Probabilistic Modelling and Reasoning
Time Series Modelling: AR, MA, ARMA and All That

School of Informatics, University of Edinburgh

Instructor: Dr Chris Williams

November 2010

One common form of real-world data is that of time series, where we observe one or more variables at a number of different times. Examples include measurements of temperature at one or more locations, the prices of stocks on the stock market, measurements from an electrocardiogram of a patient's heart beats, etc. We need appropriate models for this kind of data.

This note provides a short introduction to AR, MA and ARMA models for time series data (see below for definitions). Chatfield (1989) and Diggle (1990) are useful introductory textbooks in this area; Brockwell and Davis (1991) is a more advanced text.

A *stochastic process* can be described as "a statistical phenomenon that evolves in time according to probabilistic laws" (Chatfield, 1989, p. 27). Mathematically a stochastic process is a family of random variables $X(t)$, where t runs over an index set, which for time series can be taken either as the real line, or to run over the integers (for a discrete-time process). We can think of generating an infinite set of times series (an ensemble) from a stochastic process, and that the observed time series is a possible *realization* of the stochastic process.

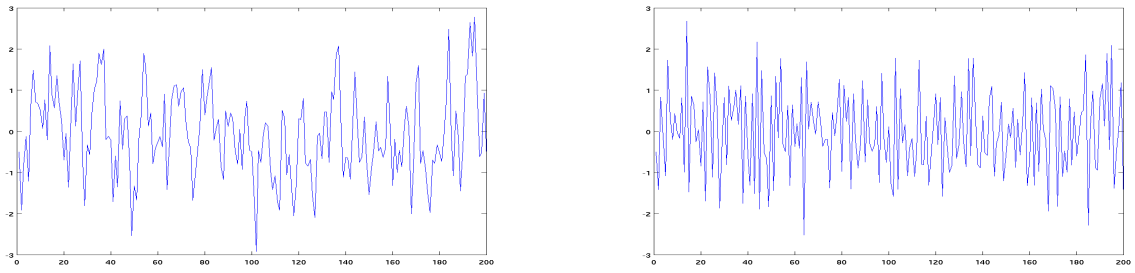
We define the *mean function* $\mu(t) = E[X(t)]$, and the autocovariance function $\gamma(t,s) = E[(X(t) - \mu(t))(X(s) - \mu(s))]$ of the stochastic process. A time series is said to be *strictly stationary* if the joint distribution of $X(t_1), \ldots, X(t_n)$ is the same as the joint distribution of $X(t_1 + \tau), \ldots, X(t_n + \tau)$ for all $t_1, \ldots, t_n, \tau$. A time series is said to be *weakly stationary* if its mean is constant and its autocovariance function depends only on the lag, i.e.

$$E[X(t)] = \mu \qquad \forall\, t,$$
$$\mathrm{Cov}[X(t)X(t+\tau)] = \gamma(\tau).$$

A special kind of stochastic process is a *Gaussian process*, which is a family of random variables, any finite number of which have a joint Gaussian distribution. Below we will consider stationary Gaussian processes, which are one of the most widely used time series models.

1 AR, MA and ARMA models

The three kinds of process you need to know about are called *autoregressive* (AR) models, *moving average* (MA) models, and autoregressive/moving average (ARMA) models; we discuss these in the following sections.



(a) (b)

Figure 1: Simulations of an AR(1) process for (a) $\alpha = 0.5$ and (b) $\alpha = -0.5$.

1.1 Autoregressive Models

We start with the simple case known as an AR(1) model, with

$$x_t = \alpha x_{t-1} + w_t, \tag{1}$$

where $w_t \sim N(0, \sigma^2)$ is a Gaussian random variable with mean zero and variance σ^2. The w's at different times are assumed independent, i.e. they are a white noise process. By repeated substitution we obtain

$$x_t = w_t + \alpha w_{t-1} + \alpha^2 w_{t-2} + \ldots \tag{2}$$

which shows that $E[X(t)] = 0$, and if $|\alpha| < 1$ the process is stationary with variance

$$\mathrm{Var}[X(t)] = (1 + \alpha^2 + \alpha^4 + \ldots)\sigma^2 = \frac{\sigma^2}{1 - \alpha^2}. \tag{3}$$

More generally, this argument shows that

$$\gamma(s) = \mathrm{Cov}[X(t)X(t-s)] = \alpha^s \mathrm{Var}[X(t-s)] = \frac{\alpha^s \sigma^2}{1 - \alpha^2}. \tag{4}$$

Note that $\gamma(s) = \alpha^s \gamma(0)$ for the AR(1) process. Samples from an AR(1) process are shown in Figure 1 for two different values of α. As expected the trace for $\alpha = -0.5$ is more oscillatory than that for $\alpha = 0.5$.

