THE BASICS OF ARMA MODELS

Stationarity

A time series in discrete time is a sequence \( \{ x_t \}_{t=-\infty}^{\infty} \) of random variables defined on a common probability space. We say that \( \{ x_t \} \) is strictly stationary if the joint distributions do not change with time, i.e., if the distribution of \( \{ x_{t_1}, \ldots, x_{t_k} \} \) is the same as the distribution of \( \{ x_{t_1+\tau}, \ldots, x_{t_k+\tau} \} \) for any integers \( t_1, \ldots, t_k \), and any integer \( \tau \).

The time series is weakly stationary (covariance stationary) if \( E[x_t] = \mu \) exists and is constant for all \( t \), \( \text{var}[x_t] = \sigma_x^2 < \infty \), and the autocovariance \( E[(x_t - \mu)(x_{t-r} - \mu)] = c_r \) depends only on the lag, \( r \). Strict stationarity, together with the assumption of finite first and second moments, implies weak stationarity. For Gaussian series (i.e., all joint distributions are Gaussian), weak stationarity is equivalent to strict stationarity. This is because the multivariate Gaussian distribution is uniquely determined by the first two moments, i.e., the mean and covariance. For convenience, and without any real loss of generality, we will take the constant mean value to be zero unless otherwise stated. In this case, we have \( c_r = E[x_r x_{r-r}] \), and the autocorrelation at lag \( r \) is \( \rho_r = E[x_r x_{r-r}] / \sqrt{\text{var} x_r \text{var} x_{r-r}} = c_r / c_0 \). Given a time series data set \( x_1, \ldots, x_n \), we estimate \( c_r \) by \( \hat{c}_r = \frac{1}{n} \sum_{t=r+1}^{n} x_t x_{t-r} \) (define \( \hat{c}_{-r} = \hat{c}_r \)), and we estimate \( \rho_r \) by \( \hat{\rho}_r = \hat{c}_r / \hat{c}_0 \).

White Noise Processes

A process \( \{ \varepsilon_t \} \) is said to be white noise if it is weakly stationary, and if the \( \varepsilon_t \) are mutually uncorrelated. In the case of zero mean, we have \( E[\varepsilon_t] = 0 \), \( E[\varepsilon_t^2] = \sigma_{\varepsilon}^2 < \infty \), and \( E[\varepsilon_t \varepsilon_u] = 0 \) if \( t \neq u \). The white noise process is a basic building block for more complicated time series models. Note that we have not assumed that the \( \varepsilon_t \) are independent. Remember that independence implies uncorrelatedness, but that it is possible to have uncorrelated random variables which are not independent. In the Gaussian case, however, independence and uncorrelatedness are equivalent properties.

The white noise process is not linearly forecastable, in the sense that the best linear forecast of \( \varepsilon_{t+1} \) based on \( \varepsilon_t, \varepsilon_{t-1} \cdots \) is simply the series mean (zero), and does not depend on the present and
past observations.

**Autoregressive Processes**

We say that the process \( \{x_t\} \) is **autoregressive of order** \( p \) (AR\((p)\)) if there exist constants \( a_1, \ldots, a_p \) such that

\[
x_t = \sum_{k=1}^{p} a_k x_{t-k} + \varepsilon_t ,
\]

where \( \{\varepsilon_t\} \) is zero-mean white noise, and \( \varepsilon_t \) is uncorrelated with \( x_{t-1}, x_{t-2} \ldots \). The AR\((p)\) process exists and is weakly stationary if and only if all the roots of the polynomial \( P(z) = 1 - a_1 z \cdots - a_p z^p \) lie outside the unit circle in the complex plane. An example of an autoregressive process which is not stationary is the **random walk**, \( x_t = x_{t-1} + \varepsilon_t \), which is an AR\((1)\) with \( a_1 = 1 \). We can verify that the random walk is not stationary by noting that the root of \( P(z) = 1 - z \) is \( z = 1 \), which is on the unit circle.

For a general AR\((p)\), if we multiply \( x_t \) by \( x_{t+r} \) \( (r > 0) \) and take expectations, we obtain

\[
c_r = E[x_t \sum_{k=1}^{p} a_k x_{t+r-k} + x_t \varepsilon_{t+r}] = \sum_{k=1}^{p} a_k c_{r-k} .
\]

The equations \( c_r = \sum_{k=1}^{p} a_k c_{r-k} \) are called the **Yule-Walker Equations**.

