quad (1.41)$$

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

$$v_0 = \kappa(1, 1) \quad (1.42)$$

$$\theta_{n,n-k} = \frac{1}{v_k} (\kappa(n+1, k+1) - \sum_{j=0}^{k-1} \theta_{k,k-j} \theta_{n,n-j} v_j) \quad (1.43)$$

for $0 \leq k < n$, and

$$v_n = \kappa(n+1, n+1) - \sum_{j=0}^{n-1} \theta_{n,n-j}^2 v_j. \quad (1.44)$$

1.5 ARMA Processes

Definition 1.25. X_t is an ARMA(p, q) process if X_t is stationary and if for every t ,

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q} \quad (1.45)$$

where $Z_t \sim WN(0, \sigma^2)$ and the polynomials $(1 - \phi_1 z - \cdots - \phi_p z^p)$ and $(1 + \theta_1 z + \cdots + \theta_q z^q)$ have no common factors.

It can be more convenient to write this in the form

$$\phi(B)X_t = \theta(B)Z_t \quad (1.46)$$

with B the back-shift operator.

ARMA($0, q$) is a moving average process of order q (MA(q)).

ARMA($p, 0$) is an autoregressive process of order p (AR(p)).

Theorem 1.26. A stationary solution of (1.45) exists (and is the unique stationary solution) if and only if

$$\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p \neq 0 \quad (1.47)$$

for all $|z| = 1$

Definition 1.27. An ARMA(p, q) process X_t is causal (or a causal function of Z_t) if there exists constants ψ_j such that $\sum_{j=0}^{\infty} |\psi_j| < \infty$ and

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j} \quad (1.48)$$

for all t .

Theorem 1.28. An ARMA(p, q) process is causal if and only if

$$\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p \neq 0 \quad (1.49)$$

for all $|z| \leq 1$.

Note that the coefficients ψ_j are determined by

$$\psi_j - \sum_{k=1}^p \theta_k \psi_{j-k} = \theta_j \quad (1.50)$$

for $j = 0, 1, \dots$ and $\theta_0 = 1$, $\theta_j = 0$ for $j > q$, and $\psi_j = 0$ for $j < 0$.

Definition 1.29. An ARMA(p, q) is invertible if there exist constants π_j such that $\sum_{j=0}^{\infty} |\pi_j| < \infty$ and

$$Z_t = \sum_{j=0}^{\infty} \pi_j X_{t-j} \quad (1.51)$$

for all t .

The coefficients π_j are determined by the equations

$$\pi_j + \sum_{k=1}^q \theta_k \pi_{j-k} = -\phi_j \quad (1.52)$$

where $\phi_0 = -1$, $\theta_j = 0$ for $j > p$, and $\pi_j = 0$ for $j < 0$.

Theorem 1.30. Invertibility is equivalent to the condition

$$\theta(z) = 1 + \theta_1 z + \dots + \theta_q z^q \neq 0 \quad (1.53)$$

for all $|z| \leq 1$.

1.5.1 ACF and PACF of an ARMA(p, q) Process

Theorem 1.31. For a causal ARMA(p, q) process defined by

$$\phi(B)X_t = \theta(B)Z_t \quad (1.54)$$

we know we can write

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j} \quad (1.55)$$

where $\sum_{j=0}^{\infty} \psi_j z^j = \theta(z)/\phi(z)$ for $|z| \leq 1$.

Thus, the ACVF γ is given as

$$\gamma(h) = \mathbb{E}(X_{t+h}X_t) = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+|h|} \quad (1.56)$$

A second approach is to multiply each side by X_{t_k} and take expectations, and obtain a sequence of m homogenous linear difference equations with constant coefficients. These can be solved to obtain the $\gamma(h)$ values.

Definition 1.32 (PACF). The partial autocorrelation function (PACF)

of an AMRA process X is the function $\alpha(\cdot)$ defined by

$$\alpha(0) = 1 \quad (1.57)$$

$$\alpha(h) = \phi_{hh}, h \geq 1 \quad (1.58)$$

where ϕ_{hh} is the last component of $\mathbf{CE}_h = \Gamma_h^{-1}\gamma_h$, where $\Gamma_h = [\gamma(i-j)]_{i,j=1}^h$, and $\gamma_h = [\gamma(1), \gamma(2), \dots, \gamma(h)]$.