We can now generalize from an AR(1) process to the AR(p) process

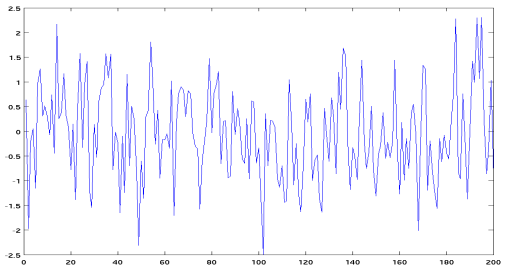
$$x_t = \sum_{i=1}^{p} \alpha_i x_{t-i} + w_t. \tag{5}$$

Notice how x_t is obtained by a (linear) regression from $x_{t-1}, \ldots, x_{t-p}$, hence the name *autoregressive*. Samples from two AR(2) processes with different parameters are shown in Fig. 2.

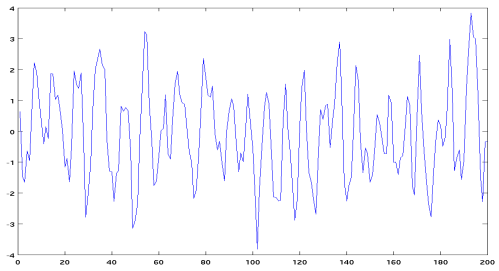
We introduce the *backward shift* operator B, so that $Bx_t = x_{t-1}$, $B^2 x_t = x_{t-2}$ etc. Then the AR(p) model can be written as

$$\phi(B)x_t = w_t, \qquad \text{with } \phi(B) = (1 - \alpha_1 B \ldots - \alpha_p B^p). \tag{6}$$

The condition for stationarity is that all the roots of $\phi(B)$ lie outside the unit circle. For an AR(1) process we have that roots of $\phi(B)$ are $\phi(B) = 1 - \alpha B = 0$ or $B = \alpha^{-1}$. The condition



(a) $\alpha_1 = 0.2$, $\alpha_2 = 0.1$



(b) $\alpha_1 = 1.0$, $\alpha_2 = -0.5$

Figure 2: Simulations of an AR(2) process.

that the roots of $\phi(B)$ lie outside the unit circle implies that $|\alpha^{-1}| > 1$, or equivalently that $|\alpha| < 1$, as discussed above.

Yule-Walker Equations: By multiplying through eq. 5 by x_{t-s}, assuming stationarity, and taking expectations we obtain

$$\gamma_s = \sum_{i=1}^{p} \alpha_i \gamma_{s-i} \qquad s = 1,\ 2, \ldots \tag{7}$$

Using p of these relations we have a system of linear equations which can be used to determine the covariances (γ's) from the α's. For the AR(1) process we obtain $\gamma_s = \alpha^s \gamma_0$ again.

1.2 Moving Average Models

Given a white noise process w, we obtain a moving average (MA) process by linear filtering, i.e.

$$x_t = \sum_{j=0}^{q} \beta_j w_{t-j} = \theta(B) w_t,$$

with scaling chosen so that $\beta_0 = 1$ and $\theta(B) = 1 + \sum_{j=1}^{q} \beta_j B^j$. We have that $E[X(t)] = 0$ (as all w's have zero mean), and that

$$\begin{aligned}
\mathrm{Var}[X(t)] &= (1 + \beta_1^2 + \ldots + \beta_q^2)\sigma^2, \\
\mathrm{Cov}[X(t)X(t-s)] &= E[\sum_{j=0}^{q} \beta_j w_{t-j}, \sum_{i=0}^{q} \beta_i w_{t-s-i}] \\
&= \begin{cases} \sigma^2 \sum_{j=0}^{q-s} \beta_{j+s}\beta_j & \text{for } s = 0,\ 1, \ldots, q \\ 0 \qquad \text{for } s > q. \end{cases}
\end{aligned}$$

Notice that covariance "cuts off" for $s > q$, in contrast to AR processes which can have infinite range correlations.

An AR(p) process can be written as a MA(∞) process

$$\begin{aligned}
\phi(B) x_t &= w_t \\
x_t &= (1 - \alpha_1 B \ldots - \alpha_p B^p)^{-1} w_t \\
&= (1 + \beta_1 B + \beta_2 B^2 \ldots) w_t.
\end{aligned}$$

Similarly a MA(q) process can be written as a AR(∞) process.

