

Métodos de Error de Predicción (PEM)

- Se selecciona una estructura de modelo \mathbf{M} , con modelos particulares $\mathbf{M}(\theta)$ parametrizados con un vector de parámetros $\theta \in \mathbf{D}_{\mathbf{M}} \subset \mathfrak{R}^p$

$$\mathbf{M} = \{\mathbf{M}(\theta) | \theta \in \mathbf{D}_{\mathbf{M}}\}$$

- Se disponen para la estimación de N pares de datos de entrada-salida

$$Z^N = \{y(n), u(n) : n = 1, \dots, N\}$$

- Cada modelo representa una forma de predecir las salidas futuras. Denotamos con $\hat{y}(n | \theta)$ a un **predictor** de la salida $y(n)$ dados los datos de entrada-salida hasta el instante $n - 1$, y basado en el vector de parámetros θ .

Predictor \Leftrightarrow Filtro Lineal o No Lineal aplicado a los datos

Ejemplos

- **Predictor Lineal General**

$$\hat{y}(n | \theta) = F_1(q, \theta)y(n) + F_2(q, \theta)u(n)$$

donde $F_1(q, \theta), F_2(q, \theta)$ son filtros lineales tal que $\hat{y}(n | \theta)$ depende sólo de datos pasados.

□ **One-Step-ahead Predictor**

Para un sistema con un modelo

$$y(n) = G(q, \theta)u(n) + H(q, \theta)e(n)$$

un predictor lineal típico es

$$\hat{y}(n | \theta) = [1 - H^{-1}(q, \theta)]y(n) + H^{-1}(q, \theta)G(q, \theta)u(n)$$

denominado **Predictor One-Step-Ahead**

Veremos más adelante que éste es un **predictor óptimo**, si $e(n)$ es ruido blanco.

- El problema que se nos plantea es cómo usar la información contenida en los datos Z^N para seleccionar un valor apropiado del vector de parámetros $\hat{\theta}_N$, y por lo tanto un elemento particular $\mathbf{M}(\hat{\theta}_N)$ en la clase de modelos \mathbf{M} .

En otras palabras se debe determinar un mapeo de los datos Z^N al conjunto \mathbf{D}_M :

$$Z^N \rightarrow \hat{\theta}_N \in \mathbf{D}_M$$

Este mapeo es el **Método de Estimación de Parámetros**.

- **Evaluación de los modelos candidatos:** se busca una forma de evaluar la **habilidad** de los distintos modelos para describir los datos observados. Como la esencia del modelo es su característica como predictor, definimos el **error de predicción** de un dado modelo $\mathbf{M}(\theta_*)$ como

$$\varepsilon(n, \theta_*) = y(n) - \hat{y}(n | \theta_*).$$

Un “**buen**” modelo será entonces aquel que sea un buen predictor, es decir aquel que produzca errores de predicción pequeños cuando es aplicado a los datos.

- **Minimización de los errores de predicción**

La secuencia de los errores de predicción $\{\varepsilon(n, \theta)\}_{n=1}^N$ puede considerarse como un vector en \mathfrak{R}^N por lo que las dimensiones del error podrían medirse usando alguna norma en \mathfrak{R}^N . Típicamente esta norma, o **función de costo**, o **criterio** es de la forma

$$V_N(\theta) = \frac{1}{N} \sum_{n=1}^N \ell(\varepsilon(n, \theta))$$

donde $\ell(\bullet)$ es una función escalar a valores reales, típicamente positiva. La estima $\hat{\theta}_N$ se obtiene minimizando el criterio $V_N(\theta)$, i.e.

$$\hat{\theta}_N = \underset{\theta \in \mathbf{D}_M}{\operatorname{argmin}} V_N(\theta)$$

Si el mínimo no es único, entonces **argmin** denota al conjunto de argumentos minimizantes del criterio.

Ejemplos de criterios

□ **Criterio cuadrático**

$$V_N(\theta) = \frac{1}{N} \sum_{n=1}^N \frac{1}{2} \varepsilon(n, \theta)^T \varepsilon(n, \theta),$$

donde $\ell(\varepsilon) = \frac{1}{2} \varepsilon(n, \theta)^T \varepsilon(n, \theta)$.