Theorem 1.33. *For an AR(p) process, the sample PACF values at lags greater than p are approximately independent $N(0, \frac{1}{n})$ random variables. Thus, if we have a sample PACF satisfying*

$$|\hat{\alpha}(h)| > \frac{1.96}{\sqrt{n}} \quad (1.59)$$

for $0 \leq h \leq p$ and

$$|\hat{\alpha}(h)| < \frac{1.96}{\sqrt{n}} \quad (1.60)$$

for $h > p$, this suggests an AR(p) model for the data.

Theorem 1.34 (PACF summary). *For an AR(p) process X_t , the PACF $\alpha(\cdot)$ has the properties that $\alpha(p) = \phi_p$, and $\alpha(h) = 0$ for $h > p$. For $h < p$ we can compute numerically from the expression that $\mathbf{CE}_h = \Gamma_h^{-1}\mathbf{f}_h$.*

1.5.2 Forecasting ARMA Processes

For the causal ARMA(p, q) process

$$\phi(B)X_t = \theta(B)Z_t, Z_t \sim WN(0, \sigma^2) \quad (1.61)$$

we can avoid using the full innovations algorithm.

If we apply the algorithm to the transformed process W_t given by

$$W_t = \begin{cases} \frac{1}{\sigma}X_t & t = 1, \dots, m \\ \frac{1}{\sigma}\phi(B)X_t & t > m \end{cases} \quad (1.62)$$

where $m = \max(p, q)$.

For notational convenience, take $\theta_0 = 1$, $\theta_j = 0$ for $j > q$.

Lemma 1.35. *The autocovariances $\kappa(i, j) = \mathbb{E}(W_i W_j)$ are found from*

$$\kappa(i, j) = \begin{cases} \sigma^2 \gamma_X(i-j) & 1 \leq i, j \leq m \\ \sigma^2 (\gamma_X(i-j) - \sum_{r=1}^p \phi_r \gamma_X(r - |i-j|)) & \min(i, j) \leq m < \max(i, j) \leq 2m \\ \sum_{r=0}^q \theta_r \theta_{r+|i-j|} & \min(i, j) > m \\ 0 & \text{otherwise} \end{cases} \quad (1.63)$$

Applying the innovations algorithm to the process W_t , we obtain

$$\hat{W}_{n+1} = \begin{cases} \sum_{j=1}^n \theta_{nj} (W_{n+1-j} - \hat{W}_{n+1-j}) & 1 \leq n < m \\ \sum_{j=1}^q \theta_{nj} (W_{n+1-j} - \hat{W}_{n+1-j}) & n \geq m \end{cases} \quad (1.64)$$

where the coefficients θ_{nj} and MSE $r_n = \mathbb{E}((W_{n+1} - \hat{W}_{n+1})^2)$ are found recursively using the innovations algorithm.

Since the equations (1.62) allow us to write X_n as a linear combination of $W_j, 1 \leq j \leq n$, and conversely, each $W_n, n \geq 1$ to be written as a linear combination of $X_j, 1 \leq j \leq n$. Thus the best linear predictor of the random variable Y in terms of $\{1, X_1, \dots, X_n\}$ is the same as the best linear predictor of Y in terms of $\{1, W_1, \dots, W_n\}$. Thus, by linearity of \hat{P}_n , we have

$$\hat{W}_t = \begin{cases} \frac{1}{\sigma} \hat{X}_t & t = 1, \dots, m \\ \frac{1}{\sigma} (\hat{X}_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p}) & t > m \end{cases} \quad (1.65)$$

