Ingeniería Eléctrica
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS UNIVERSIDAD DE CHILE

a

Análisis de Sistemas Dinámicos y Estimacióna

EL3204-1

Pauta Auxiliar 11 - Estimación

Prof Marcos Orchard - Sebastian Espinoza.
Prof Auxiliar Erik Sáez Aravena.

$\mathcal{I}(\theta)$

Resumen

Este auxiliar cubre conceptos fundamentales de la teoría de estimación de parámetros, que complementa los temas de detección vistos anteriormente. En problemas de detección, el espacio paramétrico Θ era discreto y finito, pero ahora trabajamos con un conjunto **infinito no numerable** de posibles valores para el parámetro que queremos estimar (un continuo).

1. Conceptos Básicos de Estimación

Algunos conceptos importante a considerar son:

1. **Espacio de observación ($\mathbb{X}$):** Espacio donde la variable aleatoria $X \in \mathbb{X}$ toma valores. Para vectores aleatorios, $\mathbb{X}^n$ representa el espacio producto.
2. **Espacio de decisión (Θ):** En estimación, es un conjunto **infinito no numerable** de valores (a diferencia de detección donde era finito). Ejemplos: $\Theta = \mathbb{R}^+$ para estimar amplitudes, $\Theta = \mathbb{R}$ para estimar medias.
3. **Familia de distribuciones paramétricas (J_θ):** Conjunto de distribuciones indexadas por θ:
$$J_\theta = \{P_X(x|\theta) : \theta \in \Theta\}$$
4. **Estimador:** Una función $\hat{\theta}(x_1, \ldots, x_n)$ que depende de las observaciones y entrega una estimación del parámetro θ asociado a la distribución $P_X(x|\theta)$.

2. Propiedades de los Estimadores

1. **Estimador Insesgado:** Un estimador $\hat{\theta}(x_1, \ldots, x_n)$ es insesgado si:
$$\mathbb{E}[\hat{\theta}(X_1, \ldots, X_n)] = \theta$$
Es decir, en promedio el estimador entrega el valor verdadero del parámetro.
2. **Estimador Asintóticamente Insesgado:** Si el sesgo tiende a cero cuando $n \to \infty$:
$$\lim_{n\to\infty} \mathbb{E}[\hat{\theta}(X_1, \ldots, X_n)] = \theta$$
3. **Estimador Consistente:** Un estimador es consistente si converge en probabilidad al valor verdadero:
$$\lim_{n\to\infty} P_X(|\hat{\theta}(X_1, \ldots, X_n) - \theta| > \epsilon) = 0 \quad \forall \epsilon > 0$$
4. **Observación:** Mediante la desigualdad de Chebyshev, si un estimador es insesgado y su varianza tiende a cero cuando $n \to \infty$, entonces es consistente:
$$\lim_{n\to\infty} \mathrm{Var}(\hat{\theta}(X_1, \ldots, X_n)) = 0 \implies \text{consistencia}$$

3. Condiciones de Regularidad

Las condiciones de regularidad permiten intercambiar el orden de derivación e integración en la función de verosimilitud. Bajo estas condiciones se cumple:

1. **Primera condición:** La esperanza de la derivada de la log-verosimilitud es cero:

$$\mathbb{E}_{X_1^n}\left[\frac{\partial \ln L(X_1,\ldots,X_n|\theta)}{\partial\theta}\right]=0$$

2. **Segunda condición:** La esperanza de un término específico es cero:

$$\mathbb{E}_{X_1^n}\left[\frac{1}{L(X_1,\ldots,X_n|\theta)}\frac{\partial^2 L(X_1,\ldots,X_n|\theta)}{\partial\theta^2}\right]=0$$

3. **Identidad de Bartlett:** Relaciona la varianza con su segunda derivada:

$$\mathbb{E}_{X_1^n}\left[\left(\frac{\partial \ln L}{\partial\theta}\right)^2\right]=-\mathbb{E}_{X_1^n}\left[\frac{\partial^2 \ln L}{\partial\theta^2}\right]$$

   Esta identidad permite calcular la información de Fisher de dos formas equivalentes.

