

# Estimación y Predicción en Series Temporales

Cota inferior de Cramer-Rao (CRLB)

Departamento de Procesamiento de Señales

Instituto de Ingeniería Eléctrica  
Facultad de Ingeniería

2022

# Agenda

- 1 Repaso MVU
- 2 Cota inferior de Cramér-Rao (CRLB)

## Planteo del Problema:

- Dadas  $N$  muestras de una señal discreta  $x[n]$  que depende de cierto parámetro  $\theta$  desconocido.
- Estimar  $\theta$  a partir de las  $N$  muestras  $x[0], x[1], \dots, x[N-1]$

Para ello se define un estimador de  $\theta$  que es función de los datos:

$$\hat{\theta} = g(x[0], x[1], \dots, x[N-1])$$

- $g$ : función a determinar
- $\hat{\theta}$ : estimador de  $\theta$

**Objetivo:** Encontrar una función  $g$  de forma que  $\hat{\theta}$  sea un buen estimador de  $\theta$ .

- Estimador  $\hat{\theta}$  debe ser cercano (en algún sentido a definir) al valor verdadero de  $\theta$ .
- El criterio de cercanía debe ser especificado teniendo en cuenta que  $\hat{\theta}$  es una Variable Aleatoria.

# Criterio de Mínima Varianza

- En la búsqueda de **estimadores óptimos** es necesario algún **criterio de optimalidad**.
- Uno natural es la minimización del **error cuadrático medio** (*MSE*, *Mean Square Error*)

$$\text{MSE}(\hat{\theta}) = \mathbb{E} \left[ (\hat{\theta} - \theta)^2 \right].$$

- Análisis (descomposición) del error cuadrático medio:

$$\begin{aligned}\text{MSE}(\hat{\theta}) &= \mathbb{E} \left[ (\hat{\theta} - \theta)^2 \right] \\&= \mathbb{E} \left\{ \left[ (\hat{\theta} - \mathbb{E}(\hat{\theta})) + (\mathbb{E}(\hat{\theta}) - \theta) \right]^2 \right\} \\&= \mathbb{E} \left[ (\hat{\theta} - \mathbb{E}(\hat{\theta})) \right] + \underbrace{2 \mathbb{E} \left[ (\hat{\theta} - \mathbb{E}(\hat{\theta})) (\mathbb{E}(\hat{\theta}) - \theta) \right]}_{\mathbb{E}^2(\hat{\theta}) - \mathbb{E}(\hat{\theta})\theta - \mathbb{E}^2(\hat{\theta}) + \mathbb{E}(\hat{\theta})\theta = 0} + \mathbb{E} \left[ (\mathbb{E}(\hat{\theta}) - \theta)^2 \right] \\&= \mathbb{E} \left[ (\hat{\theta} - \mathbb{E}(\hat{\theta}))^2 \right] + (\mathbb{E}(\hat{\theta}) - \theta)^2 \\&= \text{var}(\hat{\theta}) + b^2(\hat{\theta})\end{aligned}$$

- Descomposición sumamente útil *bias-variance*.

# Extensión a vector de parámetros

En el problema general de estimación de parámetros, los parámetros desconocidos pueden ser varios.

## Estimador insesgado

- Si hay  $p$  parámetros desconocidos, se construye el vector de parámetros desconocidos,

$$\boldsymbol{\theta} = [\theta_1, \theta_2, \dots, \theta_p]^T.$$

- Se dice que un estimador  $\hat{\boldsymbol{\theta}} = [\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_p]^T$  es insesgado, si

$$\mathbb{E}(\hat{\theta}_i) = \theta_i, \quad a_i < \theta_i < b_i,$$

para todo  $i=1,2,\dots,p$ .

- Si definimos la esperanza de un vector de variables aleatorias como

$$\mathbb{E}(\hat{\boldsymbol{\theta}}) = [\mathbb{E}(\hat{\theta}_1), \mathbb{E}(\hat{\theta}_2), \dots, \mathbb{E}(\hat{\theta}_p)]^T$$

- Un estimador insesgado cumple la igualdad vectorial

$$\mathbb{E}(\hat{\boldsymbol{\theta}}) = \boldsymbol{\theta}.$$

# MVU: Extensión a vector de parámetros

Estimador  $\hat{\theta}$  de parámetro vectorial  $\theta$  es **MVU** si:

- 1 Es insesgado, es decir cumple la igualdad vectorial

$$\mathbb{E}(\hat{\theta}) = \theta;$$

- 2 cumple la propiedad de que

$$\text{var}(\hat{\theta}_i) \text{ es mínima, para } i = 1, 2, \dots, p,$$

entre todos los estimadores insesgados.

# Cota inferior de Crámer-Rao (CRLB)

La **cota inferior de Cramér-Rao** establece una cota inferior teórica en la varianza de un estimador insesgado:

$$\text{var}(\hat{\theta}) \geq \mathbf{CRLB}(\theta),$$

para todo estimador **insesgado**  $\hat{\theta}$ .

## Utilidad práctica:

- Permite afirmar si un estimador insesgado es el estimador MVU.
- Esto sucede si el estimador alcanza la cota para todos los valores posibles de  $\theta$  desconocido,  $\text{var}(\hat{\theta}) = \mathbf{CRLB}(\theta)$ , para todo valor de  $\theta$ .
- Provee una referencia contra la cual comparar el desempeño de cualquier estimador insesgado.
- Indica la imposibilidad física de encontrar un estimador insesgado con varianza menor que la cota. Útil en estudios de viabilidad.
- Permite además determinar si existe un estimador que alcanza la cota.