which shows that

$$X_t - \hat{X}_t = \sigma(W_t - \hat{W}_t) \quad (1.66)$$

Substituting into (1.63) and (1.64), we obtain

$$\hat{X}_{n+1} = \begin{cases} \sum_{j=1}^n \theta_{nj} (X_{n+1-j} - \hat{X}_{n+1-j}) & 1 \leq n < m \\ \phi_1 X_n + \dots + \phi_p X_{n+1-p} + \sum_{j=1}^q \theta_{nj} (X_{n+1-j} - \hat{X}_{n+1-j}) & n \geq m \end{cases} \quad (1.67)$$

and

$$\mathbb{E}((X_{n+1} - \hat{X}_{n+1})^2) = \sigma^2 \mathbb{E}((W_{n+1} - \hat{W}_{n+1})^2) = \sigma^2 r_n \quad (1.68)$$

where θ_{nj} and r_n are found using the innovation algorithm.

1.6 Estimation of ARMA Processes

1.6.1 Yule-Walker Equations

Consider estimating a causal AR(p) process. We can write

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j} \quad (1.69)$$

where $\sum_{j=0}^{\infty} \psi_j z^j = \frac{1}{\phi(z)}$ for $z \leq 1$.

Multiplying each side by Z_{t-j} , and taking expectations, we obtain the Yule-Walker equations

$$\Gamma_p \mathbf{CE} = \mathbf{fl}_p \quad (1.70)$$

and $\sigma^2 = \gamma(0) - \langle \mathbf{CE}, \mathbf{fl}_p \rangle$ where $\Gamma_p = [\gamma(i-j)]_{i,j=1}^p$ and $\mathbf{fl}_p = (\gamma(1), \gamma(2), \dots, \gamma(p))$.

If we replace the covariances by the sample covariances $\hat{\gamma}(j)$, we obtain a set of equations for the so-called Yule-Walker estimators $\hat{\mathbf{CE}}$ and $\hat{\sigma}^2$, given by

$$\hat{\Gamma}_p \hat{\mathbf{CE}} = \hat{\mathbf{fl}}_p \quad (1.71)$$

and $\hat{\sigma}^2 = \hat{\gamma}(0) - \langle \hat{\mathbf{CE}}, \hat{\mathbf{fl}}_p \rangle$

Theorem 1.36. *If X_t is the causal AR(p) process and $\hat{\mathbf{CE}}$ is the Yule-Walker estimator of \mathbf{CE} , then*

$$n^{\frac{1}{2}}(\hat{\mathbf{CE}} - \mathbf{CE}) \xrightarrow{d} N(0, \sigma^2 \Gamma_p^{-1}) \quad (1.72)$$

Moreover, $\hat{\sigma}^2 \xrightarrow{p} \sigma^2$.

Theorem 1.37. *If X_t is a causal AR(p) process and $\hat{\mathbf{CE}}_m$ is the Yule-Walker estimate of order $m > p$, then*

$$n^{\frac{1}{2}}(\hat{\mathbf{CE}}_m - \mathbf{CE}_m) \xrightarrow{d} N(0, \sigma^2 \Gamma_m^{-1}) \quad (1.73)$$

where $\hat{\mathbf{CE}}_m$ is the coefficient vector of the best linear predictor $\langle \mathbf{CE}_m, \mathbf{X}_m \rangle$ of X_{m+1} based on X_m, \dots, X_1 . So $\mathbf{CE}_m = R_m^{-1} \mathbf{ae}_m$. In particular, for $m > p$,

$$n^{\frac{1}{2}} \hat{\phi}_{mm} \xrightarrow{d} N(0, 1) \quad (1.74)$$

Theorem 1.38 (Durbin-Levinson Algorithm for AR models). *Consider fitting an AR(m) process*

$$X_t - \hat{\theta}_{m1}X_{t-1} - \cdots - \hat{\theta}_{mm}X_{t-m} = Z_t \quad (1.75)$$

with $Z_t \sim \text{WN}(0, \hat{\nu}_m)$.

