Chapter 2

Moving Averages

2.1 White Noise Processes

There are two extreme examples of weakly stationary processes. The first is the constant process $x_t = A$ for all t, where A is a stochastic variable in $L^2(\Omega, \mathcal{F}, P)$. In this case the autocovariance function is also a constant, $\gamma_k = \text{Var}(A)$. The second is that in which, on the contrary, there is no covariance between the stochastic variables corresponding to t and $t-k$, for all $k \neq 0$.

Definition 2.1 The weakly stationary process u_t is a *white noise* if $E(u_t) = 0$ and $\gamma_k = 0$ for all $k \neq 0$.

Making heavy use of a language that will be made precise later on, we can say that the process $x_t = A$ for all t is *perfectly predictable*, this meaning that x_t may be expressed as a function of its *past values*, i.e. the stochastic variables x_{t-k} for $k > 0$, with no residual: Indeed $x_t = x_{t-1}$. On the contrary a white noise u_t is *unpredictable*, in the following sense. Given the weakly stationary process x_t, consider the projection

$$x_t = b\mathbf{1} + a_{n1}x_{t-1} + a_{n2}x_{t-2} + \cdots + a_{nn}x_{t-n} + R_{nt},$$

where $\mathbf{1}$ is the function of $L^2(\Omega, \mathcal{F}, P)$ which associates 1 with every $\omega \in \Omega$ (the unit constant). As we know, the coefficients a_{nj} are chosen in such a way that the norm of the residual $x_t - (b\mathbf{1} + a_1x_{t-1} + \cdots + a_nx_{t-n})$ be minimum. For this reason we will label the stochastic variables $b\mathbf{1} + a_{n1}x_{t-1} + a_{n2}x_{t-2} + \cdots + a_{nn}x_{t-n}$ and R_{nt} as the optimal linear predictor and the prediction error respectively. Now, if the non-zero-lag autocovariances of x_t are all zero, then $a_{nj} = 0$ for all j and the projection above becomes

$$x_t = E(x_t) + [x_t - E(x_t)]$$

(just check that if $\gamma_k = 0$ for $k > 0$ then $x_t - E(x_t)$ is orthogonal to $\mathbf{1}$ and x_{t-k} for $k > 0$). Thus the best linear predictor of x_t, based on $\mathbf{1}$ and x_{t-k}, $k = 1, \ldots, n$, is the mean of x_t. In other words, taking linear combinations of past x's does not provide any help to predict x_t. In particular, if x_t is a white noise, then $R_t = x_t$.

It is important to note that in general a white noise is unpredictable if the best predictor belongs to the class of *linear* functions of past values of x_t. We will give examples in which a white noise is predictable if non linear predictors (non linear functions of past values) are allowed. However, if u_t and u_{t-k} are not only uncorrelated but independent, then u_t is unpredictable even if non linear predictors are allowed (see Chapter 4).

Example 2.1 If the variables u_t are *identically and independently distributed* (IID), then the process is strongly stationary and white noise. The tossing coin process, Example 1.1, is the most elementary member of this family (note that you have to subtract the mean in order to fulfill definition 2.1).

Example 2.2 Normal real process. The stochastic variables x_t are real and jointly normal. This means that the distribution of the variables $x_{t_1}, x_{t_2}, \ldots, x_{t_n}$, $t_1 < t_2 < \cdots < t_n$ has the normal density with covariance function

$$\begin{pmatrix} \gamma_0 & \gamma_{k_1} & \cdots & \gamma_{k_{n-1}} \\ \gamma_{k_1} & \gamma_0 & \cdots & \gamma_{k_{n-2}} \\ \vdots & & & \\ \gamma_{k_{n-1}} & \gamma_{k_{n-2}} & \cdots & \gamma_0 \end{pmatrix}$$

where $k_j = t_j - t_{j-1}$. If the variables x_t are zero mean, then the process $\{x_t,\ t \in \mathbb{Z}\}$ is a white noise if and only if it is IID.

Example 2.3 Normal complex process. The joint normal distribution for complex stochastic variables is not obvious. Assume for simplicity that all the stochastic variables are zero mean. Let us begin with the straightforward definition: The stochastic variables $z = z_1 + i z_2$ and $w = w_1 + i w_2$ are jointly normal if the real vector $\begin{pmatrix} z_1 & z_2 & w_1 & w_2 \end{pmatrix}$ is normally distributed. Now assume that the covariance between z and w is zero, i.e. that

$$E(z\,\overline{w}) = E(z_1 w_1 + z_2 w_2 + i(z_2 w_1 - z_1 w_2)) = \sigma_{11}^{zw} + \sigma_{22}^{zw} + i(\sigma_{21}^{zw} - \sigma_{12}^{zw}) = 0.$$

This implies that the covariance matrix of the real vector $\begin{pmatrix} z_1 & z_2 & w_1 & w_2 \end{pmatrix}$ has the north-east 2×2 submatrix

$$\begin{pmatrix} \sigma_{11}^{zw} & \sigma_{12}^{zw} \\ \sigma_{12}^{zw} & -\sigma_{11}^{zw} \end{pmatrix}. \tag{2.1}$$

