Facultad de Ciencias Exactas, Ingeniería y Agrimensura
Universidad Nacional de Rosario

Identificación de Sistemas

Título: Análisis de la Estima de Mínimos Cuadrados **Fecha: Abril de 2002**
Autor: Dr. Juan Carlos Gómez

1. Introducción

Dada la estructura de **regresor lineal**

$$y(n) = \varphi^T(n)\theta + v(n) \tag{1.1}$$

la estima de mínimos cuadrados (lineales) se obtiene minimizando un criterio cuadrático de los errores de predicción, es decir

$$\hat{\theta}_N^{LS} = \underset{\theta}{\operatorname{argmin}}\, V_N(\theta) = \underset{\theta}{\operatorname{argmin}}\left\{\frac{1}{N}\sum_{n=1}^{N}\left(y(n)-\varphi^T(n)\theta\right)^T\left(y(n)-\varphi^T(n)\theta\right)\right\}$$

resultando

$$\hat{\theta}_N^{LS} = \left[\frac{1}{N}\sum_{n=1}^{N}\varphi(n)\varphi^T(n)\right]^{-1}\left[\frac{1}{N}\sum_{n=1}^{N}\varphi(n)y(n)\right]$$

Si se caracteriza a la perturbación $v(n)$ como un **proceso estocástico**, entonces la estima de mínimos cuadrados resulta una **variable aleatoria**, y tiene sentido analizar sus propiedades estadísticas. En particular interesa determinar la **consistencia**[1] de la estima, es decir analizar a qué tiende la estima $\hat{\theta}_N^{LS}$ cuando el número de datos tiende a infinito ($N \to \infty$).

Dada la estructura de modelo (1.1) se asumirá para el análisis que existe una realización particular $v_0(n)$ del ruído, y un valor θ_0 del vector de parámetros tal que los datos puedan expresarse como

$$y(n) = \varphi^T(n)\theta_0 + v_0(n)$$

A θ_0 se lo denomina **valor verdadero** del vector de parámetros.

La estimación de mínimos cuadrados puede pensarse como un mapeo entre el espacio de los datos de medición $Z^N = \{y(n), u(n)\}_{n=1}^N$ y el espacio de parámetros $D_M \subset \Re^p$, *i.e.*

$$Z^N \to \hat{\theta}_N^{LS} \in D_M \subset \Re^p$$

La idea es analizar que tanto se aproxima $\hat{\theta}_N^{LS}$ a θ_0 cuando $N \to \infty$, en un marco probabilístico.

2. Análisis de Consistencia

La estima de mínimos cuadrados viene dada por

[1] Ver la sección 5. Apéndice, por la definición de **estimador consistente**.

$$\hat{\theta}_N^{LS} = \left[\frac{1}{N}\sum_{n=1}^{N}\varphi(n)\varphi^T(n)\right]^{-1}\left[\frac{1}{N}\sum_{n=1}^{N}\varphi(n)y(n)\right]$$

$$\hat{\theta}_N^{LS} = \left[\frac{1}{N}\sum_{n=1}^{N}\varphi(n)\varphi^T(n)\right]^{-1}\left[\frac{1}{N}\sum_{n=1}^{N}\varphi(n)\varphi^T(n)\theta_0 + \varphi(n)v_0(n)\right]$$

$$\hat{\theta}_N^{LS} = \theta_0 + \left[\frac{1}{N}\sum_{n=1}^{N}\varphi(n)\varphi^T(n)\right]^{-1}\left[\frac{1}{N}\sum_{n=1}^{N}\varphi(n)v_0(n)\right]$$

y se puede, entonces, escribir el error de la estima como

$$\hat{\theta}_N^{LS} - \theta_0 = \left[\frac{1}{N}\sum_{n=1}^{N}\varphi(n)\varphi^T(n)\right]^{-1}\left[\frac{1}{N}\sum_{n=1}^{N}\varphi(n)v_0(n)\right] \tag{2.1}$$

Para asegurar la consistencia de la estima debe asumirse que la primer sumatoria en (2.1) converge a una matriz invertible cuando $N \to \infty$, es decir

$$\frac{1}{N}\sum_{n=1}^{N}\varphi(n)\varphi^T(n) \to R_\varphi(0) \quad as\, N \to \infty \quad R_\varphi(0)\, invertible$$