If $\hat{\gamma}(0) > 0$, then the fitted autoregressive models for $m = 1, 2, \dots, n-1$ can be determined recursively from the relations

$$\hat{\phi}_{11} = \hat{\rho}(1) \quad (1.76)$$

$$\hat{\nu}_1 = \hat{\gamma}(0)(1 - \hat{\rho}^2)(1) \quad (1.77)$$

$$\hat{\phi}_{mm} = \frac{\hat{\gamma}(m) - \sum_{j=1}^{m-1} \hat{\phi}_{m-1,j} \hat{\gamma}(m-j)}{\hat{\nu}_{m-1}} \quad (1.78)$$

$$\begin{Bmatrix} \hat{\phi}_{m1} \\ \vdots \\ \hat{\phi}_{m,m-1} \end{Bmatrix} = \hat{\mathbf{C}}_{m-1} - \hat{\phi}_{mm} \begin{Bmatrix} \hat{\phi}_{m-1,m-1} \\ \vdots \\ \hat{\phi}_{m-1,1} \end{Bmatrix} \quad (1.79)$$

$$\hat{\nu}_m = \hat{\nu}_{m-1}(1 - \hat{\phi}_{mm}^2) \quad (1.80)$$

Theorem 1.39 (Confidence intervals for AR(p) estimation). *Under the assumption that the order p of the fitted model is the correct value, for large sample-size n , the region*

$$\{\mathbf{C} \in \mathbb{R}^p \mid (\mathbf{C} - \hat{\phi}_p)' \hat{\Gamma}_p (\mathbf{C} - \hat{\mathbf{C}}_p) \leq \frac{1}{n} \hat{\nu}_p \chi_{1-\alpha}^2(p)\} \quad (1.81)$$

contains \mathbf{C}_p with probability close to $1 - \alpha$ where $\chi_{1-\alpha}^2(p)$ is the $(1 - \alpha)$ quantile of the chi-squared distribution with p degrees of freedom.

Similarly, if $\Phi_{1-\alpha}$ is the $(1 - \alpha)$ quantile of the standard normal distribution and $\hat{\nu}_{jj}$ is the j -th diagonal element of $\hat{\nu}_p \hat{\Gamma}_p^{-1}$, then for large n

$$\{\hat{\phi}_{pj} \pm \Phi_{1-\frac{\alpha}{2}} \frac{1}{n^{\frac{1}{2}}} \hat{\nu}_{jj}^{\frac{1}{2}}\} \quad (1.82)$$

contains ϕ_{pj} with probability close to $(1 - \alpha)$.

1.6.2 Estimation for Moving Average Processes Using the Innovations Algorithm

Consider estimating

$$X_t = Z_t + \hat{\theta}_{m1}Z_{t-1} + \cdots + \hat{\theta}_{mm}Z_{t-m} \quad (1.83)$$

with $Z_t \sim WN(0, \hat{\nu}_m)$.

Theorem 1.40. *We can apply the innovation estimates by applying the recursive relations*

$$\hat{\nu}_0 = \hat{\gamma}(0) \quad (1.84)$$

$$\hat{\theta}_{m,m-k} = \frac{1}{\hat{\nu}_k} (\hat{\gamma}(m-k) - \sum_{j=0}^{k-1} \hat{\theta}_{m,m-j} \hat{\theta}_{k,k-j} \hat{\nu}_j) \quad (1.85)$$

for $k = 0, \dots, m-1$, and

$$\hat{\nu}_m = \hat{\gamma}(0) - \sum_{j=0}^{m-1} \hat{\theta}_{m,m-j}^2 \hat{\nu}_j. \quad (1.86)$$

Theorem 1.41. *Let X_t be the causal invertible ARMA process $\phi(B)X_t = \theta(B)Z_t$ with $Z_t \sim WN(0, \sigma^2)$, $\mathbb{E}(Z_t^4) < \infty$, and let $\psi(z) = \sum_{j=0}^{\infty} \psi_j z^j = \frac{\theta(z)}{\phi(z)}$ for $|z| \leq 1$, and $\psi_0 = 1$ and $\psi_j = 0$ for $j < 0$.*