Dependencia de la PDF de los datos (modelo de datos) con el parámetro:

- Toda la información está contenida en los datos observados y en su modelo (función de densidad de probabilidad, PDF)
- La precisión de la estimación depende directamente de la PDF.
- No se puede esperar una estimación con mucha precisión si la PDF depende débilmente del parámetro.
- Cuanto mayor es la influencia del parámetros desconocido sobre la PDF, mejor debería poder estimarse el parámetro.



## Ejemplo: Dependencia de la PDF con el parámetro desconocido

Se quiere estimar el nivel de DC (parámetro  $A$ ) en WGN cuando se observa una sola muestra,

$$x[0] = A + w[0], \text{ donde } w[0] \sim \mathcal{N}(0, \sigma^2).$$

- Se espera que la estimación sea mejor si  $\sigma^2$  es pequeño (poco ruido).
- Un buen estimador insesgado es:

$$\hat{A} = x[0]$$

- La varianza del estimador es:

$$\text{var}(\hat{A}) = \sigma^2.$$

- La precisión del estimador mejora a medida que  $\sigma^2$  decrece.

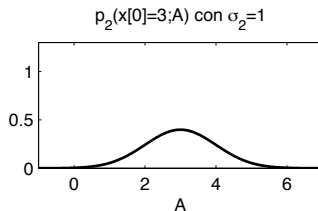
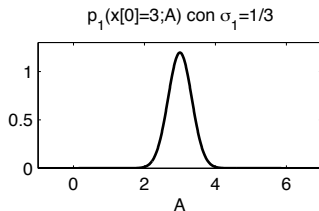
## Ejemplo: Dependencia de la PDF con el parámetro desconocido

- Se considera la PDF para dos valores distintos de varianza

$$p_i(x[0]; A) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp \left[ -\frac{1}{2\sigma_i^2} (x[0] - A)^2 \right], \quad \text{con } i = 1, 2.$$

- Supongamos  $x[0] = 3$ . Se observa la PDF en función de  $A$ , para el valor obtenido de  $x[0]$ . Se consideran los valores  $\sigma_1 = 1/3$  y  $\sigma_2 = 1$ .

**Definición:** PDF vista como una función del parámetro desconocido  $\theta$  con  $\mathbf{x}$  fijo, se denomina **función de verosimilitud** (*likelihood function*).



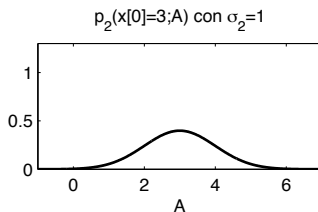
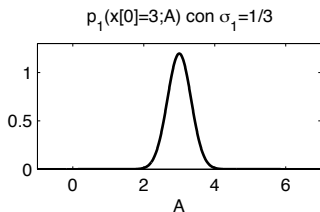
## Ejemplo: Dependencia de la PDF con el parámetro desconocido

- con  $\sigma_1 = 1/3$ , los valores de  $A > 4$  tienen una probabilidad de:

$$\Pr\{A > 4 \mid X[0] = 3\} = 1 - \Phi\left(\frac{A - x[0]}{\sigma_1}\right) = 1 - \Phi(3) \approx 0.0013$$

- con  $\sigma_2 = 1$ , los valores de  $A > 4$  tienen una probabilidad de:

$$\Pr\{A > 4 \mid X[0] = 3\} = 1 - \Phi\left(\frac{A - x[0]}{\sigma_2}\right) = 1 - \Phi(1) \approx 0.1587$$



## Ejemplo: Dependencia de la PDF con el parámetro desconocido

- Si  $x \sim \mathcal{N}(\mu, \sigma^2)$ , entonces  $\Pr\{|x - \mu| \leq 3\sigma\} \approx 0.9973$ .
- Valores de  $A$  **fuera** del intervalo  $x[0] \pm 3\sigma_i$  son muy poco probables.
  - con  $\sigma_1 = 1/3$ , candidatos probables  $A \in [2, 4]$ ,
  - con  $\sigma_2 = 1$ , candidatos probables  $A \in [0, 6]$ .

## Observaciones:

- La función de verosimilitud  $p_{\sigma_2=1}(x[0] = 3; A)$  tiene una **dependencia más débil** del parámetro  $A$  que  $p_{\sigma_1=1/3}(x[0] = 3; A)$  por lo que los candidatos probables de  $A$  se encuentran en un intervalo más amplio.
- Intuitivamente, la “**agudeza**” (**sharpness**) de la función de verosimilitud determina precisión con la cual es posible estimar el parámetro desconocido.
- Una forma de medir la agudeza de la función de verosimilitud es a través del **negativo de la derivada segunda** del logaritmo de la verosimilitud respecto al parámetro desconocido en el pico.

## Ejemplo: Dependencia de la PDF con el parámetro desconocido

Derivada segunda del logaritmo de la función de verosimilitud

- La función de verosimilitud es:

$$p(x[0]; A) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[ -\frac{1}{2\sigma^2} (x[0] - A)^2 \right].$$

- El logaritmo de la función de verosimilitud es

$$\log p(x[0]; A) = -\log \sqrt{2\pi\sigma^2} - \frac{1}{2\sigma^2} (x[0] - A)^2.$$

- Tomando la derivada primera,

$$\frac{\partial \log p(x[0]; A)}{\partial A} = \frac{1}{\sigma^2} (x[0] - A),$$

- y el opuesto de la derivada segunda queda,

$$-\frac{\partial^2 \log p(x[0]; A)}{\partial A^2} = \frac{1}{\sigma^2}.$$

$$-\frac{\partial^2 \log p(x[0]; A)}{\partial A^2} = \frac{1}{\sigma^2}$$

- La curvatura crece a medida que la varianza del ruido  $\sigma^2$  decrece.

