

Estimación y Predicción en Series Temporales

Estimadores MVU, caso lineal

Departamento de Procesamiento de Señales

Instituto de Ingeniería Eléctrica
Facultad de Ingeniería

2022

- 1 Repaso Cota inferior de Cramér-Rao (CRLB)
- 2 Estimadores MVU en modelos lineales
- 3 Estimadores MVU caso General

Cota Inferior de Cramér-Rao

Teorema: Cota Inferior de Cramér-Rao, parámetro escalar.

Se asume que la PDF $p(\mathbf{x}; \theta)$ satisface la condición de regularidad,

$$\mathbb{E}_{\mathbf{x}} \left[\frac{\partial \log p(\mathbf{x}; \theta)}{\partial \theta} \right] = 0 \quad \text{para todo } \theta.$$

Entonces,

- 1 la varianza de todo estimador insesgado $\hat{\theta}$ cumple que

$$\text{var}(\hat{\theta}) \geq \frac{1}{-\mathbb{E}_{\mathbf{x}} \left[\frac{\partial^2 \log p(\mathbf{x}; \theta)}{\partial \theta^2} \right]},$$

donde la derivada se evalúa en el valor verdadero de θ .

- 2 existe un estimador que alcanza la cota para todo θ si y solo si

$$\frac{\partial \log p(\mathbf{x}; \theta)}{\partial \theta} = I(\theta) (g(\mathbf{x}) - \theta),$$

para alguna función I y g .

Este estimador, que es el MVU, es $\hat{\theta} = g(\mathbf{x})$ y su varianza es $\frac{1}{I(\theta)}$.

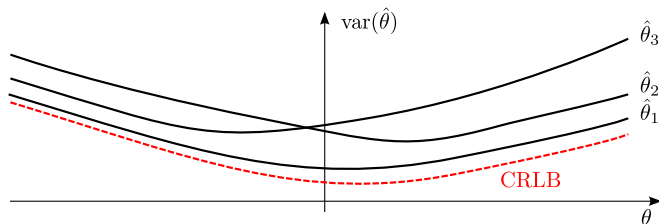
Cota Inferior de Cramér-Rao: Consideraciones

- La esperanza se toma respecto a los datos $\mathbf{x} \sim p(\mathbf{x}; \theta)$,

$$\mathbb{E}_{\mathbf{x}} \left[\frac{\partial^2 \log p(\mathbf{x}; \theta)}{\partial \theta^2} \right] = \int \frac{\partial^2 \log p(\mathbf{x}; \theta)}{\partial \theta^2} p(\mathbf{x}; \theta) d\mathbf{x}.$$

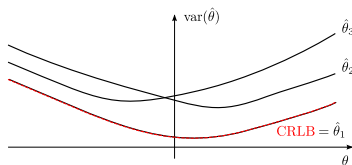
La esperanza reconoce el hecho de que la función de verosimilitud y sus derivadas son variables aleatorias por depender de los datos observados \mathbf{x} .

- La cota depende en general del parámetro desconocido θ .

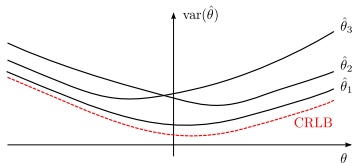


Definición (Estimador eficiente). Un estimador que es insesgado y alcanza la cota de Cramér-Rao para todos los valores del parámetro desconocido se dice que es **eficiente**.

Observación. Un estimador MVU puede ser o no ser eficiente



- $\hat{\theta}_1$ alcanza la CRLB y por lo tanto es el MVU.
- θ_1 es eficiente y MVU.



- Ningún estimador alcanza CRLB.
- Varianza de $\hat{\theta}_1$ es menor que la de los otros estimad. insesgados.
- θ_1 es MVU pero no es eficiente

CRLB general para señales con AWGN

- Sea una señal determinística con un parámetro desconocido θ observada en AWGN (ruido aditivo blanco Gaussiano),

$$x[n] = s[n; \theta] + w[n], \quad \text{con } n = 0, 1, \dots, N-1 \quad \text{y} \quad w[n] \sim \mathcal{N}(0, \sigma^2).$$

- La función de verosimilitud es

$$p(\mathbf{x}; \theta) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp \left[-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x[n] - s[n; \theta])^2 \right].$$

$$\begin{aligned} \mathbb{E}_{\mathbf{x}} \left(\frac{\partial^2 \log p(\mathbf{x}; \theta)}{\partial \theta^2} \right) &= \frac{1}{\sigma^2} \sum_{n=0}^{N-1} \left[(\mathbb{E}(x[n]) - s[n; \theta]) \frac{\partial^2 s[n; \theta]}{\partial \theta^2} - \left(\frac{\partial s[n; \theta]}{\partial \theta} \right)^2 \right] \\ &= -\frac{1}{\sigma^2} \sum_{n=0}^{N-1} \left(\frac{\partial s[n; \theta]}{\partial \theta} \right)^2. \end{aligned}$$

- La CRLB es por lo tanto:

$$\text{var}(\hat{\theta}) \geq \frac{\sigma^2}{\sum_{n=0}^{N-1} \left(\frac{\partial s[n; \theta]}{\partial \theta} \right)^2}$$

- Mejor estimación si la señal $s[n; \theta]$ cambia rápidamente con parámetro θ .

Transformación de parámetros: caso asintótico

La eficiencia es *aproximadamente* mantenida bajo transformaciones no afines si la cantidad de observaciones es suficientemente grande.