Then for any sequence of positive integers m_n , such that $m < n$, $m \rightarrow \infty$, and $m = o(n^{\frac{1}{3}})$ as $n \rightarrow \infty$, we have for each k ,

$$\frac{n^{\frac{1}{2}}}{\left(\hat{\theta}_{m1} - \psi_1, \dots, \hat{\theta}_{mk} - \psi_k \right)} \xrightarrow{d} N(0, A) \quad (1.87)$$

where $A = [a_{ij}]_{i,j=1}^k$ and

$$a_{ij} = \sum_{r=1}^{\min(i,j)} \psi_{i-r} \psi_{j-r} \quad (1.88)$$

and

$$\hat{\nu}_m \xrightarrow{p} \sigma^2. \quad (1.89)$$

Remark 1.42. *Note that for the AR(p) process, the Yule-Walker estimator is a consistent estimator of \mathbf{CE}_p . However, for an MA(q) process, the estimator $\hat{\varsigma}_q$ is not consistent for the true parameter vector as $n \rightarrow \infty$. For consistency, it is necessary to use the estimators with m satisfying the conditions given*

in Theorem 1.41.

Theorem 1.43 (Asymptotic confidence regions for the θ).

$$\{\theta \in R \mid |\theta - \hat{\theta}_{mj}| \leq \Phi_{1-\frac{\alpha}{2}} \frac{1}{n^{\frac{1}{2}}} \left(\sum_{k=0}^{j-1} \hat{\theta}_{mk}^2 \right)^{\frac{1}{2}}\} \quad (1.90)$$

is an $(1 - \alpha)$ confidence interval for θ_{mj} .

1.6.3 Maximum Likelihood Estimation

Consider X_t a gaussian time series with zero mean and autocovariance function $\kappa(i, j) = \mathbb{E}(X_i X_j)$. Let $\hat{X}_j = P_{j-1} X_j$. Let Γ_n be the covariance matrix and assume Γ_n is nonsingular. The likelihood of X_n is

$$L(\Gamma_n) = \frac{1}{(2\pi)^{\frac{n}{2}}} \frac{1}{(\det \Gamma_n)^{\frac{1}{2}}} \exp\left(-\frac{1}{2} \mathbf{X}'_n \Gamma_n^{-1} \mathbf{X}_n\right) \quad (1.91)$$

Theorem 1.44. The likelihood of the vector \mathbf{X}_n reduces to

$$L(\Gamma_n) = \frac{1}{\sqrt{(2\pi)^n \prod_{i=0}^{n-1} r_i}} \exp\left(-\frac{1}{2} \sum_{j=1}^n \frac{(X_j - \hat{X}_j)^2}{r_{j-1}}\right) \quad (1.92)$$

Remark 1.45. Even if X_t is not Gaussian, the large sample estimates are the same for $Z_t \sim \text{IID}(0, \sigma^2)$, regardless of whether or not Z_t is Gaussian.

Theorem 1.46 (Maximum Likelihood Estimators for ARMA processes).

$$\hat{\sigma}^2 = \frac{1}{n} S(\hat{\mathbf{C}}, \hat{\cdot}) \quad (1.93)$$

where $\hat{\mathbf{C}}, \hat{\cdot}$ are the values of \mathbf{C}, \cdot that minimize

$$\ell(\mathbf{C}, \cdot) = \ln\left(\frac{1}{n} S(\cdot, \cdot)\right) + \frac{1}{n} \sum_{j=0}^{n-1} \ln r_j \quad (1.94)$$

and

$$S(\hat{\mathbf{C}}, \hat{\cdot}) = \sum_{j=1}^n \frac{(X_j - \hat{X}_j)^2}{r_{j-1}} \quad (1.95)$$

Theorem 1.47 (Asymptotic Distribution of Maximum Likelihood Esti-

mators). For a large sample from an ARMA(p, q) process,

$$\hat{\mathbf{f}} = N(\mathbf{f}, \frac{1}{n} V \mathbf{f}) \quad (1.96)$$

where

$$V(\mathbf{f}) = \sigma^2 \begin{bmatrix} \mathbb{E}(U_t U_t') & \mathbb{E}(U_t V_t') \\ \mathbb{E}(V_t U_t') & \mathbb{E}(V_t V_t') \end{bmatrix}^{-1} \quad (1.97)$$

and U_t are the autoregressive process $\phi(B)U_t = Z_t$ and $\theta(B)V_t = Z_t$.