La consistencia requiere entonces que la segunda sumatoria tienda a cero cuando $N \to \infty$. Si se asume además que los procesos son **ergódicos**, en el límite cuando $N \to \infty$, puede reemplazarse la media temporal por la media muestral, es decir:

$$\frac{1}{N}\sum_{n=1}^{N}\varphi(n)v_0(n) \to E\left[\varphi(n)v_0(n)\right] \tag{2.2}$$

y por lo tanto la condición

$$\frac{1}{N}\sum_{n=0}^{N}\varphi(n)v_0(n) \xrightarrow{N\to\infty} 0 \ \Rightarrow E[\varphi(n)v_0(n)] = 0$$

Es decir, $\varphi(n)$ y $v_0(n)$ tienen que ser **no correlacionados**. Esta condición se cumple si los regresores son determinísticos o, en el caso de que no lo sean, si el ruido es ruido blanco.

Concluyendo se puede decir que si se verifica que

$$\frac{1}{N}\sum_{n=1}^{N}\varphi(n)v_0(n) \to E\left[\varphi(n)v_0(n)\right]$$

cuando $N \to \infty$ y además

$$\lim_{N\to\infty}\left[\frac{1}{N}\sum_{n=0}^{N}\varphi(n)\varphi^T(n)\right]$$

es no singular, entonces para que la estima de mínimos cuadrados sea **consistente** se debe verificar que

$$E\left[\varphi(n)v_0(n)\right] = 0,$$

es decir que los regresores estén **no correlacionados** con el ruido. Esto se da si:

- $\{v_0(n)\}$ es ruido blanco (variable aleatoria independiente e idénticamente distribuido (iid))

o bien si

- la secuencia de regresores $\{\varphi(n)\}$ es una secuencia determinística

3. Propiedades de la Estima de Mínimos Cuadrados

El siguiente Lema establece las condiciones bajo las cuales la estima de Mínimos Cuadrados es una estima **no desviada**, y además provee una expresión para la covarianza de la estima y una estima no desviada de la covarianza del ruido.

Lema ([2], [1])
Dada la estructura de modelo (1.1)

$$y(n) = \varphi^T(n)\theta + \nu(n) \rightarrow Y_N = \Phi^T\theta + \nu_N \tag{3.1}$$

con $\nu(n) = e(n)$ ruido blanco con media nula y covarianza σ^2, se asume que existe θ_0 tal que los datos verifican

$$y(n) = \varphi^T(n)\theta_0 + e(n) \rightarrow Y_N = \Phi^T\theta_0 + \nu_N$$

para alguna realización particular del ruido. Entonces se verifica:

1. $\hat{\theta}_N^{LS}$ es una estima no desviada de θ, i.e. $E\left\{\hat{\theta}_N^{LS}\right\} = \theta$

Si además los regresores son determinísticos

2. $\operatorname{cov}\left(\hat{\theta}_N^{LS}\right) = \sigma^2\left(\Phi\Phi^T\right)^{-1}$. Provee la desviación respecto a la media, es decir da una cuantificación de la incertidumbre de la estima.

3. Una estima no desviada de σ^2 viene dada por $\hat{\sigma}_N^2 = \dfrac{V_N\left(\hat{\theta}_N^{LS}\right)N}{N-p}$, donde $p = \dim\theta$.

Prueba

1. La estima de mínimos cuadrados esta dada por

$$\hat{\theta}_N^{LS} = \left(\Phi\Phi^T\right)^{-1}\Phi Y_N = \left(\Phi\Phi^T\right)^{-1}\Phi\left(\Phi^T\theta_0 + \nu_N\right) = \theta_0 + \left(\Phi\Phi^T\right)^{-1}\Phi\,\nu_N$$

entonces

$$E\left\{\hat{\theta}_N^{LS}\right\} = \theta_0 + E\left\{\left(\Phi\Phi^T\right)^{-1}\Phi\,\nu_N\right\}$$

y como ν_N es ruido blanco es no correlacionado siempre, por lo tanto

$$E\left\{\hat{\theta}_N^{LS}\right\} = \theta_0$$

2.