- Un estimador $\hat{\theta}$ del parámetro θ es **asintóticamente insesgado** si

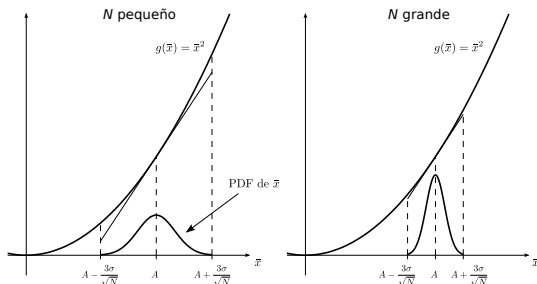
$$\lim_{N \rightarrow \infty} \mathbb{E}(\hat{\theta}) = \theta.$$

- Un estimador $\hat{\theta}$ del parámetro θ es **asintóticamente eficiente** si

$$\lim_{N \rightarrow \infty} \text{var}(\hat{\theta}) = CRLB(\theta).$$

Transformación de parámetros: Resumen

- Una transformación afín de un estimador eficiente mantiene la eficiencia.
- Es decir el estimador transformado es un estimador *eficiente* del parámetro transformado.
- Una transformación no afín de un estimador eficiente destruye la eficiencia, e incluso puede hacerlo sesgado.
- Sin embargo, estimador transformado es asint. insesgado y asint. eficiente.
- Es decir, cuando la cantidad de observaciones $N \rightarrow \infty$, el estimador límite es insesgado y eficiente.



CRLB: Extensión a vector de parámetros

- Supongamos que queremos estimar vector de parámetros $\boldsymbol{\theta} = [\theta_1, \theta_2, \dots, \theta_p]^T$.
- Si asumimos que tenemos un estimador $\hat{\boldsymbol{\theta}}$ que es insesgado, la CRLB para un vector de parámetros establece una cota en la varianza de cada elemento,

$$\text{var}(\hat{\theta}_i) \geq [\mathbf{I}^{-1}(\boldsymbol{\theta})]_{ii},$$

donde $\mathbf{I}(\boldsymbol{\theta})$ es la **matriz de información de Fisher** de tamaño $p \times p$.

- La matriz de información de Fisher se define como

$$[\mathbf{I}(\boldsymbol{\theta})]_{ij} = -\mathbb{E} \left[\frac{\partial^2 \log p(\mathbf{x}; \boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} \right], \quad \text{con } (i, j) \in \{1, 2, \dots, p\}^2,$$

en donde al evaluar la ecuación se debe emplear el valor verdadero de $\boldsymbol{\theta}$.

- En el caso escalar $p = 1$, se tiene $\mathbf{I}(\boldsymbol{\theta}) = I(\theta)$ tal como se definió previamente (CRLB escalar).

Cota Inferior de Cramér-Rao (vectorial)

Teorema: Cota Inferior de Cramér-Rao, parámetro vectorial.

Se asume que la PDF $p(\mathbf{x}; \boldsymbol{\theta})$ satisface la condición de regularidad,

$$\mathbb{E}_{\mathbf{x}} \left[\frac{\partial \log p(\mathbf{x}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right] = \mathbf{0} \quad \text{para todo } \boldsymbol{\theta}.$$

Entonces,

- 1 la matriz de covarianza de todo estimador insesgado $\hat{\boldsymbol{\theta}}$ cumple que

$$\mathbf{C}_{\hat{\boldsymbol{\theta}}} - \mathbf{I}^{-1}(\boldsymbol{\theta}) \geq \mathbf{0}, \quad \text{donde } [\mathbf{I}(\boldsymbol{\theta})]_{ij} = -\mathbb{E}_{\mathbf{x}} \left[\frac{\partial^2 \log p(\mathbf{x}; \boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} \right],$$

- $\mathbf{I}(\boldsymbol{\theta})$ es la matriz de información de Fisher,
 - “ \geq ” se interpreta en el sentido de matriz semidefinida positiva,
 - derivada se evalúa en el valor verdadero de $\boldsymbol{\theta}$.
- 2 existe un estimador que alcanza la cota para todo $\boldsymbol{\theta}$ si y solo si

$$\frac{\partial \log p(\mathbf{x}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathbf{I}(\boldsymbol{\theta}) (\mathbf{g}(\mathbf{x}) - \boldsymbol{\theta}),$$

para alguna función $\mathbf{g} : \mathbb{R}^N \rightarrow \mathbb{R}^p$ y matriz \mathbf{I} de tamaño $p \times p$. Ese

Consecuencias.

- Como en una matriz semidefinida positiva todos los elementos de la diagonal son no negativos, la cota matricial implica,

$$[\mathbf{C}_{\hat{\boldsymbol{\theta}}} - \mathbf{I}^{-1}(\boldsymbol{\theta})]_{ii} \geq 0.$$

- Por lo tanto, la varianza de cada elemento del vector estimado cumple que,

$$\text{var}(\hat{\theta}_i) = [\mathbf{C}_{\hat{\boldsymbol{\theta}}}]_{ii} \geq [\mathbf{I}^{-1}(\boldsymbol{\theta})]_{ii}$$

- Si se cumple la condición de factorización, entonces la cota se alcanza y por lo tanto

$$\text{var}(\hat{\theta}_i) = [\mathbf{I}^{-1}(\boldsymbol{\theta})]_{ii}$$

En este caso, el estimado $\hat{\boldsymbol{\theta}} = \mathbf{g}(\mathbf{x})$ es eficiente y por lo tanto MVU.

