(ii) $\mathrm{Cov}(X_t, X_{t+k})$ *is independent of* t *for each* k.

Definition 2.2 *A time series* $\{X_t, t = 0, \pm 1, \pm 2, \cdots\}$ *is strictly stationary if* $(X_1, \cdots, X_n)$ *and* $(X_{1+k}, \cdots, X_{n+k})$ *have the same joint distributions for any integer* $n \geq 1$ *and any integer* k.

The stationarity, which is often referred to as the weak stationarity in textbooks, assumes that only the first two moments of time series are time-invariant and is generally weaker than the strict stationarity, provided that the process has finite second moments. Weak stationarity is primarily used for linear time series, such as ARMA processes, where we are mainly concerned with the linear relationships among variables at different times. In fact, the assumption of stationarity suffices for most linear time series analysis, such as in spectral analysis. In contrast, we have to look beyond the first two moments if our focus is on nonlinear relationships. This explains why strict stationarity is often required in the context of nonlinear time series analysis.

2.1.2 Stationary ARMA *Processes*

It is obvious that a white noise process $\mathrm{WN}(0, \sigma^2)$ is stationary but not necessarily strictly stationary; see §1.3.1. In view of the discussion above, it is natural to use $\mathrm{WN}(0, \sigma^2)$ as a building block for general linear time series. For Gaussian time series, we need only to focus on the properties of the first two moments, too. A stationary Gaussian process is also strictly stationary.

First, we consider moving average models. It is easy to see from (1.2) that any $\mathrm{MA}(q)$ process with finite q is stationary. Let us consider an $\mathrm{MA}(\infty)$ model defined as

$$X_t = \sum_{j=0}^{\infty} a_j \varepsilon_{t-j} \quad \text{for all } t, \tag{2.1}$$

where $\{\varepsilon_t\} \sim \mathrm{WN}(0, \sigma^2)$ and $\sum_{j=0}^{\infty} |a_j| < \infty$. Therefore

$$E|X_t| \leq E|\varepsilon_1| \sum_{j=0}^{\infty} |a_j| < \infty.$$

This implies that the infinite sum on the right-hand side of (2.1) converges in probability, and also in mean of order 1 as well as order 2, as the condition $\sum_j |a_j| < \infty$ implies $\sum_j a_j^2 < \infty$. (Under the additional condition that $\{\varepsilon_t\}$ is independent, the infinite sum also converges almost surely due to the Loève theorem; see Corollary 3, p. 117 of Chow and Teicher 1997.) Furthermore, $EX_t = 0$, and

$$\mathrm{Cov}(X_t, X_{t+k}) = \sum_{j,l=0}^{\infty} a_j a_k E(\varepsilon_{t-j}, \varepsilon_{t+k-l}) = \sigma^2 \sum_{j=0}^{\infty} a_j a_{j+|k|}, \tag{2.2}$$

which is independent of t. Hence, such an MA(∞) model also defines a stationary process. Obviously, it $\{\varepsilon_t\}$ are i.i.d and $E|\varepsilon_t| < \infty$. The process $\{X_t\}$ defined by (2.1) is strictly stationary.

For a general ARMA (p, q) model defined in (1.3), we may express the process in a compact form in terms of backshift operator B as:

$$b(B)X_t = a(B)\varepsilon_t \quad \text{for all } t, \tag{2.3}$$

where B is the *backshift operator* defined as

$$B^k X_t = X_{t-k} \quad k = 0, \pm 1, \pm 2, \cdots,$$

and $b(\cdot)$ and $a(\cdot)$ are polynomials given by

$$b(z) = 1 - b_1 z - \cdots - b_p z^p, \quad a(z) = 1 + a_1 z + \cdots + a_q z^q. \tag{2.4}$$

Remark 2.1 For ARMA models as defined in (2.3), we always assume that polynomials $b(\cdot)$ and $a(\cdot)$ do not have common factors. Otherwise, a process so defined is effectively equivalent to the process with orders smaller than (p, q) after removing those common factors.

Theorem 2.1 *The process* $\{X_t,\, t = 0, \pm 1, \pm 2, \cdots\}$ *given by (2.3) is stationary if* $b(z) \neq 0$ *for all complex numbers* z *such that* $|z| \leq 1$.

Proof. Let $z_1, \cdots, z_p$ be the roots of $b(z) = 0$. Then $|z_j| > 1$ and $b(z) = \prod_{1 \leq j \leq p}(1 - z/z_j)$. It follows from some simple Taylor expansions that for any $|z| \leq 1$,

$$b(z)^{-1} = \prod_{j=1}^{p} \left(1 - z/z_j\right)^{-1} = \prod_{j=1}^{p} \left\{ \sum_{k=0}^{\infty} (z/z_j)^k \right\} \equiv \sum_{j=0}^{\infty} c_j z^j.$$

Note that

$$\sum_{j=0}^{\infty} |c_j| \leq \prod_{j=1}^{p} \left\{ \sum_{k=0}^{\infty} 1/|z_j|^k \right\} = \prod_{j=1}^{p} \left(1 - 1/|z_j|\right)^{-1} < \infty.$$

Now, write $c(z) = \sum_{j \geq 0} c_j z^j$. Then $c(z)\, b(z) \equiv 1$. Therefore

$$X_t = c(B)b(B)X_t = c(B)a(B)\varepsilon_t = d(B)\varepsilon_t, \tag{2.5}$$

where $d(z) = c(z)a(z) = \sum_{j=0}^{\infty} d_j z^j$ with $\sum_{j=0}^{\infty} |d_j| < \infty$. This indicates that $\{X_t\}$ is effectively an MA(∞) process defined as in (2.1) and is therefore stationary. ∎

Another important concept in time series is causality.

Definition 2.3 *A time series* $\{X_t\}$ *is causal if for all* t

$$X_t = \sum_{j=0}^{\infty} d_j \varepsilon_{t-j}, \quad \sum_{j=0}^{\infty} |d_j| < \infty,$$

where $\{\varepsilon_t\} \sim \mathrm{WN}(0, \sigma^2)$.