- Teniendo en cuenta que el estimador es  $\hat{A} = x[0]$ , y por lo tanto su varianza es  $\text{var}(\hat{A}) = \sigma^2$ , para este ejemplo particular se cumple que:

$$\text{var}(\hat{A}) = \frac{1}{-\frac{\partial^2 \log p(x[0]; A)}{\partial A^2}} = \sigma^2$$

- En este ejemplo, la derivada segunda no depende de los datos ( $x[0]$ ), pero en lo general lo hará. Por lo tanto una medida más apropiada de la curvatura es:

$$-\mathbb{E} \left[ \frac{\partial^2 \log p(x[0]; A)}{\partial A^2} \right]$$

- Mide la curvatura media de la función de verosimilitud logarítmica.
- La esperanza se toma sobre el modelo de los datos ( $x[0]$ , V.A.), resultando en una función únicamente de A.

## Resumen

- Se dispone de un conjunto de datos y un modelo de los datos que depende de un parámetro desconocido que queremos estimar.
- El modelo impone una PDF de los datos (con un parámetro desconocido).
- Si se considera la PDF como función del parámetro manteniendo fijos los datos, la función se denomina función de *verosimilitud* (respectivamente función de *log-verosimilitud*).
- Cuánto más fuerte es la dependencia de la función de *verosimilitud* con el parámetro, el parámetro se puede estimar con mayor precisión.
- Una forma de medir la dependencia de la *log-verosimilitud* con el parámetro es a través de la concavidad (opuesto de la derivada segunda respecto al parámetro).
- El estimador del parámetro tendrá menor varianza cuanto mayor sea la concavidad de la función de verosimilitud.

# 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  $\theta$ .

- 2 existe un estimador que alcanza la cota para todo  $\theta$  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  $\frac{1}{I(\theta)}$ .



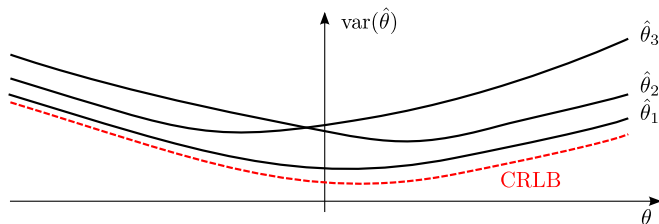
# Cota Inferior de Cramér-Rao: Consideraciones

- 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  $\theta$ .



## Condición de regularidad.

$$\begin{aligned}\mathbb{E} \left[ \frac{\partial \log p(\mathbf{x}; \theta)}{\partial \theta} \right] &= \int \frac{\partial \log p(\mathbf{x}; \theta)}{\partial \theta} p(\mathbf{x}; \theta) d\mathbf{x} \\ &\stackrel{(a)}{=} \int \frac{\partial p(\mathbf{x}; \theta)}{\partial \theta} d\mathbf{x} \\ &\stackrel{(b)}{=} \frac{\partial}{\partial \theta} \int p(\mathbf{x}; \theta) d\mathbf{x} \\ &\stackrel{(c)}{=} 0.\end{aligned}$$

- (a) Regla de la cadena
- (b) Cambio del orden de integración y diferenciación
- (c)  $\forall \theta, \int p(\mathbf{x}; \theta) d\mathbf{x} = 1$ .

- La condición de regularidad se cumple si es posible cambiar el orden de integración y diferenciación (paso (b)).
- Esto es cierto en general salvo cuando el soporte de  $p(\mathbf{x}; \theta)$  depende del parámetro desconocido  $\theta$ ,
  - se deduce de la regla de integración de Leibniz (ver apéndice I).

**Ejercicio.** Demostrar que en el caso en que  $x \sim \mathcal{U}[0, \theta]$  no se cumple la condición de

# CRLB: Prueba simple (1/3)

$$(1) \mathbb{E} [\hat{\theta}] = \int p(\mathbf{x}; \theta) g(\mathbf{x}) d\mathbf{x} = \theta \quad (2) \int p(\mathbf{x}; \theta) \frac{\partial}{\partial \theta} \log p(\mathbf{x}; \theta) d\mathbf{x} = 0$$

Derivando (1) con respecto a  $\theta$ :

$$\begin{aligned} \frac{\partial}{\partial \theta} \int p(\mathbf{x}; \theta) g(\mathbf{x}) d\mathbf{x} &\stackrel{(a)}{=} \int \frac{\partial}{\partial \theta} p(\mathbf{x}; \theta) g(\mathbf{x}) d\mathbf{x} \\ &\stackrel{(b)}{=} \int p(\mathbf{x}; \theta) \frac{\partial}{\partial \theta} \log p(\mathbf{x}; \theta) g(\mathbf{x}) d\mathbf{x} \\ &\stackrel{(c)}{=} \int p(\mathbf{x}; \theta) \frac{\partial}{\partial \theta} \log p(\mathbf{x}; \theta) (g(\mathbf{x}) - \theta) d\mathbf{x} \\ &\stackrel{(d)}{=} \frac{\partial}{\partial \theta} \theta = 1 \end{aligned}$$

- (a) intercambiando el orden de diferenciación e integración
- (b) regla de la cadena  $\left( \frac{\partial}{\partial \theta} \log p(\mathbf{x}; \theta) = \frac{1}{p(\mathbf{x}; \theta)} \frac{\partial}{\partial \theta} p(\mathbf{x}; \theta) \right)$
- (c) restando una constante ( $\theta$ ) multiplicada por algo que es igual a 0, por (2)
- (d) por (1)

# CRLB: Prueba simple (2/3)

Tenemos que

$$1 = \int p(\mathbf{x}; \theta) \frac{\partial}{\partial \theta} \log p(\mathbf{x}; \theta) (g(\mathbf{x}) - \theta) d\mathbf{x}$$

## Cauchy-Schwartz

$$\left[ \int w(\mathbf{x}) a(\mathbf{x}) b(\mathbf{x}) d\mathbf{x} \right]^2 \leq \int w(\mathbf{x}) a^2(\mathbf{x}) d\mathbf{x} \int w(\mathbf{x}) b^2(\mathbf{x}) d\mathbf{x}$$