$$E\left\{\left(\hat{\theta}_N^{LS} - \theta_0\right)\left(\hat{\theta}_N^{LS} - \theta_0\right)^T\right\} = E\left\{\left(\Phi\Phi^T\right)^{-1}\Phi\,\nu_N\nu_N^T\Phi^T\left(\Phi\Phi^T\right)^{-1}\right\} =$$

$$= \left(\Phi\Phi^T\right)^{-1}\Phi E\left\{\nu_N\nu_N^T\right\}\Phi^T\left(\Phi\Phi^T\right)^{-1}$$

(por la hipótesis de que los regresores son determinísticos)

y como $E\left\{\nu_N\nu_N^T\right\} = \sigma^2 I_N$ se tiene

$$E\left\{\left(\hat{\theta}_N^{LS} - \theta_0\right)\left(\hat{\theta}_N^{LS} - \theta_0\right)^T\right\} = \sigma^2\left(\Phi\Phi^T\right)^{-1}$$

3. Se tenía que el criterio cuadrático en los errores de predicción estaba definido como

$$V_N(\theta) = \frac{1}{N}\varepsilon_N^T(\theta)\varepsilon_N(\theta)$$

donde $\varepsilon_N(\theta) = Y_N - \Phi^T\theta \rightarrow E_N(\theta) = \Phi^T(\theta_0 - \theta) + \nu_N$, entonces

$$V_N = \frac{1}{N}\left(Y_N - \Phi^T\theta\right)^T\left(Y_N - \Phi^T\theta\right) = \frac{1}{N}\left[Y_N^TY_N - \theta^T\Phi Y_N - Y_N^T\Phi^T\theta + \theta^T\Phi\Phi^T\theta\right] =$$

$$= \frac{1}{N}\left[Y_N^TY_N - Y_N^T\Phi^T\left(\Phi\Phi^T\right)^{-1}\Phi Y_N - Y_N^T\Phi^T\left(\Phi\Phi^T\right)^{-1}\Phi Y_N + \cdots\right.$$

$$\left.\cdots + Y_N^T\Phi^T\left(\Phi\Phi^T\right)^{-1}\left(\Phi\Phi^T\right)\left(\Phi\Phi^T\right)^{-1}\Phi Y_N\right] =$$

$$= \frac{1}{N}\left[Y_N^TY_N - Y_N^T\Phi^T\left(\Phi\Phi^T\right)^{-1}\Phi Y_N\right]$$

considerando que $Y_N = \Phi^T \theta_0 + \nu_N$

$$V_N = \frac{1}{N}\left[\left(\Phi^T\theta_0 + \nu_N\right)^T \left(\Phi^T\theta_0 + \nu_N\right) - \left(\Phi^T\theta_0 + \nu_N\right)^T \Phi^T \left(\Phi\Phi^T\right)^{-1} \Phi\left(\Phi^T\theta_0 + \nu_N\right) \right] =$$

$$= \frac{1}{N}\Big[\theta_0^T\Phi\Phi^T\theta_0 + \theta_0^T\Phi\,\nu_N + \nu_N^T\Phi^T\theta_0 + \nu_N^T\nu_N - \theta_0^T\Phi\Phi^T\theta_0 - \theta_0^T\Phi\,\nu_N - \cdots$$

$$\cdots - \nu_N^T\Phi^T\theta_0 - \nu_N^T\Phi^T\left(\Phi\Phi^T\right)^{-1}\Phi\,\nu_N \Big]$$

calculando la esperanza

$$E\left\{V_N\left(\hat{\theta}_N^{LS}\right)\right\} = \frac{1}{N}E\left\{\nu_N^T\nu_N\right\} - \frac{1}{N}E\left\{\nu_N^T\Phi^T\left(\Phi\Phi^T\right)^{-1}\Phi\,\nu_N\right\} \tag{3.2}$$

donde

$$E\left\{\nu_N^T\nu_N\right\} = N\sigma^2 \tag{3.3}$$

y

$$E\left\{\nu_N^T\Phi^T\left(\Phi\Phi^T\right)^{-1}\Phi\,\nu_N\right\} = E\left\{\mathrm{Tr}\left(\nu_N^T\Phi^T\left(\Phi\Phi^T\right)^{-1}\Phi\,\nu_N\right)\right\}$$

y como $\mathrm{Tr}(AB) = \mathrm{Tr}(BA)$

$$E\left\{\nu_N^T\Phi^T\left(\Phi\Phi^T\right)^{-1}\Phi\,\nu_N\right\} = E\left\{\mathrm{Tr}\left(\nu_N\nu_N^T\Phi^T\left(\Phi\Phi^T\right)^{-1}\Phi\right)\right\}$$

