THE YULE WALKER EQUATIONS

The Yule-Walker equations arise naturally in the problem of linear prediction of any zero-mean weakly stationary process $\{x_t\}$ based on a finite number of contiguous observations. First, we will consider the case where $\{x_t\}$ is the $AR(p)$ process $\sum_{k=0}^{p} a_k x_{t-k} = \varepsilon_t$, where $a_0 = 1$, and $var\ \varepsilon_t = \sigma_p^2$. If $\{c_r\}$ are the autocovariances, then

$$E[x_t\ \varepsilon_t] = E\ [x_t\ \sum_{k=0}^{p} a_k x_{t-k}] = \sum_{k=0}^{p} a_k c_k \quad .$$

On the other hand,

$$E[x_t\ \varepsilon_t] = E[(\varepsilon_t - \sum_{k=1}^{p} a_k x_{t-k})\ \varepsilon_t] = E[\varepsilon_t^2] = \sigma_p^2 \quad .$$

Therefore,

$$\sum_{k=0}^{p} a_k c_k = \sigma_p^2 \quad . \tag{1}$$

For $l > 0$, we have

$$0 = E[x_t\ \varepsilon_{t+l}] = E[x_t \sum_{k=0}^{p} a_k x_{t+l-k}] = \sum_{k=0}^{p} a_k c_{l-k} \quad . \tag{2}$$

Writing (1) and (2) for $l = 1, \ldots, p$ in matrix form, we have the **Yule-Walker equations:**

$$\begin{bmatrix} c_0 & c_1 & \cdots & c_p \\ c_1 & c_0 & \cdots & c_{p-1} \\ \vdots & \vdots & \cdots & \vdots \\ c_p & c_{p-1} & \cdots & c_0 \end{bmatrix} \begin{bmatrix} 1 \\ a_1 \\ \vdots \\ a_p \end{bmatrix} = \begin{bmatrix} \sigma_p^2 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad . \tag{3}$$

Let Σ_p denote the $(p+1)\times(p+1)$ Toeplitz covariance matrix in (3). Note that Σ_p is nonnegative definite, since for any $p+1$-dimensional vector $b = (b_0, \ldots, b_p)'$ we have

$$b'\,\Sigma_p\ b = \sum_{j=0}^{p} \sum_{k=0}^{p} b_j c_{j-k} b_k = var\ \sum_{j=0}^{p} b_j x_{t-j}\ \geq 0 \quad .$$

Note that we did not need to assume that $\{x_t\}$ was $AR(p)$ to get this last result.

Suppose now that $\{x_t\}$ is any weakly stationary zero mean process with autocovariance sequence $\{c_r\}$, and we want to find the best linear predictor of x_t based on $x_{t-1}, \ldots, x_{t-p}$. Writing the (one-step) predictor as $\hat{x}_t = -\sum_{k=1}^{p} b_k x_{t-k}$, the mean squared prediction error is (with $b_0 = 1$)

$$E[x_t - \hat{x}_t]^2 = E[\sum_{k=0}^{p} b_k x_{t-k}]^2 = E[\sum_{j=0}^{p} b_j x_{t-j} \sum_{k=0}^{p} b_k x_{t-k}] = \sum_{j=0}^{p} \sum_{k=0}^{p} b_j c_{j-k} b_k = b' \Sigma_p b \quad .$$

We will now show that $b' \Sigma_p b$ is minimized (subject to the constraint that $b_0 = 1$) by taking $b = a$, where $a = (1, a_1, \ldots, a_p)'$ is the solution to the Yule-Walker equations (3), and that the resulting minimum attainable mean squared prediction error is σ_p^2. Thus, by solving the Yule-Walker equations for $a_1, \ldots, a_p$, and σ_p^2, we obtain the coefficients of the best linear predictor of x_t based on $x_{t-1}, \ldots, x_{t-p}$ and the corresponding minimum mean squared error of prediction, even if $\{x_t\}$ itself is not $AR(p)$.

Here is the proof of the theorem stated above. Note that

$$b' \Sigma_p b = (a + (b-a))' \Sigma_p (a + (b-a))$$

$$= a' \Sigma_p a + 2(b-a)' \Sigma_p a + (b-a)' \Sigma_p (b-a) \quad .$$

The first term is

$$a' \Sigma_p a = (1, a_1, \ldots, a_p) \begin{bmatrix} \sigma_p^2 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \sigma_p^2 \quad .$$

The second term is

$$2(b-a)' \begin{bmatrix} \sigma_p^2 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = 0 \quad ,$$

since the first entry of $b-a$ is $b_0 - a_0 = 1 - 1 = 0$. It follows that the mean squared error of the linear predictor b is

$$b' \Sigma_p b = \sigma_p^2 + (b-a)' \Sigma_p (b-a) \quad .$$

Note that $(b-a)'\Sigma_p(b-a)$ will always be nonnegative, and will be zero if we take $b=a$. Thus, the best linear predictor is a, and the resulting minimum attainable mean squared prediction error is

$$a'\Sigma_p\, a = \sigma_p^2 \ .$$