4. Información de Fisher

La información de Fisher $\mathcal{I}(\theta)$ cuantifica cuánta información sobre el parámetro θ contienen las observaciones:

$$\mathcal{I}(\theta)=\mathbb{E}\left[\left(\frac{\partial \ln f(X|\theta)}{\partial\theta}\right)^2\right]=-\mathbb{E}\left[\frac{\partial^2 \ln f(X|\theta)}{\partial\theta^2}\right]$$

Propiedad Aditiva: Para muestras i.i.d., la información de Fisher es aditiva:

$$\mathcal{I}_n(\theta)=n\cdot\mathcal{I}_1(\theta)$$

Esta propiedad refleja que más observaciones independientes proporcionan más información sobre el parámetro.

5. Cota de Cramér-Rao

La cota de Cramér-Rao establece un límite inferior para la varianza de cualquier estimador insesgado:

$$\mathrm{Var}(\hat{\theta})\geq\frac{1}{\mathcal{I}(\theta)}$$

1. **Estimador Eficiente:** Un estimador insesgado que alcanza la cota de Cramér-Rao se llama **eficiente** o **de mínima varianza**

2. **Interpretación:** La información de Fisher es inversamente proporcional a la mínima varianza posible. Más información $\implies$ menor varianza mínima.

6. Estimador de Máxima Verosimilitud (ML)

El estimador de máxima verosimilitud se obtiene maximizando la función de verosimilitud:

$$\hat{\theta}_{ML} = \arg\max_{\theta} L(x_1, \ldots, x_n|\theta) = \arg\max_{\theta} \ln L(x_1, \ldots, x_n|\theta)$$

Propiedades del estimador ML:

- Bajo condiciones de regularidad, es asintóticamente insesgado
- Es consistente
- Es asintóticamente eficiente (alcanza la cota de Cramér-Rao cuando $n \to \infty$)
- Es invariante bajo transformaciones: si $\hat{\theta}_{ML}$ es el ML de θ, entonces $g(\hat{\theta}_{ML})$ es el ML de $g(\theta)$

Método de solución: Criterio de la primera derivada:

$$\left.\frac{\partial \ln L(x_1, \ldots, x_n|\theta)}{\partial \theta}\right|_{\theta=\hat{\theta}_{ML}} = 0$$

Verificar que la segunda derivada es negativa para confirmar que es un máximo.

7. Resumen de Verificaciones para un Estimador

Para analizar completamente un estimador $\hat{\theta}$, verificar:

1. **Insesgamiento:** Calcular $\mathbb{E}[\hat{\theta}]$ y verificar si es igual a θ
2. **Consistencia:** Verificar que $\lim_{n\to\infty} \mathrm{Var}(\hat{\theta}) = 0$ (si es insesgado)
3. **Varianza:** Calcular $\mathrm{Var}(\hat{\theta})$ explícitamente
4. **Eficiencia:** Calcular $\mathcal{I}(\theta)$ y comparar $\mathrm{Var}(\hat{\theta})$ con $\frac{1}{\mathcal{I}(\theta)}$

1. Bajo las condiciones de regularidad de la Definición 3.2 del apunte, verifique si el vector aleatorio $X_1, \ldots, X_n$ es i.i.d. con distribución $P_{X_1^n}(\cdot|\theta)$ entoncesa

$$\mathbb{E}_{X_1^n}\left(\left(\frac{\partial \ln L(X_1,\ldots,X_n|\theta)}{\partial\theta}\right)^2\right) = -\mathbb{E}_{X_1^n}\left(\frac{\partial^2 \ln L(X_1,\ldots,X_n|\theta)}{\partial\theta^2}\right). \tag{1}$$