Estimadores MVU en modelos lineales

- En muchas situaciones no es posible encontrar el estimador MVU con la técnica de Cramér-Rao

$$\frac{\partial \log p(\mathbf{x}; \theta)}{\partial \theta} = I(\theta)(g(\mathbf{x}) - \theta).$$

- Puede suceder que exista el MVU, pero ninguna técnica (regla) sirva para encontrarlo.

Modelos lineales

- En el caso en que los datos pueden representarse por un modelo lineal,

$$\mathbf{x} = \mathbf{h}\theta + \mathbf{w},$$

el estimador MVU siempre existe y se puede calcular.

Modelo lineal: Ejemplo

Ejemplo: Ajuste de recta

- Se considera el problema de ajustar una recta a partir de datos contaminados con AWGN,

$$x[n] = A + Bn + w[n], \quad \text{con } n = 0, 1, \dots, N-1,$$

donde $w[n] \sim \mathcal{N}(0, \sigma^2)$ para todo n y se quiere estimar la pendiente B y el coeficiente de intersección A .

- El modelo, se puede expresar en notación matricial

$$\mathbf{x} = \mathbf{h}\boldsymbol{\theta} + \mathbf{w}$$

donde,

$$\begin{aligned}\mathbf{x} &= [x[0], x[1], \dots, x[N-1]]^T \\ \mathbf{w} &= [w[0], w[1], \dots, w[N-1]]^T \\ \boldsymbol{\theta} &= [A, B]^T\end{aligned}\quad \mathbf{h} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ \vdots & \vdots \\ 1 & N-1 \end{bmatrix}$$

- El problema de estimación de la recta es efectivamente un problema lineal

Modelos lineal en ruido AWGN

$$\mathbf{x} = \mathbf{h}\boldsymbol{\theta} + \mathbf{w}$$

- $\mathbf{x} \in \mathbb{R}^N$: datos observados.
- $\boldsymbol{\theta} \in \mathbb{R}^p$: p parámetros desconocidos.
- $\mathbf{h} \in \mathbb{R}^{N \times p}$: matriz de observación, con $N > p$ y rango p .
- $\mathbf{w} \in \mathbb{R}^N$: ruido en observación. Se asume $\mathbf{w} \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$.

Nota: El espacio \mathbb{R} puede ser remplazado por \mathbb{C} .

Propiedades:

- En el caso de modelos lineales, siempre es posible encontrar el **estimador MVU** que además es **eficiente**.
- Según teorema de CRLB, debemos probar que:

$$\frac{\partial \log p(\mathbf{x}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathbf{I}(\boldsymbol{\theta})(\mathbf{g}(\mathbf{x}) - \boldsymbol{\theta}),$$

para alguna función $\mathbf{g} : \mathbb{R}^N \rightarrow \mathbb{R}^p$ y alguna matriz $\mathbf{I} \in \mathbb{R}^{p \times p}$

Definición y propiedades

- La función de verosimilitud es:

$$\begin{aligned} p(\mathbf{x}; \boldsymbol{\theta}) &= \prod_{n=0}^{N-1} \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{1}{2\sigma^2} (x[n] - (\mathbf{h}\boldsymbol{\theta})[n])^2 \right] \\ &= \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp \left[-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x[n] - (\mathbf{h}\boldsymbol{\theta})[n])^2 \right] \\ &= \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp \left[-\frac{1}{2\sigma^2} (\mathbf{x} - \mathbf{h}\boldsymbol{\theta})^T (\mathbf{x} - \mathbf{h}\boldsymbol{\theta}) \right] \end{aligned}$$

- Tomando el logaritmo se obtiene,

$$\log p(\mathbf{x}; \boldsymbol{\theta}) = -\log(2\pi\sigma^2)^{\frac{N}{2}} - \frac{1}{2\sigma^2} (\mathbf{x} - \mathbf{h}\boldsymbol{\theta})^T (\mathbf{x} - \mathbf{h}\boldsymbol{\theta})$$

Definición y propiedades

- Derivando respecto al parámetro desconocido,

$$\begin{aligned}\frac{\partial \log p(\mathbf{x}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} &= \frac{\partial}{\partial \boldsymbol{\theta}} \left[-\log(2\pi\sigma^2)^{\frac{N}{2}} - \frac{1}{2\sigma^2} (\mathbf{x} - \mathbf{h}\boldsymbol{\theta})^T (\mathbf{x} - \mathbf{h}\boldsymbol{\theta}) \right] \\ &= -\frac{1}{2\sigma^2} \frac{\partial}{\partial \boldsymbol{\theta}} \left[\mathbf{x}^T \mathbf{x} - 2\mathbf{x}^T \mathbf{h}\boldsymbol{\theta} + \boldsymbol{\theta}^T \mathbf{h}^T \mathbf{h}\boldsymbol{\theta} \right] \\ &\stackrel{(a)}{=} -\frac{1}{2\sigma^2} \left[-2\mathbf{h}^T \mathbf{x} + 2\mathbf{h}^T \mathbf{h}\boldsymbol{\theta} \right] \\ &= \frac{1}{\sigma^2} \left[\mathbf{h}^T \mathbf{x} - \mathbf{h}^T \mathbf{h}\boldsymbol{\theta} \right] \\ &\stackrel{(b)}{=} \frac{\mathbf{h}^T \mathbf{h}}{\sigma^2} \left[(\mathbf{h}^T \mathbf{h})^{-1} \mathbf{h}^T \mathbf{x} - \boldsymbol{\theta} \right].\end{aligned}$$

Donde se utilizó,

- (a) $\frac{\partial \mathbf{B}^T \boldsymbol{\theta}}{\partial \boldsymbol{\theta}} = \mathbf{B}$, y $\frac{\partial \boldsymbol{\theta}^T \mathbf{A} \boldsymbol{\theta}}{\partial \boldsymbol{\theta}} = 2\mathbf{A}\boldsymbol{\theta}$ si \mathbf{A} es simétrica (Ver apéndice)
- (b) $\mathbf{h}^T \mathbf{h}$ invertible.

