NOTES ON ARMA PROCESSES FOR SF2943 TIME SERIES ANALYSIS

FILIP LINDSKOG

Abstract. Here are brief notes that I wrote, by borrowing material from [1], [2], and [3], to emphasize some theoretical aspects of time series analysis. Most likely, the document will grow as the course progresses. No serious proof reading has been done so there are probably a few errors somewhere.

1. Preliminaries on power series

Let C denote the complex numbers and consider functions $f : C \to C$. For such functions the derivative is define similarly as for real-valued functions defined on the real numbers:

$$f'(z) = \lim_{\Delta z \to 0} \frac{f(z + \Delta z) - f(z)}{\Delta z}$$

provided that the limit exists. The function f is analytic at z_0 if $f'(z)$ exists at every z in some neighborhood of z_0. Similarly, f is analytic in $D \subset C$ if f is analytic at every point in D. It is shown in Theorem 5 in [3] that sums, differences, and products of functions that are analytic in D are analytic. Moreover, the quotient of two functions analytic in D is analytic in D except where the denominator equals zero. The polynomials $\phi(z) = 1-\phi_1 z-\cdots-\phi_p z^p$ and $\theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q$ are analytic everywhere. The rational function $\theta(z)/\phi(z)$ is analytic in a domain D if $\phi(z) \neq 0$ for all z in D.

If f is analytic everywhere inside a circle of radius $a > 0$ centered at z_0, then there exists a power series which converges to $f(z)$ for each z inside this circle; that is

$$f(z) = \sum_{k=0}^{\infty} c_k (z - z_0)^k, \quad |z - z_0| < a,$$

where $c_k = f^{(k)}(z_0)/k!$. If $z_0 = 0$, then the series is called a Maclaurin series.

2. Linear processes

We begin with some definitions.

White noise. The process $\{Z_t\}$ is white noise with mean 0 and variance σ^2, written $\mathrm{WN}(0, \sigma^2)$, if each term has mean 0 and variance σ^2, and the terms are uncorrelated.

Stationarity. The process $\{X_t\}$ is stationary if $E[X_t]$ does not depend on t and $\mathrm{Cov}(X_s, X_t)$ depends on s and t only through $|t - s|$.

Date: 2014-05-07.

Moving average process. The process $\{X_t\}$ is a moving-average process of order q, written MA(q), if $X_t = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q}$, where $\{Z_t\}$ is WN$(0, \sigma^2)$ and $\theta_q \neq 0$.

It is easily verified that moving average processes are stationary. Clearly, $E[X_t] = 0$. Moreover, for $0 \leq h \leq q$ and with $\theta_0 = 1$,

$$\begin{aligned}
\operatorname{Cov}(X_{t+h}, X_t) &= E\Big[\sum_{j=0}^{q} \theta_j Z_{t+h-j} \sum_{k=0}^{q} \theta_k Z_{t-k}\Big] \\
&= E\Big[\sum_{j=h}^{q} \theta_j Z_{t+h-j} \sum_{k=0}^{q} \theta_k Z_{t-k}\Big] + E\Big[\sum_{j=0}^{h-1} \theta_j Z_{t+h-j} \sum_{k=0}^{q} \theta_k Z_{t-k}\Big] \\
&= \big\{ \text{ set } i = j - h \ \big\} \\
&= E\Big[\sum_{i=0}^{q-h} \theta_{i+h} Z_{t-i} \sum_{k=0}^{q} \theta_k Z_{t-k}\Big] \\
&= \sigma^2 \sum_{i=0}^{q-h} \theta_{i+h} \theta_i
\end{aligned}$$

verifying that the MA(q) process is stationary.

Linear process. The process $\{X_t\}$ is a linear process if $X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}$, where $\{Z_t\}$ is WN$(0, \sigma^2)$, and $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$.