- $w(\mathbf{x}) \geq 0, \forall \mathbf{x}$
- Igualdad si y sólo si  $a(\mathbf{x}) = c \cdot b(\mathbf{x})$  para  $c$  cte. con respecto a  $\mathbf{x}$

# CRLB: Prueba simple (2/3)

Tenemos que

$$1 = \int p(\mathbf{x}; \theta) \frac{\partial}{\partial \theta} \log p(\mathbf{x}; \theta) (g(\mathbf{x}) - \theta) d\mathbf{x}$$

## Cauchy-Schwartz

$$\left[ \int w(\mathbf{x}) a(\mathbf{x}) b(\mathbf{x}) d\mathbf{x} \right]^2 \leq \int w(\mathbf{x}) a^2(\mathbf{x}) d\mathbf{x} \int w(\mathbf{x}) b^2(\mathbf{x}) d\mathbf{x}$$

- $w(\mathbf{x}) \geq 0, \forall \mathbf{x}$
- Igualdad si y sólo si  $a(\mathbf{x}) = c \cdot b(\mathbf{x})$  para  $c$  cte. con respecto a  $\mathbf{x}$

Aplicando C.S.:

$$1 \leq \int p(\mathbf{x}; \theta) \left( \frac{\partial}{\partial \theta} \log p(\mathbf{x}; \theta) \right)^2 d\mathbf{x} \int p(\mathbf{x}; \theta) (g(\mathbf{x}) - \theta)^2 d\mathbf{x}$$
$$1 \leq \underbrace{\int p(\mathbf{x}; \theta) \left( \frac{\partial}{\partial \theta} \log p(\mathbf{x}; \theta) \right)^2 d\mathbf{x}}_{\mathbb{E}\left[\left(\frac{\partial}{\partial \theta} \log p(\mathbf{x}; \theta)\right)^2\right]} \underbrace{\int p(\mathbf{x}; \theta) (g(\mathbf{x}) - \theta)^2 d\mathbf{x}}_{\mathbb{E}[(g(\mathbf{x}) - \theta)^2]}$$

# CRLB: Prueba simple (3/3)

Finalmente:

$$\begin{aligned}\mathbb{E} [(g(\mathbf{x}) - \theta)^2] &= \text{var} [g(\mathbf{x})] \geq \frac{1}{\mathbb{E} \left[ \left( \frac{\partial}{\partial \theta} \log p(\mathbf{x}; \theta) \right)^2 \right]} \\ &= \frac{1}{-\mathbb{E} \left[ \frac{\partial^2}{\partial \theta^2} \log p(\mathbf{x}; \theta) \right]}\end{aligned}$$

□

Ejercicio: probar que  $\mathbb{E} \left[ \left( \frac{\partial}{\partial \theta} \log p(\mathbf{x}; \theta) \right)^2 \right] = -\mathbb{E} \left[ \frac{\partial^2}{\partial \theta^2} \log p(\mathbf{x}; \theta) \right]$ .

La segunda parte del teorema CRLB se prueba utilizando la condición de igualdad de Cauchy-Schwartz. (Ver Kay, apéndice 3A)

$$\underbrace{\frac{\partial \log p(\mathbf{x}; \theta)}{\partial \theta}}_{a(\mathbf{x})} = \underbrace{I(\theta)}_c \cdot \underbrace{(g(\mathbf{x}) - \theta)}_{b(\mathbf{x})}$$

# CRLB: Ejemplos (I)

## Ejemplo: CRLB nivel de DC en WGN (una única muestra)

- Estimar  $A$  a partir de  $x[0] = A + w[0]$  donde  $w[0] \sim \mathcal{N}(0, \sigma^2)$ .
- Eligiendo  $\hat{A} = x[0]$ , se tiene que  $\text{var}(\hat{A}) = \sigma^2$ .
- Derivadas primera y segunda de la función de log-verosimilitud:

$$\frac{\partial \log p(x[0]; A)}{\partial A} = \frac{1}{\sigma^2} (x[0] - A) \qquad \frac{\partial^2 \log p(x[0]; A)}{\partial A^2} = -\frac{1}{\sigma^2}$$

- Aplicando la CRLB se tiene que,

$$\text{var}(\tilde{A}) \geq \sigma^2, \forall A, \text{ para cualquier estimador insesgado } \tilde{A}.$$

- No existe un estimador insesgado de varianza menor a  $\sigma^2$ .
- **Conclusión:** Como el estimador elegido  $\hat{A}$  es insesgado y alcanza la CRLB para todo  $A$ , es el estimador MVU.

# CRLB: Ejemplos (I)

## Ejemplo: CRLB nivel de DC en WGN (una única muestra)

- En el caso de no haber descubierto a  $\hat{A} = x[0]$  como estimador, se podría haber empleado la segunda parte del teorema de la CRLB.

Es decir, si

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

entonces:

- $\hat{\theta} = g(\mathbf{x})$  es un estimador MVU
- la varianza mínima es  $1/I(\theta)$ .

