

# Estimación y Predicción en Series Temporales

Estimadores MVU, caso general

Departamento de Procesamiento de Señales

Instituto de Ingeniería Eléctrica  
Facultad de Ingeniería

2022

## Estimadores MVU caso General

- Cómo podemos hacer para encontrar, si existe, un MVU cuando éste no es eficiente (sino bastaría con el teorema de la CRLB).
- Para eso vamos a definir la noción de **estadístico suficiente**.
- Veremos como evaluar si un estadístico suficiente insesgado es un MVU (no necesariamente eficiente).
- El enfoque anterior es limitante (se requiere un estimador candidato). Veremos si existen estadísticos suficientes, y cómo encontrarlos usando el **teorema de Neyman-Fisher**.
- Veremos como usar estadísticos suficientes para encontrar un MVU (si existe). Este es el **teorema de Rao-Blackwell-Lehmann-Scheffe**.

- Vimos previamente que la evaluación de la CRLB puede resultar en la determinación de un estimador eficiente (y por lo tanto un MVU).
- Un caso especial donde sucede esto son los modelos lineales.
- ¿Qué pasa cuando no existe un estimador eficiente? Aún así puede llegar a existir el estimador MVU, y lo querríamos encontrar.
- **Estadístico suficiente**: una función de los datos que contiene **toda** la información disponible sobre éstos para realizar la estimación.

# Estadísticos suficientes

**Ejemplo:** DC en WGN de  $N$  muestras de varianza  $\sigma^2$

Vimos que  $\hat{A} = \frac{1}{N} \sum_0^{N-1} x[n]$  es el estimador MVU,  $\text{var}[\hat{A}] = \sigma^2/N$ .

Consideremos el estimador  $\check{A} = x[0]$ :

- Es insesgado
- $\text{var}(\check{A}) = \sigma^2 > \sigma^2/N$ . La performance muy pobre ¿Porqué?

No se utilizan datos que aportan información útil (muestras independientes)

## Preguntas:

- ¿Cuál es el conjunto de datos pertinentes para la estimación, i.e., que logran producir una estimación de mínima varianza?
- ¿Qué conjunto o conjuntos son **suficientes** para calcular  $\hat{A}$ ?

- Supongamos los siguientes conjuntos:

$$S_A = \{x[0], x[1], x[2], \dots, x[N-1]\}$$

$$S_B = \{x[0] + x[1], x[2], x[3], \dots, x[N-1]\}$$

$$S_C = \left\{ \sum_{n=0}^{N-1} x[n] \right\}$$

- Son los tres conjuntos suficientes para realizar la estimación?  
Si.
- ... pero  $S_C$  es el más compacto posible:  
 $\#S_C < \min(\#S_A, \#S_B)$
- Se dice que  $\sum_{n=0}^{N-1} x[n]$  es el **estadístico suficiente minimal**.
- Para estimar  $A$ , una vez conocido  $\sum_{n=0}^{N-1} x[n]$  no se precisa nada más ya que este estadístico **resume toda la información disponible**.

# Verificación de que un estadístico dado es suficiente

¿Cómo verificar que un estadístico  $T(\mathbf{x})$  es suficiente?

Verificando que una vez conocido su valor  $T(\mathbf{x}) = T_0$ , no queda información adicional para mejorar la estimación de  $A$ .

Esto sucede cuando  $p(\mathbf{x}|T(\mathbf{x}) = T_0; A)$  es **independiente de  $A$** .

**Ejemplo:** DC en WGN de  $N$  muestras de varianza  $\sigma^2$

- Consideremos el estadístico  $T(\mathbf{x}) = \sum_{n=0}^{N-1} x[n]$
- Usando la Regla del Producto:

$$p(\mathbf{x}|T(\mathbf{x})=T_0; A) = \frac{p(\mathbf{x}, T(\mathbf{x})=T_0; A)}{p(T(\mathbf{x})=T_0; A)} = \frac{p(\mathbf{x}; A)\delta(T(\mathbf{x}) - T_0)}{p(T(\mathbf{x})=T_0; A)}.$$

# Verificación de que un estadístico dado es suficiente

Se tiene que  $T(\mathbf{x}) \sim \mathcal{N}(NA, N\sigma^2)$ .