We need to verify that the condition on the coefficients ensures that the sum is convergent, i.e. that $|\sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}| < \infty$ with probability one. For a nonnegative random variable A,

$$\begin{aligned}
E[A] &= \int_0^\infty b f_A(b) db \\
&= \int_0^\infty f_A(b) \Big(\int_0^b da \Big) db \\
&= \int_0^\infty \Big(\int_a^\infty f_A(b) db \Big) da \\
&= \int_0^\infty P(A > a) da
\end{aligned}$$

(assuming that A has a density f_A - which is not really necessary to assume) so $E[A] < \infty$ implies that $P(A > a) \to 0$ as $a \to \infty$ which implies that $P(A < \infty) = 1$.

It remains to verify that $E[|\sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}|] < \infty$. Clearly,

$$\Big| \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j} \Big| \leq \sum_{j=-\infty}^{\infty} |\psi_j| |Z_{t-j}|.$$

Since the terms $|\psi_j||Z_{t-j}|$ are nonnegative, by Fubini's (Tonelli's) theorem we may interchange the order of expectation and summation to get

$$E\Big[\sum_{j=-\infty}^{\infty}|\psi_j||Z_{t-j}|\Big] = \sum_{j=-\infty}^{\infty}|\psi_j|E[|Z_{t-j}|].$$

Finally, $E[|Z_{t-j}|^2] \geq E[|Z_{t-j}|]^2$, so putting the pieces together yields

$$E\Big[\Big|\sum_{j=-\infty}^{\infty}\psi_j Z_{t-j}\Big|\Big] \leq E\Big[\sum_{j=-\infty}^{\infty}|\psi_j||Z_{t-j}|\Big] \leq \sigma\sum_{j=-\infty}^{\infty}|\psi_j| < \infty$$

which verifies that the random sum defining the linear process converges. Computations similar to those for the MA(q) process shows that linear processes are stationary.

3. ARMA processes: causality and invertibility

Here we present the ARMA process and some of its basic properties.

Autoregressive process. The process $\{X_t\}$ is an autoregressive process of order p, written AR(p), if it is stationary and $X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = Z_t$, where $\{Z_t\}$ is WN$(0, \sigma^2)$ and $\phi_p \neq 0$.

Autoregressive processes are less straightforward than moving average processes. For instance, the AR(1) process, with $|\phi_1| < 1$,

$$X_t = \phi_1 X_{t-1} + Z_t = \phi_1(\phi_1 X_{t-2} + Z_{t-1}) + Z_t = \cdots = \sum_{k=0}^{\infty}\phi_1^k Z_{t-k}$$

can be expressed as linear process (a moving average process of infinite order). On the other hand, with $|\phi_1| > 1$,

$$X_t = \phi_1^{-1}X_{t+1} - \phi_1^{-1}Z_{t+1} = \phi_1^{-1}(\phi_1^{-1}X_{t+2} - \phi_1^{-1}Z_{t+2}) - \phi_1^{-1}Z_{t+1} = \cdots = -\sum_{k=1}^{\infty}\phi_1^{-k}Z_{t+k}$$

which is also a linear process, but with the unpleasant property that the output of the system (X_t) depends on future input to the system (Z_s for $s > t$).

We now introduce the autoregressive moving average, ARMA, process and investigate under what conditions on its parameters it can be represented as a linear process that only depends on past and current input noise to the system, a property called causality.

ARMA(p,q). The process $\{X_t\}$ is an ARMA(p,q) process if it is stationary and if for every t,

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q}, \tag{1}$$

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

Now we turn to the question whether, and under what conditions, (1) has a solution of the form

$$X_t = \sum_{k=0}^{\infty} \psi_k Z_{t-k}, \quad \text{where } \sum_{k=0}^{\infty} |\psi_k| < \infty. \tag{2}$$

In that case the ARMA(p,q) process $\{X_t\}$ is causal, or more precisely, $\{X_t\}$ is obtained as the output of a causal linear filter with $\{Z_t\}$ as the input.

We will show that if the autoregressive polynomial ϕ has no roots in the unit disc, then the ARMA process is causal.

Proposition 1. *If $\phi(z) \neq 0$ for $|z| \leq 1$, then the ARMA process has the causal representation* (2).