Note that for $p = 0$, $V(\mathbf{f}) = \sigma^2 [\mathbb{E}(V_t V_t')]^{-1}$, and for $q = 0$, $V(\mathbf{f}) = \sigma^2 [\mathbb{E}(U_t U_t')]^{-1}$.

1.6.4 Order Selection

Definition 1.48 (Kullback-Leibler divergence). The Kullback-Leibler (KL) divergence between $f(\cdot; \psi)$ and $f(\cdot; \theta)$ is defined as

$$d(\psi|\theta) = \Delta(\psi|\theta) - \Delta(\theta|\theta) \quad (1.98)$$

where

$$\Delta(\psi|\theta) = \mathbb{E}_\theta(-2 \ln f(X; \psi)) \quad (1.99)$$

is the Kullback-Leibler index of $f(\cdot; \psi)$ relative to $f(\cdot; \theta)$.

Theorem 1.49 (AICC of ARMA(p, q) process).

$$AICC(\mathbf{f}) = -2 \ln L_X(\mathbf{f}, \frac{S_X(\beta)}{n}) + \frac{2(p+q+1)n}{n-p-q-2} \quad (1.100)$$

Theorem 1.50 (AIC of ARMA(p, q) process).

$$AIC(\mathbf{f}) = -2 \ln L_X(\mathbf{f}, \frac{S_X(\beta)}{n}) + 2(p+q+1) \quad (1.101)$$

Theorem 1.51 (BIC of ARMA(p, q) process).

$$BIC(\mathbf{f}) = (n-p-q) \ln \frac{n\hat{\sigma}^2}{n-p-q} + n(1 + \ln \sqrt{2\pi}) + (p+q) \ln \frac{\sum_{t=1}^n X_t^2 - n\hat{\sigma}^2}{p+q} \quad (1.102)$$

where $\hat{\sigma}^2$ is the MLE estimate of the white noise variance.

1.7 Spectral Analysis

Let X_t be a zero-mean stationary time series with autocovariance function $\gamma(\cdot)$ satisfying $\sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty$.

Definition 1.52. The spectral density of X_t is the function $f(\cdot)$ defined by

$$f(\lambda) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} e^{-ih\lambda} \gamma(h) \quad (1.103)$$

The summability implies that the series converges absolutely.

Theorem 1.53. (i) f is even

(ii) $f(\lambda) \geq 0$ for all $\lambda \in (-\pi, \pi]$.

(iii) $\gamma(k) = \int_{-\pi}^{\pi} e^{-ik\lambda} f(\lambda) d\lambda = \int_{-\pi}^{\pi} \cos(k\lambda) f(\lambda) d\lambda$.

Definition 1.54. A function f is the **spectral density** of a stationary time series X_t with autocovariance function $\gamma(\cdot)$ if

(i) $f(\lambda) \geq 0$ for all $\lambda \in (0, \pi]$,

(ii) $\gamma(h) = \int_{-\pi}^{\pi} e^{ih\lambda} f(\lambda) d\lambda$ for all integers h .

Lemma 1.55. If f and g are two spectral density corresponding to the autocovariance function γ , then f and g have the same Fourier coefficients and hence are equal.

Theorem 1.56. A real-valued function f defined on $(-\pi, \pi]$ is the spectral density of a stationary process if and only if

(i) $f(\lambda) = f(-\lambda)$,

(ii) $f(\lambda) \geq 0$

(iii) $\int_{-\pi}^{\pi} f(\lambda) d\lambda < \infty$.

Theorem 1.57. An absolutely summable function $\gamma(\cdot)$ is the autocovariance function of a stationary time series if and only if it is even and

$$f(\lambda) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} e^{-ih\lambda} \gamma(h) \geq 0 \quad (1.104)$$

for all $\lambda \in (-\pi, \pi]$, in which case $f(\cdot)$ is the spectral density of $\gamma(\cdot)$.