Definición y propiedades

CRLB en Modelos Lineales:

- El modelo cumple la condición del teorema de CRLB dada por,

$$\frac{\log p(\mathbf{x}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \underbrace{\frac{\mathbf{h}^T \mathbf{h}}{\sigma^2}}_{\mathbf{I}(\boldsymbol{\theta})} \left[\underbrace{(\mathbf{h}^T \mathbf{h})^{-1} \mathbf{h}^T \mathbf{x}}_{\mathbf{g}(\mathbf{x})} - \boldsymbol{\theta} \right].$$

- Por lo que tenemos

Estimador MVU

$$\hat{\boldsymbol{\theta}} = \mathbf{g}(\mathbf{x}) = (\mathbf{h}^T \mathbf{h})^{-1} \mathbf{h}^T \mathbf{x}$$

Matriz de Covarianza

$$\mathbf{C}_{\hat{\boldsymbol{\theta}}} = \mathbf{I}^{-1}(\boldsymbol{\theta}) = \sigma^2 (\mathbf{h}^T \mathbf{h})^{-1}$$

- Además, el estimador $\hat{\boldsymbol{\theta}}$ es **eficiente**.
- Observación:** Problema equivalente al problema de mínimos cuadrados.

Teorema: Estimador MVU en modelos lineales.

$$\mathbf{x} = \mathbf{h}\boldsymbol{\theta} + \mathbf{w}$$

- $\mathbf{x} \in \mathbb{R}^N$: datos observados.
- $\boldsymbol{\theta} \in \mathbb{R}^p$: p parámetros desconocidos.
- $\mathbf{h} \in \mathbb{R}^{N \times p}$: matriz de observación con $N > p$ y rango p
- $\mathbf{w} \in \mathbb{R}^N$: ruido en observación. Se asume $\mathbf{w} \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$.

Nota: El espacio \mathbb{R} puede ser remplazado por \mathbb{C} . Entonces,

Estimador MVU

$$\hat{\boldsymbol{\theta}} = \mathbf{g}(\mathbf{x}) = (\mathbf{h}^T \mathbf{h})^{-1} \mathbf{h}^T \mathbf{x}$$

Matriz de Covarianza

$$\mathbf{C}_{\hat{\boldsymbol{\theta}}} = \mathbf{I}^{-1}(\boldsymbol{\theta}) = \sigma^2 (\mathbf{h}^T \mathbf{h})^{-1}$$

- Además, el estimador $\hat{\boldsymbol{\theta}}$ es **eficiente** ya que alcanza la CRLB para todo $\boldsymbol{\theta}$.

Teorema: Estimador MVU en modelos lineales.

- El estimador es insesgado:

$$\begin{aligned}\mathbb{E} \left[\hat{\boldsymbol{\theta}} \right] &= \mathbb{E} \left[(\mathbf{h}^T \mathbf{h})^{-1} \mathbf{h}^T \mathbf{x} \right] \\ &= \mathbb{E} \left[(\mathbf{h}^T \mathbf{h})^{-1} \mathbf{h}^T (\mathbf{h} \boldsymbol{\theta} + \mathbf{w}) \right] \\ &= \mathbb{E} \left[\boldsymbol{\theta} + (\mathbf{h}^T \mathbf{h})^{-1} \mathbf{h}^T \mathbf{w} \right] \\ &= \boldsymbol{\theta} + (\mathbf{h}^T \mathbf{h})^{-1} \mathbf{h}^T \mathbb{E}[\mathbf{w}] \\ &= \boldsymbol{\theta}\end{aligned}$$

- Además, como $\hat{\boldsymbol{\theta}}$ es una transformación lineal del vector Gaussiano \mathbf{x} , su PDF es Gaussiana y está completamente especificada,

$$\hat{\boldsymbol{\theta}} \sim \mathcal{N}(\boldsymbol{\theta}, \sigma^2 (\mathbf{h}^T \mathbf{h})^{-1}).$$

Ejemplo I: Análisis de Fourier

- Muchas señales tienen un comportamiento periódico
- Presencia de componentes sinusoidales puede ser detectada mediante análisis de Fourier
- Grandes coeficientes indican presencia de fuertes componentes sinusoidales
- Supongamos como modelo la superposición de sinusoides en AWGN,

$$x[n] = \sum_{k=1}^M a_k \cos\left(\frac{2\pi kn}{N}\right) + b_k \sin\left(\frac{2\pi kn}{N}\right) + w[n], \quad n = 0, 1, \dots, N-1,$$

donde $w[n] \sim \mathcal{N}(0, \sigma^2 I)$. Las amplitudes a_k y b_k son desconocidas. El modelo asume que las frecuencias se encuentran en relación armónica con frecuencia fundamental $f_1 = 1/N$ (frecuencia normalizada)

- ¿Es un modelo lineal en los parámetros?