Proof. (Theorem 3.1.1 in [1].) Let B be the backward shift operator: $B^j X_t = X_{t-j}$ and $B^j Z_t = Z_{t-j}$. We may rewrite (1) more concisely as

$$\phi(B) X_t = \theta(B) Z_t. \tag{3}$$

Formal manipulation gives $X_t = \phi(B)^{-1}\theta(B) Z_t$. If the polynomial ϕ satisfies $\phi(z) \neq 0$ for $|z| \leq 1$, then there is some $\delta > 0$ such that $\phi(z) \neq 0$ for $|z| < 1 + \delta$, which ensures that $\psi(z) = \phi(z)^{-1}\theta(z)$ is analytic everywhere inside a circle of radius $1 + \delta$ centered at 0. In particular, for $|z| < 1 + \delta$, $\psi(z)$ can be represented as convergent the power series

$$\psi(z) = \sum_{k=0}^{\infty} \psi_k z^k.$$

Hence, $\psi_k (1 + \delta/2)^k \to 0$ as $k \to \infty$, which implies that $|\psi_k| < K(1 + \delta/2)^{-k}$ for some $K > 0$ and every nonnegative integer k. It follows that $\sum_{k=0}^{\infty} |\psi_k| < \infty$. Hence,

$$X_t = \phi(B)^{-1}\theta(B) Z_t = \sum_{k=0}^{\infty} \psi_k Z_{t-k}, \quad \text{where } \sum_{k=0}^{\infty} |\psi_k| < \infty.$$

We know that the condition $\sum_{k=0}^{\infty} |\psi_k| < \infty$ ensures that the series representation of X_t converges absolutely with probability one. Moreover, the condition on the absolute summability of the coefficients also implies that the linear process is stationary. □

Causality implies stationarity: for a processes $\{X_t\}$ satisfying (2) and it holds that $E[X_t] = 0$ and

$$\mathrm{Cov}(X_{t+h}, X_t) = E[X_{t+h} X_t] = E\Big[\sum_{j,k=0}^{\infty} \psi_j \psi_k Z_{t+h-j} Z_{t-k}\Big] = \sigma^2 \sum_{k=0}^{\infty} \psi_k \psi_{k+h}.$$

Since

$$\Big|\sum_{k=0}^{\infty} \psi_k \psi_{k+h}\Big| \leq \sum_{k=0}^{\infty} |\psi_k||\psi_{k+h}| \leq \sum_{j=0}^{\infty} |\psi_j| \sum_{k=0}^{\infty} |\psi_k| < \infty$$

$E[X_{t+h} X_t]$ exists finitely.

Invertibility means that the system input noise variable Z_t can be expressed as a linear function of past system output values X_s for $s \leq t$. From the ARMA representation $\phi(B)X_t = \theta(B)Z_t$ and the study of causality above, it is clear (or at least very plausible) that the ARMA process is invertible if $\theta(z) \neq 0$ for $|z| \leq 1$.

4. Yule-Walker estimation

If $\{X_t\}$ is stationary, then $\text{Cov}(X_{t+h}, X_t)$ does not depend on t and we define the autocovariance function (ACVF) of $\{X_t\}$ at lag h as $\gamma(h) = \text{Cov}(X_{t+h}, X_t)$. Similarly, we define the autocorrelation function (ACF) of $\{X_t\}$ at lag h as $\gamma(h) = \text{Cor}(X_{t+h}, X_t) = \gamma(h)/\gamma(0)$.