introduciendo la esperanza dentro de la traza

$$\begin{aligned} E\left\{\nu_N^T\Phi^T\left(\Phi\Phi^T\right)^{-1}\Phi\,\nu_N\right\} &= \mathrm{Tr}\left(E\left\{\nu_N\nu_N^T\right\}\Phi^T\left(\Phi\Phi^T\right)^{-1}\Phi\right) = \\ &= \mathrm{Tr}\left(\sigma^2\Phi^T\left(\Phi\Phi^T\right)^{-1}\Phi\right) = \mathrm{Tr}\left(\sigma^2 I_p\right) = \sigma^2 p \end{aligned} \tag{3.4}$$

Introduciendo (3.3) y (3.4) en (3.2) resulta

$$E\left\{V_N\left(\hat{\theta}_N^{LS}\right)\right\} = \frac{1}{N}\left(N\sigma^2 - \sigma^2 p\right) = \sigma^2\left(1 - \frac{p}{N}\right) = \left(\frac{N-p}{N}\right)\sigma^2$$

Es decir que una estima no desviada de σ^2 es $\hat{\sigma}_N^2 = V_N\left(\hat{\theta}_N^{LS}\right)\dfrac{N}{N-p}$ ya que

$$E\left\{\hat{\sigma}_N^2\right\} = E\left\{V_N\left(\hat{\theta}_N^{LS}\right)\right\}\frac{N}{N-p} = \sigma^2$$

□

4. Apéndice (Definiciones)

Procesos (estocásticos) estacionarios

Un proceso aleatorio $\{\nu(n)\}$ se dice "débilmente" estacionario (o estacionario hasta momentos de segundo orden) si:

$$m_\nu \overset{\Delta}{=} E\{\nu(n)\} = \text{constante}$$

$$R_\nu(\tau) \overset{\Delta}{=} E\left\{(\nu(n) - m_\nu)(\nu(n-\tau) - m_\nu)^T\right\}$$ depende sólo de la diferencia temporal τ

Procesos cuasi-estacionarios

Un proceso $\{u(k)\}$ se dice cuasi-estacionario si los siguientes límites definiendo la media m_u y covarianza $R_u(\tau)$ existen

$$m_u \overset{\Delta}{=} \lim_{N\to\infty} \frac{1}{N}\sum_{k=1}^{N} E\{u(k)\} = \text{constante}$$

$$R_u(\tau) \overset{\Delta}{=} \lim_{N\to\infty} \frac{1}{N}\sum_{k=1}^{N} E\left\{(u(k) - m_u)(u(k-\tau) - m_u)^T\right\}$$ depende sólo de la diferencia temporal τ

Por simplicidad suele usarse la notación

$$\overline{E}\{\bullet\} \overset{\Delta}{=} \lim_{N\to\infty} \frac{1}{N}\sum_{k=1}^{N} E\{\bullet\}$$

Ergodicidad

Un proceso aleatorio se dice ergódico (hasta momento de segundo orden) si

$$\frac{1}{N}\sum_{k=1}^{N} \nu(n) \overset{a.s.}{\rightarrow} E\{\nu(n)\}$$

$$\frac{1}{N}\sum_{k=1}^{N} \nu(k+\tau)\nu(k) \overset{a.s.}{\rightarrow} E\{\nu(k+\tau)\nu(k)\}$$

para $N\rightarrow\infty$.

Estimador Consistente

Una estima $\hat{\theta}_N$ se dice **consistente** si se verifica

$$\hat{\theta}_N \xrightarrow{as} \theta_0 \quad \text{para } N \rightarrow \infty$$

Estimador No Desviado (unbiased)

Una estima $\hat{\theta}_N$ del vector de parámetros θ_0 se dice **no desviada** si se verifica

$$E\{\hat{\theta}_N\} = \theta_0$$

En caso que no se verifique esta condición, a la diferencia $E\{\hat{\theta}_N\} - \theta_0$ se la denomina **desvio** de la estima.

5. Referencias

[1] Ljung, L. *System Identification: Theory for the User*, second edition, Prentice Hall, N.J., 1999.
[2] Söderström, T. and Stoica, P.. *System Identification* , Prentice Hall, N.J., 1989.