Ejemplo I: Análisis de Fourier

- Superposición de sinusoides en AWGN:

$$x[n] = \sum_{k=1}^M a_k \cos\left(\frac{2\pi kn}{N}\right) + b_k \sin\left(\frac{2\pi kn}{N}\right) + w[n], \quad n = 0, 1, \dots, N-1,$$

donde las amplitudes a_k y b_k son desconocidas.

- El modelo es lineal, $\mathbf{x} = \mathbf{h}\boldsymbol{\theta} + \mathbf{w}$, con

$$\boldsymbol{\theta} = [a_1, a_2, \dots, a_M, b_1, b_2, \dots, b_M]^T$$

$$\mathbf{h} = \begin{bmatrix} 1 & \dots & 1 & 0 & \dots & 0 \\ \cos\left(\frac{2\pi}{N}\right) & \dots & \cos\left(\frac{2\pi M}{N}\right) & \sin\left(\frac{2\pi}{N}\right) & \dots & \sin\left(\frac{2\pi M}{N}\right) \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ \cos\left(\frac{2\pi(N-1)}{N}\right) & \dots & \cos\left(\frac{2\pi M(N-1)}{N}\right) & \sin\left(\frac{2\pi(N-1)}{N}\right) & \dots & \sin\left(\frac{2\pi M(N-1)}{N}\right) \end{bmatrix}$$

- $\boldsymbol{\theta}$ es de dimensión $2M \times 1$ y \mathbf{h} es de dimensión $N \times 2M$.

Ejemplo I: Análisis de Fourier

Determinación del MVU

- El estimador MVU es $\hat{\theta} = (\mathbf{h}^T \mathbf{h})^{-1} \mathbf{h}^T \mathbf{x}$.
- Si se representa \mathbf{h} en columnas,

$$\mathbf{h} = [\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_{2M}],$$

entonces se tiene que

$$\mathbf{h}^T \mathbf{h} = \begin{bmatrix} \mathbf{h}_1^T \\ \mathbf{h}_2^T \\ \vdots \\ \mathbf{h}_{2M}^T \end{bmatrix} [\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_{2M}] = \begin{bmatrix} \mathbf{h}_1^T \mathbf{h}_1 & \mathbf{h}_1^T \mathbf{h}_2 & \dots & \mathbf{h}_1^T \mathbf{h}_{2M} \\ \mathbf{h}_2^T \mathbf{h}_1 & \mathbf{h}_2^T \mathbf{h}_2 & \dots & \mathbf{h}_2^T \mathbf{h}_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{h}_{2M}^T \mathbf{h}_1 & \mathbf{h}_{2M}^T \mathbf{h}_2 & \dots & \mathbf{h}_{2M}^T \mathbf{h}_{2M} \end{bmatrix}.$$

- Los cálculos se simplifican notando que las columnas de \mathbf{h} son ortogonales. Es decir si \mathbf{h}_i es la columna i -ésima de \mathbf{h} , se cumple que

$$\mathbf{h}_i^T \mathbf{h}_j = 0 \quad \text{para } i \neq j$$

- Esto implica que $\mathbf{h}^T \mathbf{h}$ es diagonal, haciéndola fácil de invertir.

Ejemplo I: Análisis de Fourier

- La ortogonalidad de las columnas viene de la DFT (ejercicio),

$$\sum_{n=0}^{N-1} \cos\left(\frac{2\pi in}{N}\right) \cos\left(\frac{2\pi jn}{N}\right) = \frac{N}{2} \delta_{ij}$$

$$\sum_{n=0}^{N-1} \sin\left(\frac{2\pi in}{N}\right) \sin\left(\frac{2\pi jn}{N}\right) = \frac{N}{2} \delta_{ij}$$

$$\sum_{n=0}^{N-1} \cos\left(\frac{2\pi in}{N}\right) \sin\left(\frac{2\pi jn}{N}\right) = 0 \quad \forall i, j$$

- Con lo cual

$$\mathbf{h}^T \mathbf{h} = \begin{bmatrix} \frac{N}{2} & 0 & \dots & 0 \\ 0 & \frac{N}{2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{N}{2} \end{bmatrix} = \frac{N}{2} \mathbf{I}$$

Ejemplo I: Análisis de Fourier

- El estimador MVU queda entonces

$$\hat{\boldsymbol{\theta}} = (\mathbf{h}^T \mathbf{h})^{-1} \mathbf{h}^T \mathbf{x} = \frac{2}{N} \mathbf{I} \mathbf{h}^T \mathbf{x} = \frac{2}{N} \begin{bmatrix} \mathbf{h}_1^T \\ \mathbf{h}_2^T \\ \vdots \\ \mathbf{h}_{2M}^T \end{bmatrix} \mathbf{x} = \frac{2}{N} \begin{bmatrix} \mathbf{h}_1^T \mathbf{x} \\ \mathbf{h}_2^T \mathbf{x} \\ \vdots \\ \mathbf{h}_{2M}^T \mathbf{x} \end{bmatrix}$$

- De donde se obtiene,

$$\hat{a}_k = \frac{2}{N} \sum_{n=0}^{N-1} x[n] \cos\left(\frac{2\pi kn}{N}\right) \quad \hat{b}_k = \frac{2}{N} \sum_{n=0}^{N-1} x[n] \sin\left(\frac{2\pi kn}{N}\right),$$

es decir, \hat{a}_k y \hat{b}_k son los coeficientes de Fourier de $x[n]$.

