ID6

ARMAX Identification

6.12 MAXIMUM LIKELIHOOD ESTIMATES

Maximum likelihood (ML) estimates are based on statistical considerations that lead, however, to an estimate which minimizes the prediction error, i.e. to a PEM (Prediction Error Method) estimate. Consider vector $y^N \in \mathcal{R}^N$ whose entries are N output observations performed on the ARMAX process (6.1.2)

$$y^N = \left[\, y(1)\;\; y(2)\;\; \dots\;\; y(N)\, \right]^T; \tag{6.12.1}$$

the entries of this vector, because of the presence of $w(t)$ in (6.1.2), are stochastic variables. Denote then with

$$p\big(y_1, y_2, \dots, y_N; \theta\big) = p\big(y^N; \theta\big) \tag{6.12.2}$$

the associated probability density function (pdf), conditioned by θ. Considering a subset, $\mathcal{S} \subseteq \mathcal{R}^N$, the probability that $y^N \in \mathcal{S}$ is given by

$$p\big(y^N \in \mathcal{S}\big) = \int_{x^N \in \mathcal{S}} p\big(x^N; \theta\big) \mathrm{d}x^N. \tag{6.12.3}$$

The maximum likelihood method consists in determining the estimate $\hat{\theta}(y^N)$ which maximizes the probability of observations y^N; this method requires the knowledge of the conditioned probability density function (6.12.2). It can be observed that the probability of the observations is conditioned not only by the parameter vector θ, but also by the order of the process. To apply the ML method in estimating the parameters of ARMAX processes we will introduce the assumption, often realistic, that the pdf of $w(\cdot)$ is Gaussian. The one to one transformation between $w(t)$ and $y(t)$ described by (6.1.12) implies that, neglecting the effect of initial conditions, the pdf of $y(t)$ is also Gaussian; furthermore it is possible to make reference to the probability density function of the residuals, $\varepsilon(t, \theta) = y(t) - y(t|t-1; \theta)$, instead of $p(y^N; \theta)$. We will refer thus to the pdf

$$L(\theta) = \frac{1}{\sqrt{(2\pi\sigma_\varepsilon^2(\theta))^N}} \exp\left[-\frac{1}{2\sigma_\varepsilon^2(\theta)} \sum_{t=1}^{N} \varepsilon^2(t, \theta) \right], \tag{6.12.4}$$

which is called the Likelihood Function; in (6.12.4) $\sigma_\varepsilon^2(\theta)$ denotes the variance of $\varepsilon(t,\theta)$. It is usually preferred to consider the logarithmic likelihood function that can be obtained taking the natural logarithms of both members of (6.12.4),

$$\log L(\theta) = -\frac{1}{2\sigma_\varepsilon^2(\theta)}\sum_{t=1}^{N}\varepsilon^2(t,\theta) - \frac{N}{2}\log\sigma_\varepsilon^2(\theta) - \frac{N}{2}\log 2\pi; \tag{6.12.5}$$

since the logarithm is a monotone function, we obtain the same estimate of θ maximizing (6.12.5) or (6.12.4). Expression (6.12.5) will be now used to estimate the α_i, β_i and γ_i parameters and $\sigma_\varepsilon^2(\theta)$. It is convenient to rely on a simpler notation, to consider $\sigma_\varepsilon^2(\theta)$ as a further variable to be estimated beyond θ; it will be denoted, in the following with λ. The likelihood function can thus be written in the form

$$\begin{aligned}\log L(\theta,\lambda) &= -\frac{1}{\lambda}\frac{N}{2}\frac{1}{N}\sum_{t=1}^{N}\varepsilon^2(t,\theta) - \frac{N}{2}\log\lambda + \text{const.}\\ &= -\frac{N}{2}\left[\frac{J_N(\theta)}{\lambda} + \log\lambda\right] + \text{const.}\end{aligned} \tag{6.12.6}$$

To determine the stationary points of $\log L(\theta,\lambda)$ we will annihilate its derivative with respect to λ, given by

$$\frac{\partial \log L(\theta,\lambda)}{\partial\lambda} = -\frac{N}{2}\left[-\frac{J_N(\theta)}{\lambda^2} + \frac{1}{\lambda}\right]. \tag{6.12.7}$$

The only stationary point is

$$\lambda^\circ = J_N(\theta) \tag{6.12.8}$$

where the second–order derivative,

$$\frac{\partial^2 \log L(\theta,\lambda)}{\partial\lambda^2} = \frac{N}{2}\left[-\frac{2J_N(\theta)}{\lambda^3} + \frac{1}{\lambda^2}\right] \tag{6.12.9}$$

is negative; it corresponds thus to a maximum where σ_ε^2 is

$$\sigma_\varepsilon^2 = \lambda^\circ = J_N(\theta) \tag{6.12.10}$$

can be computed after the estimate of θ. Substituting $J_N(\theta)$ in (6.12.6) with expression (6.12.10) we obtain

$$\log L(\theta,\lambda^\circ) = -\frac{N}{2}\log J_N(\theta) + \text{const.}; \tag{6.12.11}$$

it follows that the ML estimate of θ can be obtained minimizing the mean square prediction error $J_N(\theta)$ and that PEM estimates are also ML estimates when $w(\cdot)$ is Gaussian.