$$\begin{aligned} p(\mathbf{x}; A) \delta(T(\mathbf{x}) - T_0) &= (2\pi\sigma^2)^{-\frac{N}{2}} \exp \left[ -\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x[n] - A)^2 \right] \delta(T(\mathbf{x}) - T_0) \\ &= (2\pi\sigma^2)^{-\frac{N}{2}} \exp \left[ -\frac{1}{2\sigma^2} \left( \sum_{n=0}^{N-1} x^2[n] - 2AT(\mathbf{x}) + NA^2 \right) \right] \delta(T(\mathbf{x}) - T_0) \\ &= (2\pi\sigma^2)^{-\frac{N}{2}} \exp \left[ -\frac{1}{2\sigma^2} \left( \sum_{n=0}^{N-1} x^2[n] - 2AT_0 + NA^2 \right) \right] \delta(T(\mathbf{x}) - T_0). \end{aligned}$$

Por lo que,

$$\begin{aligned} p(\mathbf{x} | T(\mathbf{x}) = T_0; A) &= \frac{(2\pi\sigma^2)^{-\frac{N}{2}} \exp \left[ -\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} x^2[n] \right] \exp \left[ -\frac{1}{2\sigma^2} (-2AT_0 + NA^2) \right]}{\frac{1}{\sqrt{2\pi N\sigma^2}} \exp \left[ -\frac{1}{2N\sigma^2} (T_0 - NA)^2 \right]} \delta(T(\mathbf{x}) - T_0) \\ &= \frac{\sqrt{N}}{(2\pi\sigma^2)^{(N-1)/2}} \exp \left[ -\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} x[n]^2 \right] \exp \left[ \frac{T_0^2}{2N\sigma^2} \right] \delta(T(\mathbf{x}) - T_0). \end{aligned}$$

- Como no depende de  $A$ , se concluye que  $T(\mathbf{x}) = \sum_{n=0}^{N-1} x[n]$  es un estimador suficiente.
- **Limitante.** Primero hay que tener el estadístico, para luego verificar que es suficiente.



# Teorema de factorización de Neyman-Fisher

**Teorema Factorización de Neyman-Fisher.** Si se puede factorizar  $p(\mathbf{x}; \theta)$  como

$$p(\mathbf{x}; \theta) = g(T(\mathbf{x}), \theta) h(\mathbf{x}), \quad (1)$$

dónde  $g$  solo depende de  $\mathbf{x}$  a través de  $T(\mathbf{x})$  y  $h$  es función únicamente de  $\mathbf{x}$ ; entonces  $T(\mathbf{x})$  es un estadístico suficiente (E.S.) de  $\theta$ .

Además, si  $T(\mathbf{x})$  es un estadístico suficiente de  $\theta$  entonces  $p(\mathbf{x}; \theta)$  puede ser factorizada como (1).

**Prueba:** Ver [Kay 1993], Apéndice 5A

**Ejemplo:** DC en WGN de  $N$  muestras de varianza  $\sigma^2$

Veamos si podemos factorizar  $p(\mathbf{x}; A)$ :

$$p(\mathbf{x}; A) = \underbrace{\frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} x[n]^2\right)}_{h(\mathbf{x})} \cdot \underbrace{\exp\left(-\frac{1}{2\sigma^2} (-2A \sum_{n=0}^{N-1} x[n] + NA^2)\right)}_{g(T(\mathbf{x}), A)}.$$
$$\Rightarrow T(\mathbf{x}) = \sum_{n=0}^{N-1} x[n] \text{ es un E.S. de } A.$$

**OBS:**  $T_2(\mathbf{x}) = 2T(\mathbf{x})$  es también un E.S. de  $A$ . Se puede ver que cualquier transformación biyectiva de un E.S. es también un E.S.

## Estadístico Completo.

- Un estadístico  $T(\mathbf{x})$  es completo si existe una única función  $g$  tal que  $g(T(\mathbf{x}))$  es insesgado.
- Un estadístico suficiente es completo: si es suficiente y es completo.