1.3 ARMA models

The ARMA(p,q) process is the natural generalization of the AR and MA processes,

$$x_t = \sum_{i=1}^{p} \alpha_i x_{t-i} + \sum_{j=0}^{q} \beta_j w_{t-j}$$
$$\phi(B)x_t = \theta(B)w_t$$

The utility of the ARMA process is its parsimony; an ARMA(p,q) process could be written as an infinite order AR or MA process, but the ARMA process gives a compact description.

1.4 The Fourier View

AR/MA/ARMA models are linear time-invariant systems, and thus sinusoids are their eigenfunctions. We can think about the corresponding process in the Fourier domain, where the power spectrum $S(k)$ is (roughly speaking) the amount of power allocated on average to the eigenfunction $e^{2\pi ikt}$. For ARMA processes the power spectrum $S(k)$ can be calculated from the $\{\alpha\}$, $\{\beta\}$ coefficients. The Fourier view is a useful way to understand some properties of ARMA processes, but we will not pursue it further here. If you want to know more, see e.g. Chatfield (1989, chapter 7) or Diggle (1990, chapter 4).

1.5 Vector AR processes

It is not necessary to restrict x_t to be a scalar quantity. For $\mathbf{x}_t$ being a vector we write

$$\mathbf{x}_t = \sum_{i=1}^{p} A_i \mathbf{x}_{t-i} + G\mathbf{w}_t \tag{8}$$

where the A_is and G are square matrices, and $\mathbf{w}_t$ is a white noise vector of the same dimension as $\mathbf{x}_t$. In general we can consider modelling multivariate (as opposed to univariate) time series, for example we might model the location of a bat in 3D space.

We can also write a scalar AR(p) process as a vector AR(1) process. For example an AR(2) process can be written as

$$\begin{pmatrix} x_t \\ x_{t-1} \end{pmatrix} = \begin{pmatrix} \alpha_1 & \alpha_2 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x_{t-1} \\ x_{t-2} \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} w_t \\ w_{t-1} \end{pmatrix} \tag{9}$$

In general an AR(p) process can be written as a vector AR(1) process with a p-dimensional state vector; this is similar to the way one can create a system of first order ordinary differential equations (ODEs) from a single higher-order ODE.

2 Parameter Estimation for ARMA models

The section above considers the properties of AR, MA and ARMA stochastic processes driven by Gaussian white noise. We now consider the statistical question of how to estimate the parameters from data. This breaks down into two parts: (i) parameter estimation for a given ARMA(p, q) model, and (ii) model order selection to choose p and q.

First of all, if the original data does not have zero mean, we estimate that as $\hat{\mu} = \sum_{i=1}^{n} x_i/n$ and subtract it off to get a zero-mean series.

For AR(p) models we recognize that eq. 5 is just a linear regression between the inputs $x_{t-1}, \ldots, x_{t-p}$ and the target x_t. Thus the coefficients $\boldsymbol{\alpha}$ and noise level σ^2 can readily estimated. This viewpoint also shows how more general AR models that are a *nonlinear* function of the inputs can be fitted in a similar fashion, using nonlinear regression.

For other ARMA models driven by Gaussian white noise, it is probably easiest to recognize that the likelihood $L(\mathbf{x}; \boldsymbol{\alpha}, \boldsymbol{\beta})$ is a multivariate Gaussian, where $\mathbf{x} = (x(t_1), \ldots, x(t_n))^T$. The likelihood can be optimized wrt the parameters with numerical optimization techniques, e.g. by gradient ascent.

We now turn to model order selection. An observation for pure MA(q) processes is that there should be a cut-off in the autocorrelation function for lags greater than q, so the empirical autocorrelation would drop to near zero. If the data comes from an AR(p) model, then we would expect that if we fit an AR($p+1$) model, then the last ($p+1$th) coefficient should be near zero. For general ARMA models we simply point out the usual problem of model selection holds, i.e. that larger models having more free parameters are likely to fit the data better (including perhaps overfitting to the noise on the data). This issue is discussed elsewhere in the PMR course.

References

Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods*. Springer-Verlag, New York, second edition.

Chatfield, C. (1989). *The Analysis of Time Series: An Introduction*. Chapman and Hall, London, 4th edition.

Diggle, P. J. (1990). *Time Series: A Biostatistical Introduction*. Clarendon Press, Oxford.