- La matriz de covarianza es:

$$\mathbf{C}_{\hat{\boldsymbol{\theta}}} = \sigma^2 (\mathbf{h}^T \mathbf{h})^{-1} = \frac{2\sigma^2}{N} \mathbf{I}.$$

Dado que la matriz de covarianza es diagonal, los estimadores son independientes.

Modelo lineal con ruido gaussiano coloreado

$$\mathbf{x} = \mathbf{h}\boldsymbol{\theta} + \mathbf{w}$$

- $\mathbf{x} \in \mathbb{R}^N$: datos observados.
- $\boldsymbol{\theta} \in \mathbb{R}^p$: p parámetros desconocidos.
- $\mathbf{h} \in \mathbb{R}^{N \times p}$: matriz de observación.
- $\mathbf{w} \in \mathbb{R}^N$: ruido en observación es **coloreado**, muestras no son independientes. Se asume $\mathbf{w} \sim \mathcal{N}(0, \mathbf{C})$, con \mathbf{C} simétrica definida positiva.

- Para encontrar el estimador MVU, se puede proceder de forma análoga al caso con ruido blanco.
- En el caso con ruido coloreado, la PDF del ruido es una normal multivariada

$$p(\mathbf{w}) = \frac{1}{(2\pi|\mathbf{C}|)^{N/2}} \exp \left[-\frac{1}{2} \mathbf{w}^T \mathbf{C}^{-1} \mathbf{w} \right],$$

por lo que la PDF paramétrica de los datos es

$$p(\mathbf{x}; \boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi}^N |\mathbf{C}|} \exp \left[-\frac{1}{2} (\mathbf{x} - \mathbf{h}\boldsymbol{\theta})^T \mathbf{C}^{-1} (\mathbf{x} - \mathbf{h}\boldsymbol{\theta}) \right].$$

Modelo lineal con ruido gaussiano coloreado

- \mathbf{C} es una matriz de covarianza,

$$\mathbf{C} = \mathbb{E} \left((\mathbf{w} - \mathbb{E}(\mathbf{w}))(\mathbf{w} - \mathbb{E}(\mathbf{w}))^T \right) = \mathbb{E}(\mathbf{w}\mathbf{w}^T),$$

por lo que es simétrica y definida positiva (ver apéndice).

- Además \mathbf{C}^{-1} también es simétrica y definida positiva (Ejercicio). Por lo cual

$$\mathbf{C}^{-1} = \mathbf{D}^T \mathbf{D},$$

con \mathbf{D} de dimensión $N \times N$ invertible.

- La matriz \mathbf{D} actúa como un blanqueador (*whitening*),

$$\begin{aligned} \mathbb{E}[(\mathbf{D}\mathbf{w})(\mathbf{D}\mathbf{w})^T] &= \mathbf{D}\mathbb{E}[\mathbf{w}\mathbf{w}^T]\mathbf{D} \\ &= \mathbf{D}\mathbf{C}\mathbf{D}^T \\ &= \mathbf{D}\mathbf{D}^{-1}(\mathbf{D}^T)^{-1}\mathbf{D}^T \\ &= \mathbf{I}. \end{aligned}$$

- Como consecuencia se tiene que,

$$\mathbf{w}' = \mathbf{D}\mathbf{w} \sim \mathcal{N}(0, \mathbf{I}).$$

Modelo lineal con ruido gaussiano coloreado

- Transformando el modelo original con operación de blanqueado, se obtiene

$$\begin{aligned}\mathbf{x}' &= \mathbf{D}\mathbf{x} \\ &= \mathbf{D}\mathbf{h}\boldsymbol{\theta} + \mathbf{D}\mathbf{w} \\ &= \mathbf{h}'\boldsymbol{\theta} + \mathbf{w}',\end{aligned}$$

donde se definió $\mathbf{h}' = \mathbf{D}\mathbf{h}$. Notar que $\mathbf{w}' \sim \mathcal{N}(0, \mathbf{I})$.

- Como se vio anteriormente, el MVU en este caso (AWGN) es

$$\begin{aligned}\hat{\boldsymbol{\theta}} &= (\mathbf{h}'^T \mathbf{h}')^{-1} \mathbf{h}'^T \mathbf{x}' \\ &= (\mathbf{h}^T \mathbf{D}^T \mathbf{D} \mathbf{h})^{-1} \mathbf{h}^T \mathbf{D}^T \mathbf{D} \mathbf{x} \\ &= (\mathbf{h}^T \mathbf{C}^{-1} \mathbf{h})^{-1} \mathbf{h}^T \mathbf{C}^{-1} \mathbf{x}\end{aligned}$$

- y su matriz de covarianza,

$$\mathbf{C}_{\hat{\boldsymbol{\theta}}} = (\mathbf{h}'^T \mathbf{h}')^{-1} = (\mathbf{h}^T \mathbf{C}^{-1} \mathbf{h})^{-1}.$$

- Si $\mathbf{C} = \sigma^2 \mathbf{I}$, se obtiene el resultado para el caso AWGN.

Nota: podríamos haber calculado la derivada de la función de verosimilitud y procedido igual que antes.

El resultado es el mismo (ejercicio).