Theorem 1.58 (Spectral Representation of the ACVF). *A function $\gamma(\cdot)$ defined on the integers is the ACVF of a stationary time series if and only if there exists a right-continuous, nondecreasing, bounded function F on $[-\pi, \pi]$ with $F(-\pi) = 0$ such that*

$$\gamma(h) = \int_{-\pi}^{\pi} e^{ih\lambda} dF(\lambda) \quad (1.105)$$

for all integers h .

Remark 1.59. *The function F is a **generalized distribution function** on $[-\pi, \pi]$ in the sense that $G(\lambda) = \frac{F(\lambda)}{F(\pi)}$ is a probability distribution function on $[-\pi, \pi]$. Note that since $F(\pi) = \gamma(0) = \mathbb{V}(X_1)$, the ACF of X_t has the spectral representation function*

$$\rho(h) = \int_{-\pi}^{\pi} e^{ih\lambda} dG(\lambda) \quad (1.106)$$

The function F is called the *spectral distribution function* of $\gamma(\cdot)$. If $F(\lambda)$ can be expressed as $F(\lambda) = \int_{-\pi}^{\lambda} f(y)dy$ for all $\lambda \in [-\pi, \pi]$, then f is the *spectral density function* and the time series is said to have a *continuous spectrum*. If F is a discrete function, then the time series is said to have a *discrete spectrum*.

Theorem 1.60. *A complex valued function $\gamma(\cdot)$ is the autocovariance function of a stationary process X_t if and only if either*

- (i) $\gamma(h) = \int_{-\pi}^{\pi} e^{-ihv} dF(v)$ for all $h = 0, \pm 1, \dots$ where F is a right-continuous, non-decreasing, bounded function on $[-\pi, \pi]$ with $F(-\pi) = 0$, or
- (ii) $\sum_{i,j=1}^n a_i \gamma(i-j) \bar{a}_j \geq 0$ for all positive integers n and all $a = (a_1, \dots, a_n \in \mathbb{C}^n)$.

1.7.1 The Spectral Density of an ARMA Process

Theorem 1.61. *If Y_t is any zero-mean, possibly complex-valued stationary process with spectral distribution function $F_Y(\cdot)$ and X_t is the process*

$X_t = \sum_{j=-\infty}^{\infty} \psi_j Y_{t-j}$ where $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$, then X_t is stationary with spectral distribution function $F_X(\lambda) = \int_{-\pi, \lambda}^{\infty} |\sum_{j=-\infty}^{\infty} \psi_j e^{-ijv}|^2 dF_Y(v)$ for $-\pi \leq \lambda \leq \pi$.

If Y_t has a spectral density $f_Y(\cdot)$, then X_t has a spectral density $f_X(\cdot)$ given by $f_X(\lambda) = |\Psi(e^{-i\lambda})|^2 f_Y(\lambda)$ where $\Psi(e^{-i\lambda}) = \sum_{j=-\infty}^{\infty} \psi_j e^{-ij\lambda}$.

Theorem 1.62. Let X_t be an ARMA(p, q) process, not necessarily causal or invertible satisfying $\phi(B)X_t = \theta(B)Z_t$, $Z_t \sim \text{WN}(0, \sigma^2)$ where $\phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p$ and $\theta(z) = 1 + \theta_1 z + \dots + \theta_q z^q$ have no common zeroes and $\phi(z)$ has no zeroes on the unit circle. Then X_t has spectral density

$$f_X(\lambda) = \frac{\sigma^2}{2\pi} \frac{|\theta(e^{-i\lambda})|^2}{|\phi(e^{-i\lambda})|^2} \quad (1.107)$$

for $-\pi \leq \lambda \leq \pi$.

Theorem 1.63. The spectral density of the white noise process is constant, $f(\lambda) = \frac{\sigma^2}{2\pi}$.