**Moving Average Models**

We say that \( \{x_t\} \) is a **moving average of order** \( q \) (MA\((q)\)) if there exist constants \( b_1, \ldots, b_q \) such that

\[
x_t = \sum_{k=0}^{q} b_k \varepsilon_{t-k} ,
\]

where \( b_0 = 1 \). The MA\((q)\) process has a finite memory, in the sense that observations spaced more than \( q \) time units apart are uncorrelated. In other words, the autocovariance function of the MA\((q)\) cuts off beyond lag \( q \). This contrasts with the autocovariance function of the AR\((p)\), which decays exponentially, but does not cut off.
Autoregressive Moving Average Models

We say that \( \{x_t\} \) is an **autoregressive moving average process of order** \( p, q \) \((ARMA(p, q))\) if there exist constants \( a_1, \ldots, a_p, b_1, \ldots, b_q \) such that

\[
x_t = \sum_{j=1}^{p} a_j x_{t-j} + \sum_{j=1}^{q} b_j \varepsilon_{t-j} + \varepsilon_t ,
\]

where \( \{\varepsilon_t\} \) is zero-mean white noise, and \( \varepsilon_t \) is uncorrelated with \( x_{t-1}, x_{t-2}, \ldots \). The \( ARMA(p, q) \) process exists and is weakly stationary if and only if all the roots of the polynomial \( P(z) \) defined earlier are outside the unit circle. The process is said to be **invertible** if all the roots of the polynomial \( Q(z) = 1 + b_1 z + \cdots + b_q z^q \) lie outside the unit circle. A time series is invertible if and only if it has an infinite-order autoregressive \((AR(\infty))\) representation of the form

\[
x_t = \sum_{j=1}^{\infty} \pi_j x_{t-j} + \varepsilon_t ,
\]

where \( \pi_j \) are constants with \( \sum \pi_j^2 < \infty \).

**Linear Forecasting of ARMA Processes**

In the linear prediction problem, we want to forecast \( x_{t+h} \) based on a linear combination of \( x_t, x_{t-1}, \ldots \), where \( h > 0 \) is the lead time. It can be shown that the best linear forecast, i.e., the one which makes the mean squared error of prediction as small as possible, denoted by \( \hat{x}_{t+h} \), is determined by two characteristics: (1) \( \hat{x}_{t+h} \) can be expressed as a linear combination of \( x_t, x_{t-1}, \ldots \), and (2) The prediction error \( x_{t+h} - \hat{x}_{t+h} \) is uncorrelated with all linear combinations of \( x_t, x_{t-1}, \ldots \). Thus, for example, the best one-step linear forecast in the \( AR(p) \) process is \( \hat{x}_{t+1} = \sum_{k=1}^{p} a_k x_{t+1-k} \). The best one-step linear forecast in the \( MA(q) \) model is \( \hat{x}_{t+1} = \sum_{k=1}^{q} b_k \varepsilon_{t+1-k} \), which, in order to be useful in practice, must be expressed as a linear combination of the observations \( x_t, x_{t-1}, \ldots \), by inverting the process to its \( AR(\infty) \) representation.
Optimal Forecasting

It is not necessarily true that the best linear forecast is the best possible forecast. It may be that some nonlinear function of the observations gives a smaller mean squared prediction error than the best linear forecast. It can be shown that the optimal forecast, in terms of mean squared error, is the conditional expectation, $E [x_{t+h} | x_t, x_{t-1}, \cdots ]$. This is a function of the present and past observations. If the $\{x_t\}$ are Gaussian, then this function will be linear. Thus, in the Gaussian case, the optimal predictor is the same as the best linear predictor. In fact, for any zero-mean process with a one-sided $MA(\infty)$ representation, $x_t = \sum_{k=0}^{\infty} b_k \varepsilon_{t-k}$, if the conditional expectation of $\varepsilon_{t+1}$ given the available information is zero, it will be true that the optimal predictor is the same as the best linear predictor. In general, however, the optimal predictor may be nonlinear.