Ejemplo II

Ejemplo: Nivel de DC en ruido coloreado

Como generalización del ejemplo de base, en este caso se observan N muestras de nivel de continua en **ruido coloreado**,

$$x[n] = A + w[n], \quad \text{con } n = 0, 1, \dots, N-1, \text{ y } \mathbf{w} \sim \mathcal{N}(0, \mathbf{C}),$$

donde \mathbf{C} es la matriz de covarianza del ruido (simétrica, definida positiva).

- El modelo es lineal,

$$\mathbf{x} = \mathbf{1} \cdot A + \mathbf{w}, \quad \mathbf{h} = \mathbf{1} = [1, 1, \dots, 1]^T$$

- Empleando el resultado del modelo lineal con ruido coloreado, el estimador MVU y la varianz es:

$$\hat{A} = (\mathbf{h}^T \mathbf{C}^{-1} \mathbf{h})^{-1} \mathbf{h}^T \mathbf{C}^{-1} \mathbf{x} = \frac{\mathbf{1}^T \mathbf{C}^{-1} \mathbf{x}}{\mathbf{1}^T \mathbf{C}^{-1} \mathbf{1}}$$

y su varianza es,

$$\text{var}(\hat{A}) = (\mathbf{h}^T \mathbf{C}^{-1} \mathbf{h})^{-1} = \frac{1}{\mathbf{1}^T \mathbf{C}^{-1} \mathbf{1}}$$

- En el caso en que el ruido es blanco y gaussiano, el MVU es la media muestra con varianza σ^2/N .

Ejemplo II

Ejemplo: Nivel de DC en ruido coloreado

- Interpretación. Considerando la factorización $\mathbf{C}^{-1} = \mathbf{D}^T \mathbf{D}$, se obtiene

$$\hat{A} = \frac{\mathbf{1}^T \mathbf{D}^T \mathbf{D} \mathbf{x}}{\mathbf{1}^T \mathbf{D}^T \mathbf{D} \mathbf{1}} = \frac{(\mathbf{D} \mathbf{1})^T \mathbf{x}'}{(\mathbf{D} \mathbf{1})^T (\mathbf{D} \mathbf{1})} = \sum_{n=0}^{N-1} d_n x'[n], \quad \text{con } d_n = \frac{[\mathbf{D} \mathbf{1}]_n}{(\mathbf{D} \mathbf{1})^T (\mathbf{D} \mathbf{1})}$$

- El estimador consiste en primero blanquear los datos ($x'[n]$) y luego hacer un promedio ponderado de los datos blanqueados.
- El blanqueado tiene el efecto de decorrelacionar e igualar las varianzas de las muestras de ruido

Caso particular: muestras no correlacionadas pero de varianza distinta

$$\mathbf{C} = \text{diag}(\sigma_0^2, \sigma_1^2, \dots, \sigma_{N-1}^2)$$

- En este caso,

$$\hat{A} = \frac{\mathbf{1}^T \mathbf{C}^{-1} \mathbf{x}}{\mathbf{1}^T \mathbf{C}^{-1} \mathbf{1}} = \frac{\sum_{n=0}^{N-1} \frac{x[n]}{\sigma_n^2}}{\sum_{n=0}^{N-1} \frac{1}{\sigma_n^2}}.$$

- Estimador le da más importancia a muestras con ruido de menor varianza
- ¿Qué sucede si alguna muestra tiene ruido con varianza $\epsilon \approx 0$?

Modelo lineal + señal conocida + ruido coloreado

$$\mathbf{x} = \mathbf{h}\boldsymbol{\theta} + \mathbf{s} + \mathbf{w}$$

- $\mathbf{x} \in \mathbb{R}^N$: datos observados.
- $\boldsymbol{\theta} \in \mathbb{R}^p$: p parámetros desconocidos.
- $\mathbf{h} \in \mathbb{R}^{N \times p}$: matriz de observación.
- $\mathbf{s} \in \mathbb{R}^N$: muestras de señal conocida.
- $\mathbf{w} \in \mathbb{R}^N$: Se asume $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \mathbf{C})$ (ruido coloreado).

Nota: El espacio \mathbb{R} puede ser remplazado por \mathbb{C} .

- En este caso es fácil ver que el estimador MVU es,

$$\hat{\boldsymbol{\theta}} = (\mathbf{h}^T \mathbf{C}^{-1} \mathbf{h})^{-1} \mathbf{h}^T \mathbf{C}^{-1} (\mathbf{x} - \mathbf{s})$$

y la matriz de covarianza del estimador es

$$\mathbf{C}_{\hat{\boldsymbol{\theta}}} = (\mathbf{h}^T \mathbf{C}^{-1} \mathbf{h})^{-1}.$$

- El estimador es **eficiente** ya que alcanza la CRLB para todo $\boldsymbol{\theta}$.

Ejemplo III

Ejemplo: Nivel de DC y exponencial en ruido coloreado

- Supongamos el siguiente modelo de observación

$$x[n] = A + r^n + w[n] \quad \text{con } n = 0, 1, \dots, N-1, \quad \text{y } \mathbf{w} \sim \mathcal{N}(0, \mathbf{C}),$$

donde r es conocido. Se quiere determinar el MVU de A y su varianza.