## Proposición

Un estadístico suficiente  $T(\mathbf{x})$  es completo si y solo si:

$$\int_{-\infty}^{+\infty} v(T)p(T;\theta)dT = 0 \quad \forall \theta,$$

se cumple únicamente para la función  $v(T) = 0$  para todo  $T$ .

**Ejercicio:** pensar la demostración.

## **Teorema de Rao-Blackwell-Lehemann-Scheffe.**

Sea  $\check{\theta}$  un estimador insesgado de  $\theta$  y  $T(\mathbf{x})$  un estadístico suficiente de  $\theta$ .

Entonces:

- $\hat{\theta} := \mathbb{E}[\check{\theta}|T(\mathbf{x})]$  es un estimador insesgado de  $\theta$ , y  $\text{var}(\hat{\theta}) \leq \text{var}(\check{\theta}) \forall \theta$ .
- Si además  $T(\mathbf{x})$  es completo,  $\hat{\theta}$  es el MVU de  $\theta$ .

**Prueba:** Ver [Kay 1993], Apéndice 5B

## **Observaciones:**

- Si  $T(\mathbf{x})$  no es completo, no sabemos si  $\hat{\theta}$  es o no MVU.
- Reducción en varianza de  $\hat{\theta}$  con respecto a  $\check{\theta}$  es consecuencia de extraer toda la info disponible en los datos via el estadístico suficiente.

## Observaciones (cont.)

- Tenemos que,

$$\hat{\theta} := \mathbb{E}[\check{\theta}|T(\mathbf{x})] = \int \check{\theta} p(\check{\theta}|T(\mathbf{x})) d\check{\theta} = g(T(\mathbf{x})).$$

donde además  $\hat{\theta}$  es un estimador insesgado (RBLS).

- Si  $T$  es completo, entonces la  $g$  tal que el estimador  $\hat{\theta} = g(T(\mathbf{x}))$  es insesgado, es única, independientemente del  $\check{\theta}$  que elijamos.

$$\Rightarrow \forall \theta, \text{ var}(\hat{\theta}) \leq \text{var}(\check{\theta}) \quad \forall \check{\theta} \text{ insesgado (i.e. } \hat{\theta} \text{ es el MVU).}$$

- Por la observación anterior, si tenemos un E.S.  $T(\mathbf{x})$  y encontramos una  $g$  tal que  $g(T(\mathbf{x}))$  es insesgado, y verificamos que es única, entonces ya sabemos que  $\hat{\theta} = g(T(\mathbf{x}))$  es el MVU.

# Resumen: Procedimiento para encontrar el MVU

- 1 Aplicar el teorema de Neyman-Fisher para encontrar  $T(\mathbf{x})$  estadístico suficiente.
- 2 Determinar si  $T(\mathbf{x})$  es completo:
  - Tratando de encontrar  $g$  tal que  $g(T(\mathbf{x}))$  es insesgado y verifica ser única. Si es así,  $\hat{\theta} = g(T(\mathbf{x}))$  es el MVU (fin, terminamos).
  - Usando la proposición (nos dice si  $T$  es completo pero no da  $g$ ).
- 3 Calcular el MVU como

$$\hat{\theta} = \mathbb{E}[\check{\theta}|T(\mathbf{x})],$$

con  $\check{\theta}$  cualquier estimador insesgado.

# Ejemplo con media y varianza a estimar

**Ejemplo:** Nivel de DC en ruido blanco de varianza desconocida

$$x[n] = A + w[n], \quad n = 0, 1, \dots, N-1,$$

con  $w[n] \sim \mathcal{N}(0, \sigma^2)$  i.i.d. Tanto  $A$  como  $\sigma^2$  son desconocidos y deben ser estimados.

$$p(\mathbf{x}; (A, \sigma^2)) = \underbrace{\frac{1}{(2\pi\sigma^2)^{N/2}} \exp \left( -\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} x[n]^2 + \frac{2A}{2\sigma^2} \sum_{n=0}^{N-1} x[n] - \frac{NA^2}{2\sigma^2} \right)}_{f([T_1(\mathbf{x}), T_2(\mathbf{x})], [A, \sigma^2])} \cdot \underbrace{1}_{h(\mathbf{x})}$$

donde

$$\mathbf{T}(\mathbf{x}) = \begin{bmatrix} T_1(\mathbf{x}) \\ T_2(\mathbf{x}) \end{bmatrix} = \begin{bmatrix} \sum_{n=0}^{N-1} x[n], \sum_{n=0}^{N-1} x^2[n] \end{bmatrix}^T$$

es un E.S. de acuerdo al teorema de factorización de Neyman-Fisher.