Thus zero correlation in the complex case does not imply independence. To obtain the usual implication we have to introduce restrictions in the definition of jointly normal complex variables. An elegant way is the following: z and w are jointly normally distributed if the real vector $\begin{pmatrix} z_1 & z_2 & w_1 & w_2 \end{pmatrix}$ is normally distributed and

$$E(z^2) = E(w^2) = E(zw) = 0, \tag{2.2}$$

i.e. z is orthogonal to $\bar{z}$, w to $\bar{w}$ and z to $\bar{w}$ (see [14], p. 147-8). As a consequence, the covariance matrix of the real vector $\begin{pmatrix} z_1 & z_2 & w_1 & w_2 \end{pmatrix}$ is

$$\begin{pmatrix} \sigma_z^2 & 0 & \sigma_{11}^{zw} & 0 \\ 0 & \sigma_z^2 & 0 & \sigma_{11}^{zw} \\ \sigma_{11}^{zw} & 0 & \sigma_w^2 & 0 \\ 0 & \sigma_{11}^{zw} & 0 & \sigma_w^2 \end{pmatrix}.$$

As zero correlation implies (2.1), then zero correlation, under (2.2), implies $\sigma_{11}^{zw} = 0$ and therefore independence between z and w. Along this line of reasoning, a complex stationary process $x_t = x_{1t} + i x_{2t}$ is normal if

$$E(x_t \; x_{t-k}) = 0 \text{ for all } k \text{ (including } k = 0) \tag{2.3}$$

(zero covariance between x_t and $\overline{x_{t-k}}$), so that a normal complex white noise, under (2.3), is IID and the covariance matrix of the real vector $(x_{1t} \; x_{2t})$ is

$$\begin{pmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_x^2 \end{pmatrix}.$$

Example 2.4 Consider again Exercise 1.28 and, as in Example 1.7, assume that ϕ is uniformly distributed in the interval $[-\pi \; \pi]$. Assume also that $P(A = 0) = 0$. We have

$$\begin{aligned} \gamma_k &= \sigma_A^2 E(e^{i\phi k}) = \sigma_A^2 \int_{-\pi}^{\pi} e^{i\phi k} d\phi = \int_{-\pi}^{\pi} \cos(\phi k) d\phi + i \int_{-\pi}^{\pi} \sin(\phi k) d\phi \\ &= \frac{1}{k} [\sin(\phi k)]_{-\pi}^{\pi} + i \frac{1}{k} [-\cos(\phi k)]_{-\pi}^{\pi} . \end{aligned}$$

The sine and cosine functions appearing in this expression are periodic with period $2\pi/k$ and therefore also with period 2π, so that $\gamma_k = 0$ for $k \neq 0$ (see Example 1.7). Thus x_t is a white noise. However, x_t is perfectly predictable. For, from the definition of x_t,

$$e^{i\phi} = \cos\phi + i \sin\phi = x_{t-1}/x_{t-2}$$

(the assumption on A ensures that the right hand side makes sense almost surely). Thus the stochastic variable ϕ is determined, up to an integer multiple of 2π, by x_{t-1} and x_{t-2}. But this is all we need to determine $e^{i\phi\tau}$, because $e^{i\phi\tau} = e^{i(\phi+2s\pi)\tau}$ for integer values of τ and s (see Observation 1.15, 8). Then, almost surely,

$$A = x_{t-1}e^{-i\phi(t-1)}.$$

In conclusion, A and ϕ are completely determined as *non linear* functions of x_{t-1} and x_{t-2}. Therefore the white noise x_t can be perfectly predicted, though of course non linearly, using past values.

The lesson from Example 2.4 should be carefully kept in mind:
(i) The second moments of a weakly stationary process say little on the probability distributions $\mu_{t_1,\dots,t_n}$ or, which is equivalent, on the probability distribution of the process realizations in $\mathbb{C}^{\mathbb{Z}}$ (see p. 7); in the case of Example 2.4 the probability is concentrated on a subset of very regular realizations.
(ii) If non-linear operations are allowed on past values of x_t, then a white noise can be partially or even perfectly predicted.

Observation 2.1 Obviously, the definition of white noise does not rule out the possibility that $\sigma_u^2 = \int_\Omega |x_t(\omega)|^2 dP(\omega) = 0$. But this implies that $x_t = 0$ almost surely. Conversely, $u_t = 0$ almost surely implies $\sigma_u^2 = 0$ (see Observation 1.9). We shall use indifferently $\sigma_u^2 = 0$ or $u_t = 0$.

Summary. A white noise has been defined by assuming that the autocorrelation between the variables at times t and $t-k$ is zero for $k \neq 0$. Though this definition is sufficient for all the results that are based on second moments, we have shown by an example that "white noise" does not implies "unpredictable", but only "linearly unpredictable". This will be a theme of detailed discussion in Chapter 4.