

### 3. Representations-2

An alternative canonical representation of a time series is

$$x_m = \sum_{k=0}^M a_k \theta_{m-k}, \quad (3.1)$$

( $M$  can be infinite), which called a moving average process of order  $M$  (an  $MA(M)$ ). Here again  $\theta_t$  is zero-mean white noise of variance  $\sigma_\theta^2$ . Notice that only positive indices  $a_k$  exist, so that  $x_m$  involves only

Continued on next page...

the *past* values of  $\theta_k$ . We can determine these  $a_k$  by the same least-squares approach of minimizing an error,

$$J = \sum_{m=0}^{N-1} \left( x_m - \sum_{k=0}^M a_k \theta_{m-k} \right)^2, \quad (3.2)$$

and leading to a set of normal equations, again producing a Toeplitz matrix. As with the AR problem, the real issue is deciding how large  $M$  should be.

*Exercise.* Derive the normal equations for (3.2) and for  $J = \langle \left( x_m - \sum_{k=0}^M a_k \theta_{m-k} \right)^2 \rangle$ .

Various theorems exist to show that any stationary discrete time series can be represented with arbitrary accuracy as either an MA or AR form. Given one form, it is easy to generate the other. Consider for example (3.1). We recognize that  $x_m$  is being represented as the convolution of the finite sequence  $a_k$  with the sequence  $\theta_m$ . Taking the Fourier (or  $z$ ) transform of  $x_m$  produces,

$$\hat{x}(z) = \hat{a}(z) \hat{\theta}(z) \quad (3.3)$$

or,

$$\hat{\theta}(z) = \frac{\hat{x}(z)}{\hat{a}(z)}. \quad (3.4)$$

Assuming that  $\hat{a}(z)$  has a stable, causal, convolution inverse, such that  $\hat{b}(z) = 1/\hat{a}(z)$ , we can write, by taking the inverse transform of (3.4)

$$\theta_m = \sum_{k=0}^L b_k x_{m-k} \quad (3.5)$$

Normalizing  $b_0 = 1$ , by dividing both sides of the last equation, we can recognize that (3.5) is in exactly the form of (1.12).

*Exercise.* Convert the moving average process  $x_m = \theta_m - 1/3\theta_{m-1} + 1/4\theta_{m-2}$  into an AR process. What order is the AR process?

Because of the reciprocal nature of the vectors  $\mathbf{a}$ ,  $\mathbf{b}$  in the AR and MA forms, it is generally true that a finite length  $\mathbf{a}$  generates a formally infinite length  $\mathbf{b}$ , and vice-versa (although in practice, one may well be able to truncate the formally infinite representation without significant loss of accuracy). The question arises as to whether a combined form, usually called an autoregressive-moving-average process (or ARMA), might produce the *most efficient* representation? That is, one might try to represent a given time series as

$$x_m - a_1 x_{m-1} - a_2 x_{m-2} - \dots - a_N x_{m-N} = \theta_t + b_1 \theta_{m-1} + \dots + b_M \theta_{m-M} \quad (3.6)$$

in such a way that the fewest possible coefficients are required. Taking the  $z$  transform of both sides of (3.6), we obtain

$$\hat{x}(z) \hat{a}(z) = \hat{\theta}(z) \hat{b}(z) \quad (3.7)$$

(defining  $a_0 = b_0 = 1$ ). One can again use least-squares in either time or frequency domains. The major issues are once again the best choice of  $M, N$  and this problem is discussed at length in the various references.

*Exercise.* An ARMA is given as

$$x_m - \frac{1}{2}x_{m-1} = \theta_m - \frac{1}{8}\theta_{m-1} \quad (3.8)$$

Convert it to (a) an AR, and (b) a MA.

If one has the simplest AR,

$$x_m = ax_{m-1} + \theta_m \quad (3.9)$$

and takes its Fourier or  $z$ -transform, we have

$$\hat{x}(z)(1 - az) = \hat{\theta}(z) \quad (3.10)$$

and dividing

$$\hat{x}(z) = \frac{\hat{\theta}(z)}{1 - az} \quad (3.11)$$

The Taylor Series about the origin is

$$\hat{x}(z) = \hat{\theta}(z) (1 + az + az^2 + az^3 + \dots) \quad (3.12)$$

which converges on  $|z| = 1$  if and only if  $|a| < 1$  and the corresponding MA form is

$$x_m = \sum_{k=0}^{\infty} a_k \theta_{m-k} \quad (3.13)$$

where the magnitude of the contribution from the remote past of  $\theta_m$  diminishes to arbitrarily small values. If we nonetheless take the limit  $a \rightarrow 1$ , we have a process

$$x_m = \sum_{k=0}^{\infty} \theta_{m-k} \quad (3.14)$$

with an apparent infinite memory of past random forcing. Note that the AR equivalent of (3.14) is simply

$$x_m = x_{m-1} + \theta_m \quad (3.15)$$

—a more efficient representation.

This last process is not stationary and is an example of what is sometimes called an ARIMA (autoregressive integrated moving average).

*Exercise:* Show that the process (3.14 or 3.15) has a variance which grows with time.

Despite the non-stationarity, (3.15) is a very simple rule to implement. The resulting time series has a number of very interesting properties, some of which are described by Wunsch (1999) and Stephenson et al. (2000) including some discussion of their applicability as a descriptor of climate change.

*Exercise.* Using a pseudo-random number generator, form a 10,000 point realization of (3.15). Calculate the mean and variance as a function of sample length  $N$ . How do they behave? What is the true mean and variance? Find the power density spectrum of the realization and describe it. Compare the results to realizations from (3.9) with  $a = 0.9999, 0.99, 0.9$  and describe the behaviors of the sample averages and variance with sample length and the changes in the spectrum.

*Exercise.* We can generalize the various representations to vector processes. Let

$$\mathbf{x}_m = [x_1(m), x_2(m), \dots, x_L(m)]^T$$

be a vector time series of dimension  $L$ . Then a vector MA form is

$$\mathbf{x}_m = \sum_{k=0}^K \mathbf{A}_k \boldsymbol{\theta}_{m-k}, \quad (3.16)$$

where the  $\mathbf{A}_k$  are matrices, and  $\boldsymbol{\theta}_m$  are vectors of white noise elements. Find the normal equations for determining  $\mathbf{A}_k$  and discuss any novel problems in their solution. Discuss the question of whether the  $\mathbf{A}_k$  should be square matrices or not. Define a vector AR form, and find the normal equations. What might the advantages be of this approach over treating each of the scalar elements  $x_j(m)$  on its own?