Solución:

Para demostrar esta identidad, utilizaremos las siguientes condiciones de regularidad que permiten intercambiar el orden de derivación e integración:

Primera condición de regularidad: La esperanza de la primera derivada de la log-verosimilitud es cero:

$$\mathbb{E}_{X_1^n}\left[\frac{\partial \ln L(X_1,\ldots,X_n|\theta)}{\partial\theta}\right] = 0 \tag{2}$$

Segunda condición de regularidad: La esperanza de un término específico es cero:

$$\mathbb{E}_{X_1^n}\left[\frac{1}{L(X_1,\ldots,X_n|\theta)}\frac{\partial^2 L(X_1,\ldots,X_n|\theta)}{\partial\theta^2}\right] = 0 \tag{3}$$

Partimos de la primera condición de regularidad:

$$\mathbb{E}_{X_1^n}\left[\frac{\partial \ln L(X_1,\ldots,X_n|\theta)}{\partial\theta}\right] = 0$$

Y derivamos ambos lados de esta ecuación respecto a θ. Del lado izquierdo, bajo condiciones de regularidad podemos intercambiar la derivada con la esperanza (que es una integral):

$$\frac{\partial}{\partial\theta}\underbrace{\mathbb{E}_{X_1^n}\left[\frac{\partial \ln L(X_1,\ldots,X_n|\theta)}{\partial\theta}\right]}_{=0} = \mathbb{E}_{X_1^n}\left[\frac{\partial^2 \ln L(X_1,\ldots,X_n|\theta)}{\partial\theta^2}\right]$$

Del lado derecho, derivamos la constante 0:

$$\frac{\partial}{\partial\theta}(0) = 0$$

Por lo tanto:

$$\mathbb{E}_{X_1^n}\left[\frac{\partial^2 \ln L(X_1,\ldots,X_n|\theta)}{\partial\theta^2}\right] = 0$$

Este resultado nos dice que la esperanza de la segunda derivada de la log-verosimilitud es cero. Ahora necesitamos conectar estos dos resultados. Partimos de la propiedad fundamental del logaritmo aplicada a la función de verosimilitud:

$$\frac{\partial \ln L(X_1,\ldots,X_n|\theta)}{\partial\theta} = \frac{1}{L(X_1,\ldots,X_n|\theta)}\frac{\partial L(X_1,\ldots,X_n|\theta)}{\partial\theta}$$

Ahora derivamos ambos lados respecto a θ para obtener la segunda derivada. Aplicamos la regla del producto en el lado derecho, donde tenemos un producto de dos funciones: $\frac{1}{L(X_1,\ldots,X_n|\theta)}$ y $\frac{\partial L(X_1,\ldots,X_n|\theta)}{\partial\theta}$.

Recordemos la regla del producto: si $h(\theta) = f(\theta)\cdot g(\theta)$, entonces $\frac{dh}{d\theta} = \frac{df}{d\theta}\cdot g + f\cdot\frac{dg}{d\theta}$.

Aplicando esto a lo siguiente:

$$\frac{\partial^2 \ln L(X_1,\ldots,X_n|\theta)}{\partial\theta^2} = \frac{\partial}{\partial\theta}\left[\frac{1}{L(X_1,\ldots,X_n|\theta)}\frac{\partial L(X_1,\ldots,X_n|\theta)}{\partial\theta}\right]$$

Desarrollamos usando la regla del producto:

$$= \underbrace{\frac{\partial}{\partial\theta}\left(\frac{1}{L(X_1,\ldots,X_n|\theta)}\right)}_{\text{derivada del primer factor}}\cdot\frac{\partial L(X_1,\ldots,X_n|\theta)}{\partial\theta} + \frac{1}{L(X_1,\ldots,X_n|\theta)}\cdot\underbrace{\frac{\partial^2 L(X_1,\ldots,X_n|\theta)}{\partial\theta^2}}_{\text{derivada del segundo factor}}$$