- El modelo es,

$$\mathbf{x} = \mathbf{1} \cdot A + \mathbf{s} + \mathbf{w},$$

y por lo tanto

$$\mathbf{h} = \mathbf{1} = [1, 1, 1, \dots, 1]^T$$

$$\mathbf{s} = [1, r, r^2, \dots, r^{N-1}]^T.$$

- El estimador y su varianza se obtienen con las fórmulas de la slide anterior,

$$\hat{A} = \frac{\mathbf{1}^T \mathbf{C}^{-1} (\mathbf{x} - \mathbf{s})}{\mathbf{1}^T \mathbf{C}^{-1} \mathbf{1}}$$

y su varianza es

$$\text{var}(\hat{A}) = \frac{1}{\mathbf{1}^T \mathbf{C}^{-1} \mathbf{1}}.$$

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 σ^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 σ^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 estadístico 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 θ .

Además, si $T(\mathbf{x})$ es un estadístico suficiente de θ 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 σ^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 θ y $T(\mathbf{x})$ un estadístico suficiente de θ .

Entonces:

- $\hat{\theta} := \mathbb{E}[\check{\theta}|T(\mathbf{x})]$ es un estimador insesgado de θ , 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 θ .

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 Si $T(\mathbf{x})$ es completo, 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 σ^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 σ^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, asumiendo que $T(\mathbf{x})$ es completo, 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, Capítulos 4 y 5.

Apéndice I

- Derivada de una función escalar $f : \mathbb{R} \rightarrow \mathbb{R}^n$ respecto a un vector $\boldsymbol{\theta} \in \mathbb{R}^n$,

$$\frac{\partial f}{\partial \boldsymbol{\theta}} = \begin{bmatrix} \frac{\partial f}{\partial \theta_1} \\ \frac{\partial f}{\partial \theta_2} \\ \vdots \\ \frac{\partial f}{\partial \theta_n} \end{bmatrix}$$

- Si $f(\boldsymbol{\theta}) = \mathbf{w}^T \boldsymbol{\theta} = \sum_{i=1}^n w_i \theta_i$. Entonces

$$\frac{\partial \mathbf{w}^T \boldsymbol{\theta}}{\partial \boldsymbol{\theta}} = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} = \mathbf{w}.$$

- Si $g(\boldsymbol{\theta}) = \boldsymbol{\theta}^T \mathbf{A} \boldsymbol{\theta} = \sum_{i,j} a_{ij} \theta_i \theta_j$. Entonces

$$\frac{\partial \boldsymbol{\theta}^T \mathbf{A} \boldsymbol{\theta}}{\partial \boldsymbol{\theta}} = \begin{bmatrix} \sum_i a_{i1} \theta_i + \sum_j a_{1j} \theta_j \\ \sum_i a_{i2} \theta_i + \sum_j a_{2j} \theta_j \\ \vdots \\ \sum_i a_{in} \theta_i + \sum_j a_{nj} \theta_j \end{bmatrix} = \mathbf{A} \boldsymbol{\theta} + \mathbf{A}^T \boldsymbol{\theta}.$$

Ortogonalidad de la base DFT

$$\sum_{n=0}^{N-1} \cos\left(\frac{2\pi kn}{N}\right) \cos\left(\frac{2\pi ln}{N}\right) = \frac{1}{2} \sum_{n=0}^{N-1} \cos\left(\frac{2\pi(k+l)n}{N}\right) + \frac{1}{2} \sum_{n=0}^{N-1} \cos\left(\frac{2\pi(k-l)n}{N}\right)$$

Por otro lado,

$$\begin{aligned} \sum_{n=0}^{N-1} \cos\left(\frac{2\pi Mn}{N}\right) &= \operatorname{Re} \left[\sum_{n=0}^{N-1} \exp\left(\frac{2\pi Mn}{N}\right) \right] \\ &= \operatorname{Re} \left[\frac{1 - \exp(2\pi i M)}{1 - \exp(2\pi i M/N)} \right] \\ &= \begin{cases} N & \text{si } M = 0 \\ 0 & \text{si } M \in \mathbb{N}, M \neq 0 \end{cases} \end{aligned}$$

Con lo cual,

$$\sum_{n=0}^{N-1} \cos\left(\frac{2\pi kn}{N}\right) \cos\left(\frac{2\pi ln}{N}\right) = \frac{N}{2} \delta_{kl}.$$

- Propiedades de la matriz de covarianza

$$\mathbf{C} = \mathbb{E} \left[(\mathbf{w} - \mathbb{E}(\mathbf{w}))(\mathbf{w} - \mathbb{E}(\mathbf{w}))^T \right].$$

- Supongamos para simplificar que $\mathbb{E}[\mathbf{w}] = 0$. Entonces hay que probar que

$$\mathbb{E}[\mathbf{w}\mathbf{w}^T]$$

es simétrica (trivial) y definida positiva.

- Para ser semi-definida positiva hay que probar que si $\mathbf{b} \neq 0$, entonces $\mathbf{b}^T \mathbf{C} \mathbf{b} \geq 0$,

$$\begin{aligned} \mathbf{b}^T \mathbf{C} \mathbf{b} &= \mathbf{b}^T \mathbb{E}[\mathbf{w}\mathbf{w}^T] \mathbf{b} \\ &= \mathbb{E}[\mathbf{b}^T \mathbf{w}\mathbf{w}^T \mathbf{b}] \\ &= \mathbb{E}[(\mathbf{b}^T \mathbf{w})(\mathbf{b}^T \mathbf{w})^T] \\ &= \mathbb{E}[|\mathbf{b}^T \mathbf{w}|^2] \\ &\geq 0. \end{aligned}$$