Consider a causal AR(p) process

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + Z_t, \quad \{Z_t\} \sim \text{WN}(0, \sigma^2).$$

Multiplying each side of the above expression by X_{t-k} and taking expectations yield

$$\begin{aligned}
\gamma(k) &= \text{Cov}(X_t, X_{t-k}) \\
&= \phi_1 \text{Cov}(X_{t-1}, X_{t-k}) + \cdots + \phi_p \text{Cov}(X_{t-p}, X_{t-k}) + \text{Cov}(Z_t, X_{t-k}) \\
&= \phi_1 \gamma(k-1) + \cdots + \phi_p \gamma(k-p).
\end{aligned}$$

Here causality is used to ensure that $\text{Cov}(Z_t, X_{t-k}) = 0$. For $k = 1, \ldots, p$ this operation results in the matrix equation

$$\Gamma_p \phi = \gamma_p, \quad \text{where} \quad \Gamma_p = [\gamma(i-j)]_{i,j=1}^p, \quad \phi = (\phi_1, \ldots, \phi_p)', \quad \gamma_p = (\gamma(1), \ldots, \gamma(p))'.$$

Similarly, multiplying each side of the expression

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

by X_t and taking expectations yield $\sigma^2 = \gamma(0) - \phi' \gamma_p$. The covariance matrix Γ_p is invertible for any sensible choice of AR(p) process (see Prop. 5.1.1 in [1] for details), which yields

$$\phi = \Gamma_p^{-1} \gamma_p \quad \text{and} \quad \sigma^2 = \gamma(0) - \gamma_p \Gamma_p^{-1} \gamma_p. \tag{4}$$

Notice that (4) can be expressed in terms of autocorrelations instead of autocovariances as

$$\phi = R_p^{-1} \rho_p \quad \text{and} \quad \sigma^2 = \gamma(0)(1 - \rho_p R_p^{-1} \rho_p), \tag{5}$$

where $\rho_p = \gamma(0)^{-1} \gamma_p$ and $R_p = \gamma(0)^{-1} \Gamma_p$. The usefulness of expressing the coefficient of the AR(p) process in terms of autocovariances (or autocorrelations) as in (4) (or (5)) comes from the fact that the autocovariances (and autocorrelations) can be approximated by their sample analogs

$$\widehat{\gamma}(h) = \frac{1}{n} \sum_{t=0}^{n-|h|} (X_{t+|h|} - \overline{X}_n)(X_t - \overline{X}_n) \quad \text{and} \quad \widehat{\rho}(h) = \frac{\widehat{\gamma}(h)}{\widehat{\gamma}(0)} \tag{6}$$

which results in the Yule-Walker estimates. It is shown in Section 2.4.2 in [2] that the sample autocovariance and autocorrelation matrices $\widehat{\Gamma}_p$ and $\widehat{R}_p$ are in fact invertible if $\widehat{\gamma}(0) > 0$. Yule-Walker estimation of the parameters of a AR(p) process is relatively straightforward if we know the number p. In practice, we do not. However, there are solutions to this problem.

5. Forecasting

Consider a stationary time series with mean μ and autocovariance function $\gamma(h)$ and the following situation: given observations of $X_1, \dots, X_n$ we want to form the best linear predictor $P_n X_{n+h} = a_0 + a_1 X_n + \cdots + a_n X_1$ of the future process value X_{n+h}, $h > 0$. What should be meant by "best"? Here we take "best" to mean that the mean squared error

$$E\big[(X_{n+h} - P_n X_{n+h})^2\big] \tag{7}$$

is minimized. Fortunately, minimizing (7) is relatively straightforward since (7) is a differentiable convex function of the vector of coefficients $(a_0, a_1, \dots, a_n)$. The coefficients minimizing the mean squared error are found by computing the partial derivatives of (7) with respect to the a_ks, setting the partial derivatives to zero, and finally solving the obtained linear equation system. With $X = (X_n, X_{n-1}, \dots, X_1)'$, $\Sigma_{X,X_{n+h}} = (\mathrm{Cov}(X_n, X_{n+h}), \dots, \mathrm{Cov}(X_1, X_{n+h}))'$ and Σ_X and μ_X denoting the covariance matrix and mean vector of X, respectively, the unique solution takes the form

$$a = \Sigma_X^{-1} \Sigma_{X,X_{n+h}} \quad \text{and} \quad a_0 = \mu - a' \mu_X. \tag{8}$$