Para el primer término, calculamos la derivada de $\frac{1}{L(X_1,\ldots,X_n|\theta)}$ usando la regla de la cadena:

$$\frac{\partial}{\partial\theta}\left(\frac{1}{L(X_1,\ldots,X_n|\theta)}\right) = -\frac{1}{L(X_1,\ldots,X_n|\theta)^2}\frac{\partial L(X_1,\ldots,X_n|\theta)}{\partial\theta}$$

Sustituyendo esto en la expresión anterior:

$$\frac{\partial^2 \ln L(X_1,\ldots,X_n|\theta)}{\partial\theta^2} = -\frac{1}{L(X_1,\ldots,X_n|\theta)^2}\left(\frac{\partial L(X_1,\ldots,X_n|\theta)}{\partial\theta}\right)^2 + \frac{1}{L(X_1,\ldots,X_n|\theta)}\frac{\partial^2 L(X_1,\ldots,X_n|\theta)}{\partial\theta^2}$$

Ahora observamos que el primer término se puede reescribir. Notemos que:

$$\frac{1}{L(X_1,\ldots,X_n|\theta)}\frac{\partial L(X_1,\ldots,X_n|\theta)}{\partial\theta} = \frac{\partial \ln L(X_1,\ldots,X_n|\theta)}{\partial\theta}$$

Por lo tanto:

$$\frac{1}{L(X_1,\ldots,X_n|\theta)^2}\left(\frac{\partial L(X_1,\ldots,X_n|\theta)}{\partial\theta}\right)^2 = \left(\frac{\partial \ln L(X_1,\ldots,X_n|\theta)}{\partial\theta}\right)^2$$

Finalmente, sustituyendo esto obtenemos:

$$\boxed{\frac{\partial^2 \ln L(X_1,\ldots,X_n|\theta)}{\partial\theta^2} = -\left(\frac{\partial \ln L(X_1,\ldots,X_n|\theta)}{\partial\theta}\right)^2 + \frac{1}{L(X_1,\ldots,X_n|\theta)}\frac{\partial^2 L(X_1,\ldots,X_n|\theta)}{\partial\theta^2}}$$

Tomando esperanza en ambos lados y usando que $\mathbb{E}\left[\frac{1}{L(X_1,\ldots,X_n|\theta)}\frac{\partial^2 L(X_1,\ldots,X_n|\theta)}{\partial\theta^2}\right] = 0$ (otra condición de regularidad):

$$\mathbb{E}_{X_1^n}\left[\frac{\partial^2 \ln L(X_1,\ldots,X_n|\theta)}{\partial\theta^2}\right] = -\mathbb{E}_{X_1^n}\left[\left(\frac{\partial \ln L(X_1,\ldots,X_n|\theta)}{\partial\theta}\right)^2\right]$$

2. Compruebe que si el vector aleatorio $X_1, \ldots, X_n$ es i.i.d. con distribución $P_{X_1^n}(\cdot|\theta)$ entonces la información de Fisher es aditiva, es decir:

$$\mathcal{I}_n(\theta) \triangleq \mathbb{E}_{X_1^n}\left(\left(\frac{\partial \ln L(X_1, X_2, \ldots, X_n|\theta)}{\partial\theta}\right)^2\right) = n \cdot \mathcal{I}_1(\theta). \tag{4}$$

Solución:

Dado que $X_1, \ldots, X_n$ son variables i.i.d., la función de verosimilitud se puede expresar como el producto de las densidades individuales:

$$L(X_1, \ldots, X_n|\theta) = \prod_{i=1}^{n} f(X_i|\theta)$$

Tomando logaritmo en ambos lados:

$$\ln L(X_1, \ldots, X_n|\theta) = \sum_{i=1}^{n} \ln f(X_i|\theta)$$

Derivando respecto a θ y usando la linealidad de la derivada:

$$\frac{\partial}{\partial\theta} \ln L(X_1, \ldots, X_n|\theta) = \sum_{i=1}^{n} \frac{\partial \ln f(X_i|\theta)}{\partial\theta}$$