Identificando términos:

- $\theta = A$
- $I(\theta) = \frac{1}{\sigma^2}$
- $g(x[0]) = x[0]$

En este caso,

$$\frac{\partial \log p(x[0]; A)}{\partial A} = \frac{1}{\sigma^2}(x[0] - A).$$

Por lo que:

- $\hat{A} = g(x[0]) = x[0]$  estimador MVU.
- $\text{var}(\hat{A}) = \frac{1}{I(\theta)} = \sigma^2$ .
- Además, se cumple la igualdad

$$I(\theta) = -\mathbb{E} \left[ \frac{\partial^2 \log p(x[0]; A)}{\partial A^2} \right].$$



# CRLB: Ejemplos (II)

## Ejemplo: CRLB nivel de DC en WGN (múltiples muestras)

Como generalización del ejemplo anterior, en este caso se observan múltiples muestras del nivel de continua en WGN,

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

Se quiere determinar la CRLB del problema de estimar  $A$ ,

- La función de verosimilitud es:

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

- Tomando el logaritmo, obtenemos la log-verosimilitud,

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

## Ejemplo: CRLB nivel de DC en WGN (múltiples muestras)

- Calculando la derivada primera de la log-verosimilitud, obtenemos

$$\begin{aligned}\frac{\partial \log p(\mathbf{x}; A)}{\partial A} &= \frac{\partial}{\partial A} \left\{ -\log \left[ (2\pi\sigma^2)^{\frac{N}{2}} \right] - \frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x[n] - A)^2 \right\} \\ &= \frac{1}{\sigma^2} \sum_{n=0}^{N-1} (x[n] - A) \\ &= \frac{N}{\sigma^2} \left( \frac{1}{N} \sum_{n=0}^{N-1} x[n] - A \right) \\ &= \frac{N}{\sigma^2} (\bar{x} - A)\end{aligned}$$

- Diferenciando nuevamente respecto de  $A$ ,

$$\frac{\partial^2 \log p(\mathbf{x}; A)}{\partial A^2} = -\frac{N}{\sigma^2}.$$

## Ejemplo: CRLB nivel de DC en WGN (múltiples muestras)

- Teniendo en cuenta que la derivada segunda es constante, mediante Teo. de la CRLB se obtiene,

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

- Además, asociando términos podemos reconocer,

$$g(\mathbf{x}) = \bar{x} = \frac{1}{N} \sum_{n=0}^{N-1} x[n] \quad (\text{media muestral})$$

alcanza la CRLB, y siendo insesgado, es un estimador MVU.

- La cota se cumple con igualdad

$$\text{var}(\hat{A}) = \frac{1}{I(A)} = \frac{\sigma^2}{N}.$$

- La varianza del estimador es inversamente proporcional a la cantidad de datos observados.

# CRLB: Ejemplos (III)

## Ejemplo: CRLB estimación de fase

Se quiere estimar la fase  $\phi$  de una senoide contaminada con AWGN,

$$x[n] = A \cos(2\pi f_0 n + \phi) + w[n], \quad \text{con } n = 0, 1, \dots, N-1,$$

donde  $w[n] \sim \mathcal{N}(0, \sigma^2)$  para todo  $n$ .

- (Ejercicio) Mostrar que la derivada de la función de log-verosimilitud es,

$$\frac{\partial \log p(\mathbf{x}; \phi)}{\partial \phi} = -\frac{A}{\sigma^2} \sum_{n=0}^{N-1} \left[ x[n] \sin(2\pi f_0 n + \phi) - \frac{A}{2} \sin(4\pi f_0 n + 2\phi) \right].$$

- Además,

$$\mathbb{E}_{\mathbf{x}} \left[ \frac{\partial^2 \log p(\mathbf{x}; \phi)}{\partial \phi^2} \right] \approx -\frac{NA^2}{2\sigma^2},$$

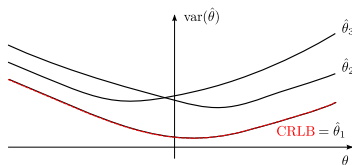
con lo cual

$$\text{var}(\hat{\phi}) \geq \frac{2\sigma^2}{NA^2}.$$

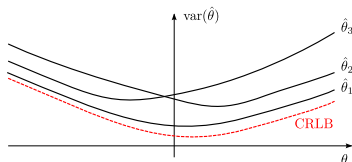
- No se cumple la condición para alcanzar la cota porque la derivada de la log-verosimilitud no se puede factorizar como  $I(\theta)(g(\mathbf{x}) - \theta)$ .
- Puede que exista un estimador MVU, pero no es posible determinar su existencia ni encontrarlo mediante el Teorema de la CRLB.

**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.
- $\theta_1$  es eficiente y MVU.



- Ningún estimador alcanza CRLB.
- Varianza de  $\hat{\theta}_1$  es menor que la de los otros estimad. insesgados.
- $\theta_1$  es MVU pero no es eficiente

**Definición (Información de Fisher).** La información de Fisher para los datos  $\mathbf{x} \sim p(\mathbf{x}; \theta)$  se define como:

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

Cuando un estimador alcanza la CRLB, su varianza es:

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

**Propiedades.**  $I(\theta)$  tiene las propiedades de una medida de información.

- Es **no-negativa**. Esto puede verse a partir de la siguiente igualdad

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

(Ejercicio, ver Apéndice 3A en Kay, Vol I.)

- Es **aditiva** para observaciones independientes. Si  $I(\theta)$  es la información de  $N$  observaciones IID e  $i(\theta)$  de una única observación, se tiene que  $I(\theta) = Ni(\theta)$ .

- La densidad de probabilidad de  $N$  observaciones IID cumple que

$$p(\mathbf{x}; \theta) = p(x[0], x[1], \dots, x[N-1]; \theta) = \prod_{n=0}^{N-1} p(x[n]; \theta).$$

- La **información de Fisher** es entonces:

$$I(\theta) = -\mathbb{E} \left[ \frac{\partial^2 \log p(\mathbf{x}; \theta)}{\partial \theta^2} \right] = -\sum_{n=0}^{N-1} \mathbb{E} \left[ \frac{\partial^2 \log p(x[n]; \theta)}{\partial \theta^2} \right] = Ni(\theta),$$

con  $i(\theta) = -\mathbb{E} \left[ \frac{\partial^2 \log p(x[n]; \theta)}{\partial \theta^2} \right]$  la información de Fisher de una muestra.