Notice that minimizing the mean squared error is equivalent to solving a standard linear regression problem: regressing X_{n+h} onto the regressors $X_n, X_{n-1}, \dots, X_1$. The solution is identified as the solution to the so-called normal equations. Moreover, from (8) it follows that

$$\mathrm{Cov}(X_{n+h} - P_n X_{n+h}, X_k) = 0 \quad \text{for } k = 1, \dots, n. \tag{9}$$

This statement can be verified as follows. Clearly, $\Sigma'_{X,X_{n+h}} = \Sigma'_{X,X_{n+h}} \Sigma_X^{-1} \Sigma_X = 0$. In particular, the kth component of the row vector on the left-hand side equals zero:

$$(\Sigma'_{X,X_{n+h}})_k - \Sigma'_{X,X_{n+h}} \Sigma_X^{-1} \Sigma_{X,X_k} = 0,$$

which is a way of writing $\mathrm{Cov}(X_{n+h}, X_k) - \mathrm{Cov}((\Sigma_X^{-1}\Sigma_{X,X_{n+h}})'X, X_k) = 0$ or equivalently $\mathrm{Cov}(X_{n+h} - P_n X_{n+h}, X_k) = 0$.

Since $E[P_n X_{n+h}] = E[X_{n+h}]$ the minimal mean squared error is

$$\begin{aligned}
E\big[(X_{n+h} - P_n X_{n+h})^2\big] &= \mathrm{Var}(X_{n+h} - P_n X_{n+h}) \\
&= \mathrm{Var}(X_{n+h}) + \mathrm{Var}(P_n X_{n+h}) - 2\mathrm{Cov}(X_{n+h}, P_n X_{n+h}) \\
&= \mathrm{Var}(X_{n+h}) + (\Sigma_X^{-1}\Sigma_{X,X_{n+h}})' \Sigma_X \Sigma_X^{-1} \Sigma_{X,X_{n+h}} \\
&\quad - 2(\Sigma_X^{-1}\Sigma_{X,X_{n+h}})' \Sigma_{X,X_{n+h}} \\
&= \mathrm{Var}(X_{n+h}) - (\Sigma_{X,X_{n+h}})' \Sigma_X^{-1} \Sigma_{X,X_{n+h}}. \qquad (10)
\end{aligned}$$

Since the process is stationary with constant mean μ and autocovariance function $\gamma(h)$, (8) can be written as

$$a = \Gamma_n^{-1} \gamma_n(h) \quad \text{and} \quad a_0 = \mu\Big(1 - \sum_{k=1}^{n} a_k\Big), \tag{11}$$

where $\Gamma_n = [\gamma(j-k)]_{j,k=1}^n$ and $\gamma_n(h) = (\gamma(h), \gamma(h+1), \ldots, \gamma(h+n-1))'$, and (10) can be written as

$$\begin{aligned} E\big[(X_{n+h} - P_n X_{n+h})^2\big] &= \gamma(0) - \gamma_n(h)'\Gamma_n^{-1}\gamma_n(h) \\ &= \gamma(0)(1 - \rho_n(h)' R_n^{-1}\rho_n(h)), \end{aligned}$$

where $R_n = [\rho(j-k)]_{j,k=1}^n$ and $\rho_n(h) = (\rho(h), \rho(h+1), \ldots, \rho(h+n-1))'$. We notice that, hardly surprising, large absolute values for the autocorrelations make the mean squared prediction error small.

Example 5.1. Recall from (4) that $\Gamma_p\phi = \gamma_p$ describes the relation between the coefficients and the autocovariances of a causal AP(p) process

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + Z_t, \quad \{Z_t\} \sim \mathrm{WN}(0, \sigma^2).$$

However, from (11) we know that the best linear predictor of X_{n+1} based on $X_n, \ldots, X_1$ is $P_n X_{n+1} = a_0 + a_1 X_n + \cdots + a_n X_1$, where $\Gamma_n a = \gamma_n$, where $\gamma_n = \gamma_n(1)$. If $n \geq p$, then $a = (\phi_1, \ldots, \phi_p, 0, \ldots, 0)'$ therefore gives the best linear predictor of X_{n+1} based on $X_n, \ldots, X_1$, i.e.