- **Normas dependientes del tiempo:** A veces las mediciones en distintos instantes de tiempo pueden tener distinta confiabilidad. En esos casos es conveniente usar un criterio que depende explícitamente del tiempo, que permita ponderar con distinto peso los datos. Por ejemplo,

$$V_N(\theta, n) = \frac{1}{N} \sum_{n=1}^N \ell(\varepsilon(n, \theta), n),$$

o a veces se prefiere hacer explícita la ponderación con una función de ponderación $\beta(n)$, resultando

$$V_N(\theta, n) = \frac{1}{N} \sum_{n=1}^N \beta(n) \ell(\varepsilon(n, \theta)).$$

- Los **Métodos de Error de Predicción (PEM)** pueden resumirse en:

Métodos de Error de Predicción

- Elección de la Estructura de Modelo \mathbf{M} , parametrizada por el parámetro θ .
- Selección de la estructura del predictor $\hat{y}(n | \theta)$.
- Cómputo del EP $\varepsilon(n, \theta) = y(n) - \hat{y}(n | \theta)$.
- Selección del criterio $V_N(\theta)$ (o $V_N(\theta, n)$), a minimizar.
- Obtención de la estima $\hat{\theta}_N$ por minimización del criterio (implica la elección del método de minimización).

• Análisis Asintótico de las estimas

Es de interés determinar las propiedades de las estimas de parámetros $\hat{\theta}_N$ cuando el número de datos N tiende a infinito.

□ Hipótesis

1. Los datos $\{u(n), y(n)\}$ son procesos estacionarios.
2. La entrada es una **excitación persistente** (se definirá precisamente más adelante en el Curso).
3. El Hessiano $V_N''(\theta)$ es no singular localmente alrededor de los mínimos del criterio $V_N(\theta)$.
4. Las (matrices) transferencias $G(q, \theta)$ y $H(q, \theta)$ son funciones diferenciables del vector de parámetros θ .

Algunos de los resultados requerirán también la hipótesis

5. El conjunto

$$\mathbf{D}_T = \{\theta : G(q, \theta) = G(q); H(q, \theta) = H(q)\}$$

consistente de aquellos valores de los parámetros para los cuales el modelo provee una descripción exacta del sistema real consta de exactamente un punto, que denotaremos θ_0 .

□ **Estima Asintótica**

Sea

$$\hat{\theta}_N = \operatorname{argmin}_{\theta \in \mathbf{D}_M} V_N(\theta)$$

y suponga que las hipótesis 1. a 4. se verifican. Entonces el criterio $V_N(\theta)$ converge uniformemente en $\theta \in \mathbf{D}_M$ a la función límite $V_\infty(\theta)$, i.e.

$$\sup_{\theta \in \mathbf{D}_M} |V_N(\theta) - V_\infty(\theta)| \xrightarrow{\text{as}} 0 \quad \text{cuando } N \rightarrow \infty,$$

donde $V_\infty(\theta) = \lim_{N \rightarrow \infty} V_N(\theta)$. Además

$$\hat{\theta}_N \xrightarrow{\text{as}} \theta_* \quad \text{cuando } N \rightarrow \infty,$$

donde $\theta_* \in \mathbf{D}_C$, siendo \mathbf{D}_C el conjunto de valores del parámetro que minimiza el criterio límite $V_\infty(\theta)$.

Este resultado implica que cuando el conjunto \mathbf{D}_T es vacío, entonces la estima asintótica será 'biased', pero será la mejor aproximación posible del sistema que puede obtenerse con la estructura de modelo. Si el sistema es **parámetro identificable** (se verifica la condición 5.), el conjunto \mathbf{D}_T es no vacío y consiste de un único elemento. Bajo ciertas hipótesis no muy restrictivas sobre el conjunto de datos es posible mostrar que en este caso $\mathbf{D}_T = \mathbf{D}_C = \{\theta_0\}$, y la estima es **consistente en sentido fuerte** (strongly consistent).

□ **Distribución asintótica de las estimas**

Asumiendo que se verifican las hipótesis 1. a 4., puede probarse que la distribución de la variable aleatoria

$$\sqrt{N}(\hat{\theta}_N - \theta_*)$$

converge a una distribución Gaussiana con media cero y matriz de covarianza

$$P = [V''_{\infty}(\theta_*)]^{-1} \left[\lim_{N \rightarrow \infty} N \mathbf{E} \{ V'_N(\theta_*)^T V'_N(\theta_*) \} \right] [V''_{\infty}(\theta_*)]^{-1}$$

donde $V'_N(\theta_*)$ es el gradiente de $V_N(\theta)$ calculado en $\theta = \theta_*$, y $V''_{\infty}(\theta_*)$ es el Hessiano de $V_{\infty}(\theta)$ calculado en $\theta = \theta_*$. Es decir

$$\sqrt{N}(\hat{\theta}_N - \theta_*) \xrightarrow{\text{dist}} N(0, P)$$

Este resultado es importante porque da una expresión de matriz de covarianza asintótica que puede ser usada para cuantificar el error de estimación. La expresión de la covarianza asintótica puede también usarse para computar intervalos de confianza para cada estima particular $\hat{\theta}_N$ obtenida de los datos. Como en general la estima asintótica θ_* no se conoce, la matriz de covarianza P puede aproximarse reemplazando θ_* por $\hat{\theta}_N$ y el operador esperanza matemática por la media muestral.