La CRLB al observar  $N$  muestras IID es  $N$  veces menor que al observar una muestra. En general,

- Independencia:  $I(\theta) = Ni(\theta)$
- No Independencia:  $I(\theta) < Ni(\theta)$
- Dependencia completa:  $I(\theta) = i(\theta)$

# CRLB general para señales con AWGN

## Señales con AWGN: Caso muy frecuente en la práctica

- Sea una señal determinística con un parámetro desconocido  $\theta$  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

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

- Tomando el logaritmo queda

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

- Diferenciando una vez se tiene que

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



# CRLB general para señales con AWGN

## Señales con AWGN: Caso muy frecuente en la práctica

- Diferenciando una segunda vez obtenemos

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

- y tomando el valor esperado,

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

donde en (a) se empleó que  $\mathbb{E}(x[n]) = s[n; \theta]$  (ruido tiene media nula).

- La CRLB es por lo tanto:

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

- Observar que se obtiene una mejor estimación cuando la señal  $s[n; \theta]$  cambia más rápidamente con el parámetro  $\theta$ .

## Ejemplo: Estimación de la frecuencia de una senoide

- Se considera una señal sinusoidal en AWGN y se quiere estimar su frecuencia. Es decir,

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

donde

$$s[n; f_0] = A \cos(2\pi f_0 n + \phi) \quad \text{con } 0 < f_0 < \frac{1}{2},$$

con la amplitud  $A$  y la fase  $\phi$  conocida.

- Usando la ecuación general de la CRLB en señales con AWGN, tenemos:

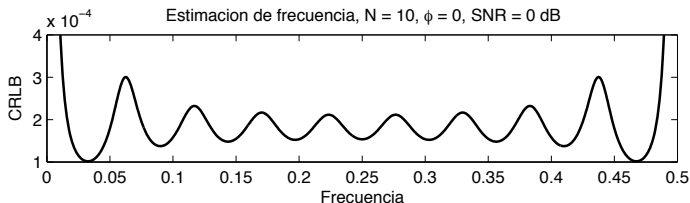
$$\text{var}(\hat{f}_0) \geq \frac{\sigma^2}{A^2 \sum_{n=0}^{N-1} [2\pi n \sin(2\pi f_0 n + \phi)]^2}.$$

# CRLB general para señales con AWGN

## Ejemplo: Estimación de la frecuencia de una senoide

### Observaciones:

- En la precisión del estimador hay frecuencias preferidas.
- Cuando  $f_0 \rightarrow 0$ ,  $\text{CRLB} \rightarrow \infty$ . Esto es porque para  $f_0 \approx 0$ , pequeños cambios en  $f_0$  no alteran la señal significativamente.
- Mediante el teorema de Cramér-Rao, se encontró una cota de la varianza de todo estimador insesgado, sin embargo no es posible encontrar el estimador en este ejemplo.



# Transformación de parámetros

- En la práctica es bastante común que querremos estimar una función de algún parámetro más fundamental.
- Nos interesa estimar un parámetro  $\alpha = f(\theta)$  estimando primero  $\theta$ .
- ¿Es suficiente con aplicar la transformación al resultado estimado  $\hat{\theta}$ ?
- Es, decir es  $\hat{\alpha} = f(\hat{\theta})$  un buen estimador de  $\alpha$ , si  $\hat{\theta}$  es buen estimador de  $\theta$ ?

## Ejemplo: Potencia de DC en WGN

- Supongamos que en lugar de estimar el de DC dado por  $A$  en WGN,

$$x[n] = A + w[n], \quad n = 0, 1, \dots, N-1 \quad \text{donde } w[n] \sim \mathcal{N}(0, \sigma^2),$$

nos interesa estimar la potencia  $A^2$ .

- Vimos que  $\hat{A} = \frac{1}{N} \sum_{n=0}^{N-1} x[n]$  es un estimador *eficiente* de  $A$ .
- ¿Es entonces  $\hat{A}^2$  buen estimador de  $A^2$ ? ¿Es eficiente?
- ¿Cuál es la CRLB para la estimación del parámetro  $A^2$ ?

# Transformación de parámetros

## Ejemplo: Potencia de DC en WGN

- Recordar que  $\hat{A}$  es *eficiente* y por lo tanto  $\text{var}(\hat{A}) = \text{CRLB}(A) = \sigma^2/N$ .
- La varianza se puede escribir como:

$$\text{var}(\hat{A}) = \mathbb{E}(\hat{A}^2) - \mathbb{E}(\hat{A})^2$$

Es decir:

$$\begin{aligned}\mathbb{E}(\hat{A}^2) &= \text{var}(\hat{A}) + \mathbb{E}(\hat{A})^2 \\ &= \frac{\sigma^2}{N} + A^2\end{aligned}$$

- Por lo tanto,  $\hat{A}^2$  ni siquiera es un estimador insesgado. de  $A^2$ .
- Una transformación **no lineal** destruye la eficiencia de un estimador.

**Teorema.** La CRLB para cualquier estimador insesgado de  $\alpha = f(\theta)$ , es

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

Demostración en Kay, 1993 (Apéndice 3A)

**Ejemplo:** Potencia de DC en WGN

En el caso del ejemplo,  $\alpha = f(A) = A^2$  y por lo tanto,

$$\text{var}(\widehat{A^2}) \geq \frac{(2A)^2}{N/\sigma^2} = \frac{4A^2\sigma^2}{N}.$$

# Transformación *afín* de parámetros

- Supongamos que  $\hat{\theta}$  es un estimador *eficiente* de  $\theta$  y queremos estimar  $\alpha = f(\theta) = a\theta + b$ . Cualquier estimador insesgado  $\hat{\alpha}$  de  $\alpha$  cumple,

$$\text{var}(\hat{\alpha}) \stackrel{(a)}{\geq} \frac{\left(\frac{\partial f}{\partial \theta}\right)^2}{I(\theta)}$$

(a) CRLB para  $\alpha = f(\theta)$ .

$$\stackrel{(b)}{=} \left(\frac{\partial f}{\partial \theta}\right)^2 \text{var}(\hat{\theta})$$

(b)  $\hat{\theta}$  es estimador *eficiente* de  $\theta$ .

$$\stackrel{(c)}{=} a^2 \text{var}(\hat{\theta})$$

(c) definición de  $\frac{\partial f}{\partial \theta}$ .