$$P_n X_{n+1} = \phi_1 X_n + \cdots + \phi_p X_{n-p+1}. \tag{12}$$

Is there a similar nice expression for $P_n X_{n+h}$ for $h > 1$? The answer is yes. Similar to (9) it holds that

$$\mathrm{Cov}(X_{n+h} - P_{n+h-1}X_{n+h}, X_k) = 0 \quad \text{for } k = 1, \ldots, n.$$

Therefore, $P_n(X_{n+h} - P_{n+h-1}X_{n+h}) = E[X_{n+h} - P_{n+h-1}X_{n+h}]$. Since $E[P_j X_{n+h}] = E[X_{n+h}]$ for $j = 1, \ldots, n+h-1$, we find that

$$P_n X_{n+h} = P_n P_{n+h-1} X_{n+h}. \tag{13}$$

From (12) we know that

$$P_{n+h-1}X_{n+h} = \phi_1 X_{n+h-1} + \cdots + \phi_p X_{n+h-p}.$$

Inserting this expression into (13) gives

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

which provides us with a nice recursive formula for the computing the h-step best linear predictor for an AR(p)-process.

6. The partial autocorrelation function

In Section 5 we used the notation $P_n X_{n+h}$ for the best linear predictor of X_{n+h} based on $X_1, \ldots, X_n$, i.e. the random variable of the form $a_0 + a_1 X_n + \cdots + a_n X_1$ that minimizes the mean squared error $E\big[(X_{n+h} - P_n X_{n+h})^2\big]$. It is more convenient to use the notation $P(X_{n+h} \mid X_1, \ldots, X_n)$ instead of $P_n X_{n+h}$ if we want also to consider the best linear predictor based on some subset of $\{X_1, \ldots, X_n\}$ or some other set of random variables. Now this notation is used to define the partial autocorrelation function $\alpha(h)$ of a stationary time series.

The partial autocorrelation function $\alpha(h)$ at lag h of a stationary time series $\{X_t\}$ is given by

$$\alpha(h) = \begin{cases} \mathrm{Cor}(X_2, X_1), & h = 1, \\ \mathrm{Cor}(X_{h+1} - P(X_{h+1} \mid X_2, \dots, X_h), X_1 - P(X_1 \mid X_2, \dots, X_h)), & h \geq 2. \end{cases} \tag{14}$$

From (11) we know that the best linear predictor of X_{n+1} based on $X_n, \dots, X_1$ is $P_n X_{n+1} = a_{n0} + a_{n1} X_n + \cdots + a_{nn} X_1$, where $\Gamma_n a_n = \gamma_n$ with $\gamma_n = \gamma_n(1)$ and $a_n = (a_{n1}, \dots, a_{nn})'$. Corollary 5.2.1 in [1] establishes that $\alpha(n) = a_{nn}$ for $n \geq 1$.

The partial autocorrelation function is a useful tool for determining the order of an AR(p) process. From Example 5.1 we know that for a causal AR(p) process,

$$P(X_{h+1} \mid X_2, \dots, X_h) = \phi_1 X_h + \cdots + \phi_p X_{h-p+1} \quad \text{if } h - p + 1 \geq 2.$$

Hence, for $h > p$,

$$\begin{aligned} \alpha(h) &= \mathrm{Cor}(X_{h+1} - P(X_{h+1} \mid X_2, \dots, X_h), X_1 - P(X_1 \mid X_2, \dots, X_h)) \\ &= \mathrm{Cor}(Z_{h+1}, X_1 - P(X_1 \mid X_2, \dots, X_h)) \\ &= 0. \end{aligned}$$

References

[1] Brockwell, P. J. and Davis R. A. *Time Series: Theory and Methods.* Springer, 1998.

[2] Brockwell, P. J. and Davis R. A. *Introduction to Time Series and Forecasting,* 2nd edition, 8th corrected printing. Springer, 2010.

[3] Wunsch, A. D. *Complex Variables with Applications.* 2nd ed. Addison Wesley, 1994.