A continuación buscaremos una función  $g$  que haga que  $g(\mathbf{T}(\mathbf{x}))$  sea insesgado,

$$\mathbb{E}[g(\mathbf{T}(\mathbf{x}))] = \begin{bmatrix} A \\ \sigma^2 \end{bmatrix}.$$

# Ejemplo con media y varianza desconocidas

Primero observar que

$$\mathbb{E}[\mathbf{T}(\mathbf{x})] = \begin{bmatrix} \mathbb{E}[T_1(\mathbf{x})] \\ \mathbb{E}[T_2(\mathbf{x})] \end{bmatrix} = \begin{bmatrix} \sum_{n=0}^{N-1} \mathbb{E}[x[n]] \\ \sum_{n=0}^{N-1} \mathbb{E}[x^2[n]] \end{bmatrix} = \begin{bmatrix} NA \\ N(\sigma^2 + A^2) \end{bmatrix}.$$

Una función  $g$  candidata natural es la siguiente

$$\mathbf{g}(\mathbf{T}(\mathbf{x})) = \begin{bmatrix} \frac{1}{N} T_1(\mathbf{x}) \\ \frac{1}{N} T_2(\mathbf{x}) - \left[ \frac{1}{N} T_1(\mathbf{x}) \right]^2 \end{bmatrix} = \begin{bmatrix} \bar{x} \\ \frac{1}{N} \sum_{n=0}^{N-1} x^2[n] - \bar{x}^2 \end{bmatrix}$$

Por un lado,

$$\mathbb{E}(\bar{x}) = A,$$

y por otro,

$$\mathbb{E} \left( \frac{1}{N} \sum_{n=0}^{N-1} x^2[n] - \bar{x}^2 \right) = \sigma^2 + A^2 - \mathbb{E}(\bar{x}^2).$$

# Ejemplo con media y varianza desconocidas

Por otro lado sabemos que  $\bar{x} \sim \mathcal{N}(A, \sigma^2/N)$ , con lo cual,  $\mathbb{E}(\bar{x}^2) = A^2 + \frac{\sigma^2}{N}$ .  
Por lo que

$$\mathbb{E} \left( \frac{1}{N} \sum_{n=0}^{N-1} x^2[n] - \bar{x}^2 \right) = \sigma^2 + A^2 - \left( A^2 + \frac{\sigma^2}{N} \right) = \frac{N-1}{N} \sigma^2.$$

Si multiplicamos este estadístico por  $N/(N-1)$  entonces obtenemos un estimador insesgado de  $\sigma^2$ ,

$$\mathbf{g}(\mathbf{T}(\mathbf{x})) = \begin{bmatrix} \frac{1}{N} T_1(\mathbf{x}) \\ \frac{1}{N-1} \left[ T_2(\mathbf{x}) - N \left( \frac{1}{N} T_1(\mathbf{x}) \right)^2 \right] \end{bmatrix} = \begin{bmatrix} \bar{x} \\ \frac{1}{N-1} \left[ \sum_{n=0}^{N-1} x^2[n] - N \bar{x}^2 \right] \end{bmatrix}$$

Observar que esto puede ser escrito como

$$\hat{\theta} = \mathbf{g}(\mathbf{T}(\mathbf{x})) = \begin{bmatrix} \bar{x} \\ \frac{1}{N-1} \sum_{n=0}^{N-1} (x[n] - \bar{x})^2 \end{bmatrix},$$

que es por lo tanto el estimador MVU de  $\theta = [A, \sigma^2]^T$ .

**Ejercicio.** Probar que  $\hat{\theta}$  no es eficiente (sugerencia, calcular la CRLB y mostrar que la varianza del estimador MVU es estrictamente superior).



- **Kay, S. M.** (1993)  
*Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory*, Capítulo 5.