- Vamos a estudiar  $\widehat{f(\theta)} = a\hat{\theta} + b$  como estimador de  $\alpha = f(\theta)$

## Esperanza

$$\begin{aligned}\mathbb{E}(\widehat{f(\theta)}) &= \mathbb{E}(a\hat{\theta} + b) \\ &= a\mathbb{E}(\hat{\theta}) + b \\ &= a\theta + b = f(\theta)\end{aligned}$$

## Varianza

$$\begin{aligned}\text{var}(\widehat{f(\theta)}) &= \text{var}(a\hat{\theta} + b) \\ &= a^2 \text{var}(\hat{\theta})\end{aligned}$$

- Estimador  $\widehat{f(\theta)}$  es insesgado y alcanza la CRLB. Por lo tanto es eficiente.
- Transformaciones afines mantienen la eficiencia de los estimadores.

# 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  $\theta$  es **asintóticamente insesgado** si

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

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

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

## Ejemplo: Potencia de DC en WGN

- Previamente vimos que  $\hat{A}^2 = \left( \sum_{n=0}^{N-1} x[n] \right)^2$  es un estimador sesgado de  $A^2$ .
- Sin embargo, la esperanza es,

$$\mathbb{E}(\hat{A}^2) = A^2 + \frac{\sigma^2}{N} \xrightarrow{N \rightarrow \infty} A^2$$

por lo que  $\hat{A}^2$  es un estimador *asintóticamente insesgado* de  $A^2$ .



# Transformación de parámetros: caso asintótico

## Ejemplo: Potencia de DC en WGN

- Además, como  $\hat{A} \sim \mathcal{N}(A, \sigma^2/N)$ , podemos calcular su varianza

$$\text{var}(\hat{A}^2) = \mathbb{E}(\hat{A}^4) - \mathbb{E}^2(\hat{A}^2)$$

**Observación.** Si  $\zeta \sim \mathcal{N}(\mu, \sigma^2)$ , entonces

$$\mathbb{E}(\zeta^2) = \mu^2 + \sigma^2, \quad \mathbb{E}(\zeta^4) = \mu^4 + 6\mu^2\sigma^2 + 3\sigma^4,$$

y por lo tanto

$$\text{var}(\zeta^2) = \mathbb{E}(\zeta^4) - \mathbb{E}(\zeta^2)^2 = 4\mu^2\sigma^2 + 2\sigma^4.$$

Entonces,

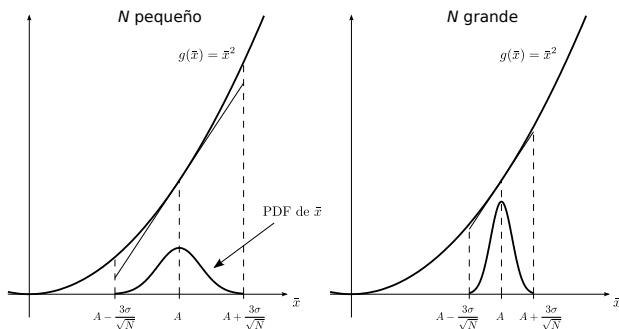
$$\text{var}(\hat{A}^2) = \frac{4A^2\sigma^2}{N} + 2\frac{\sigma^4}{N^2} \xrightarrow{N \rightarrow \infty} \frac{4A^2\sigma^2}{N} = CRLB_{\hat{A}^2}(A^2)$$

- Es decir,  $\hat{A}^2$  es un estimador **asintóticamente eficiente** de  $A^2$ .

# Transformación de parámetros: caso asintótico

## Linealidad estadística de una transformación no afín

- A medida que crece  $N$ , la PDF de  $\hat{A}$  se concentra alrededor de  $A$ .
- Valores observados de  $\hat{A}$  están en pequeño intervalo en torno de  $A$ .
- Transformación no afín es aproximadamente afín en intervalo pequeño.
- Los valores de  $\hat{A}$  en la región no afín ocurren raramente.



# 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, el estimador transformado es asintóticamente insesgado y eficiente.
- Es decir, cuando la cantidad de observaciones  $N \rightarrow \infty$  , el estimador limite 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).

# CRLB: Ejemplos (IV)

## Ejemplo: CRLB para nivel de DC en WGN

Como una extensión al Ejemplo II, se considera la observación de  $N$  muestras de continua en WGN,

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

pero ahora, además de desconocerse  $A$  también se desconoce  $\sigma^2$ .

- En este caso, el vector de parámetros a estimar es  $\boldsymbol{\theta} = [A, \sigma^2]^T$ , y  $p = 2$ .
- La matriz de información de Fisher (tamaño  $2 \times 2$ ),

$$\mathbf{I}(\boldsymbol{\theta}) = \begin{bmatrix} -\mathbb{E} \left[ \frac{\partial^2 \log p(\mathbf{x}; \boldsymbol{\theta})}{\partial (A)^2} \right] & -\mathbb{E} \left[ \frac{\partial^2 \log p(\mathbf{x}; \boldsymbol{\theta})}{\partial A \partial (\sigma^2)} \right] \\ -\mathbb{E} \left[ \frac{\partial^2 \log p(\mathbf{x}; \boldsymbol{\theta})}{\partial (\sigma^2) \partial A} \right] & -\mathbb{E} \left[ \frac{\partial^2 \log p(\mathbf{x}; \boldsymbol{\theta})}{\partial (\sigma^2)^2} \right] \end{bmatrix},$$

es simétrica y semi-definida positiva (ejercicio).