1.7.2 The Periodogram

Definition 1.64. The periodogram of (x_1, \dots, x_n) is the function

$$I_n(\lambda) = \frac{1}{n} \left| \sum_{t=1}^n x_t e^{-it\lambda} \right|^2 \quad (1.108)$$

Theorem 1.65. If x_1, \dots, x_n are any real numbers and ω_k is any of the nonzero Fourier Frequencies $\frac{2\pi k}{n}$ in $(-\pi, \pi]$, then $I_n(\omega_k) = \sum_{|h| < n} \hat{\gamma}(h) e^{-ih\omega_k}$ where $\hat{\gamma}(h)$ is the sample ACVF of x_1, \dots, x_n .

Theorem 1.66. Let X_t be the linear process $X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}$, $Z_t \sim \text{IID}(0, \sigma^2)$, with $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$. Let $I_n(\lambda)$ be the periodogram of X_1, \dots, X_n , and let $f(\lambda)$ be the spectral density of X_t .

(i) If $f(\lambda) > 0$ for all $\lambda \in [-\pi, \pi]$ and if $0 < \lambda_1 < \dots < \lambda_m < \pi$, then the random vector $(I_n(\lambda_1), \dots, I_n(\lambda_m))$ converges in distribution to a vector of independent and exponentially distributed random variables, the i -th component which has mean $2\pi f(\lambda_i)$, $i = 1, \dots, m$.

(ii) If $\sum_{j=-\infty}^{\infty} |\psi_j| |j|^{\frac{1}{2}} < \infty$, $\mathbb{E}(Z_1^4) = v\sigma^4 < \infty$, $\omega_j = \frac{2\pi j}{n} \geq 0$, and

$\omega_k = \frac{2\pi k}{n} \geq 0$, then

$$\text{Cov}(I_n(\omega_j), I_n(\omega_k), -NoValue-) = \begin{cases} 2(2\pi)^2 f^2(\omega_j) + O(n^{-\frac{1}{2}}) & \omega_j = \omega_k = \{0, \pi\} \\ (2\pi)^2 f^2(\omega_j) + O(n^{-\frac{1}{2}}) & 0 < \omega_j = \omega_k < \pi \\ O(n^{-1}) & \omega_j \neq \omega_k \end{cases} \quad (1.109)$$

Definition 1.67. The estimator $\hat{f}(\omega) = \hat{f}(g(n, \omega))$ with $\hat{f}(\omega_j)$ defined by $\frac{1}{2\pi} \sum_{|k| \leq m_n} W_n(k) I_n(\omega_{j+k})$ with $m \rightarrow \infty$, $\frac{m}{n} \rightarrow 0$, $W_n(k) = W_n(-k)$, $W_n(k) \geq 0$ for all k , and $\sum_{|k| \leq m} W_n(k) = 1$, and $\sum_{|k|} W_n^2(k) \rightarrow 0$ as $n \rightarrow \infty$ is called a **discrete spectral average estimator** of $f(\omega)$.

Theorem 1.68. Let X_t be the linear process $X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}$, $Z_t \sim IID(0, \sigma^2)$, with $\sum_{j=-\infty}^{\infty} |\psi_j| |j|^{\frac{1}{2}} < \infty$ and $\mathbb{E}(Z_1^4) < \infty$. If \hat{f} is a discrete spectral average estimator of the spectral density f , then for $\lambda, \omega \in [0, \pi]$,

(i) $\lim_{n \rightarrow \infty} \mathbb{E}(\hat{f}(\omega)) = f(\omega)$

(ii)

$$\lim_{n \rightarrow \infty} \frac{1}{\sum_{|j| \leq m} W_n^2(j)} \text{Cov}(\hat{f}(\omega), \hat{f}(\lambda), -NoValue-) = \begin{cases} 2f^2(\omega) & \omega = \lambda = \{0, \pi\} \\ f^2(\omega) & 0 < \omega = \lambda < \pi \\ 0 & \omega \neq \lambda. \end{cases} \quad (1.110)$$

Bibliography

Peter J Brockwell and Richard A Davis. *Introduction to time series and forecasting*. Taylor & Francis US, 2002.

Peter J Brockwell and Richard A Davis. *Time series: theory and methods*. Springer, 2009.