# CRLB: Ejemplos (IV)

## Ejemplo: CRLB para nivel de DC en WGN

- La función de log-verosimilitud, es

$$\log p(\mathbf{x}; \boldsymbol{\theta}) = -\frac{N}{2} \log 2\pi - \frac{N}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x[n] - A)^2.$$

- Las respectivas derivadas son (ejercicio),

$$\frac{\partial^2 \log p(\mathbf{x}; \boldsymbol{\theta})}{\partial(A)^2} = -\frac{N}{\sigma^2}$$

$$\frac{\partial^2 \log p(\mathbf{x}; \boldsymbol{\theta})}{\partial \partial A(\sigma^2)} = -\frac{1}{\sigma^4} \sum_{n=0}^{N-1} (x[n] - A)$$

$$\frac{\partial^2 \log p(\mathbf{x}; \boldsymbol{\theta})}{\partial(\sigma^2)^2} = \frac{N}{2\sigma^4} - \frac{1}{\sigma^6} \sum_{n=0}^{N-1} (x[n] - A)^2.$$

# CRLB: Ejemplos (IV)

## Ejemplo: CRLB para nivel de DC en WGN

- Tomando el negativo de la esperanza, se construye la matriz de Fisher

$$\mathbf{I}(\boldsymbol{\theta}) = \begin{bmatrix} \frac{N}{\sigma^2} & 0 \\ 0 & \frac{N}{2\sigma^4} \end{bmatrix}.$$

- En este caso la Matriz de Fisher es diagonal (en general no es así), por lo que invertirla es directo

$$\mathbf{I}(\boldsymbol{\theta})^{-1} = \begin{bmatrix} \frac{\sigma^2}{N} & 0 \\ 0 & \frac{2\sigma^4}{N} \end{bmatrix}.$$

- Entonces la CRLB para  $\hat{A}$  y  $\widehat{\sigma^2}$ ,

$$\text{var}(\hat{A}) \geq \frac{\sigma^2}{N}, \quad \text{var}(\widehat{\sigma^2}) \geq \frac{2\sigma^4}{N}.$$

## Observaciones:

- La CRLB de  $\hat{A}$  es la misma que en el caso en que  $\sigma^2$  es conocido (Ejemplo II).
- Este es un caso particular (matriz de Fisher diagonal), en general no es así. En este caso la información de un parámetro respecto del otro en media es nula.

# 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 (ejercicio), la cota matricial implica,

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

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

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

- 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.

## Regla de Integración de Leibniz

$$\frac{d}{d\theta} \int_{a(\theta)}^{b(\theta)} f(x, \theta) dx = \int_{a(\theta)}^{b(\theta)} \frac{\partial f(x, \theta)}{\partial \theta} dx + f(b(\theta), \theta) \cdot b'(\theta) - f(a(\theta), \theta) \cdot a'(\theta).$$

- De la regla surge que el orden de derivación e integración puede cambiarse si los límites de integración no dependen de  $\theta$ .
- Es decir, si  $a(\theta) = a$  y  $b(\theta) = b$ , entonces,

$$\frac{d}{d\theta} \int_a^b f(x, \theta) dx = \int_a^b \frac{\partial f(x, \theta)}{\partial \theta} dx.$$

- Notar que, por ejemplo, la distribución  $\mathcal{U}[0, \theta]$  va a generar inconvenientes.

## Covarianza

- La covarianza entre dos variables aleatorias  $x \in \mathbb{R}$  e  $y \in \mathbb{R}$  se define como:

$$\begin{aligned}\text{cov}(x, y) &= \mathbb{E}[(x - \mathbb{E}(x))(y - \mathbb{E}(y))] \\ &= \mathbb{E}[xy] - \mathbb{E}[x]\mathbb{E}[y].\end{aligned}$$

- Es una medida de la dependencia entre variables aleatorias.
- Dos variables aleatorias  $x$  e  $y$  se dicen no correlacionadas si  $\text{cov}(x, y) = 0$ .
- Si dos variables aleatorias  $x$  e  $y$  son independientes se cumple que,

$$\begin{aligned}\text{cov}(x, y) &= \mathbb{E}[xy] - \mathbb{E}[x]\mathbb{E}[y] \\ &= \mathbb{E}[x]\mathbb{E}[y] - \mathbb{E}[x]\mathbb{E}[y] \\ &= 0\end{aligned}$$

Recíproco no es cierto (salvo si  $(x, y)$  tiene distribución normal bivalente).

## Matriz de Covarianza

- Sea el vector de variables aleatorias  $\mathbf{x} = [x_1, x_1, \dots, x_n]^T$ , la matriz de covarianza se define como

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

donde la esperanza de un vector (matriz) con entradas aleatorias se define como el vector (matriz) con la esperanza de cada variable aleatoria.

- Ejemplo  $\mathbf{x} = [x_1, x_2]^T$ ,

$$\begin{aligned} \mathbf{C} &= \begin{bmatrix} \mathbb{E}[(x_1 - \mu_1)(x_1 - \mu_1)] & \mathbb{E}[(x_1 - \mu_1)(x_2 - \mu_2)] \\ \mathbb{E}[(x_2 - \mu_2)(x_1 - \mu_1)] & \mathbb{E}[(x_2 - \mu_2)(x_2 - \mu_2)] \end{bmatrix} \\ &= \begin{bmatrix} \text{var}(x_1) & \text{cov}(x_1, x_2) \\ \text{cov}(x_1, x_2) & \text{var}(x_2) \end{bmatrix} \end{aligned}$$

- **Kay, S. M.** (1993)  
*Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory*, Capítulo 3.