Ahora calculamos la esperanza del cuadrado de esta derivada. Notemos que:

$$\mathbb{E}_{X_1^n}\left[\left(\frac{\partial \ln L}{\partial\theta}\right)^2\right] = \mathbb{E}_{X_1^n}\left[\left(\sum_{i=1}^{n} \frac{\partial \ln f(X_i|\theta)}{\partial\theta}\right)^2\right]$$

Expandiendo el cuadrado:

$$= \mathbb{E}_{X_1^n}\left[\sum_{i=1}^{n}\left(\frac{\partial \ln f(X_i|\theta)}{\partial\theta}\right)^2 + 2\sum_{i<j} \frac{\partial \ln f(X_i|\theta)}{\partial\theta} \cdot \frac{\partial \ln f(X_j|\theta)}{\partial\theta}\right]$$

Por linealidad de la esperanza:

$$= \sum_{i=1}^{n} \mathbb{E}_{X_i}\left[\left(\frac{\partial \ln f(X_i|\theta)}{\partial\theta}\right)^2\right] + 2\sum_{i<j} \mathbb{E}_{X_i,X_j}\left[\frac{\partial \ln f(X_i|\theta)}{\partial\theta} \cdot \frac{\partial \ln f(X_j|\theta)}{\partial\theta}\right]$$

Como las variables son independientes, la esperanza del producto es el producto de las esperanzas. Por las condiciones de regularidad, $\mathbb{E}_{X_i}\left[\frac{\partial \ln f(X_i|\theta)}{\partial\theta}\right] = 0$, por lo que todos los términos cruzados se anulan:

$$2\sum_{i<j} \mathbb{E}_{X_i}\left[\frac{\partial \ln f(X_i|\theta)}{\partial\theta}\right] \cdot \mathbb{E}_{X_j}\left[\frac{\partial \ln f(X_j|\theta)}{\partial\theta}\right] = 0$$

Por lo tanto:

$$\mathcal{I}_n(\theta) = \sum_{i=1}^{n} \mathbb{E}_{X_i}\left[\left(\frac{\partial \ln f(X_i|\theta)}{\partial\theta}\right)^2\right]$$

Como todas las variables son idénticamente distribuidas, cada término es igual a $\mathcal{I}_1(\theta)$:

$$\mathcal{I}_n(\theta) = n \cdot \mathcal{I}_1(\theta)$$

3. [**Propuesto**]Muestre que para cualquier estimador τ_n de θ su error de estimación se puede descomponer como varianza más sesgo, es decir:

$$\mathbb{E}_{X_1^n}\left((\tau_n(X_1^n) - \theta)^2\right) = \mathrm{Var}(\tau_n(X_1^n)) + \left(\mathbb{E}_{X_1^n}(\tau_n(X_1^n)) - \theta\right)^2. \tag{5}$$

Solución:

Para demostrar esta descomposición, partimos del error cuadrático medio (MSE) y realizamos manipulaciones algebraicas. Sea $\hat{\theta} = \tau_n(X_1^n)$ el estimador y denotemos $\mathbb{E}[\hat{\theta}]$ como su esperanza.Comenzamos con el MSE:

$$\mathrm{MSE}(\hat{\theta}) = \mathbb{E}_{X_1^n}\left[(\hat{\theta} - \theta)^2\right]$$

Sumamos y restamos la esperanza del estimador $\mathbb{E}[\hat{\theta}]$ dentro del cuadrado:

$$\mathrm{MSE}(\hat{\theta}) = \mathbb{E}_{X_1^n}\left[\left((\hat{\theta} - \mathbb{E}[\hat{\theta}]) + (\mathbb{E}[\hat{\theta}] - \theta)\right)^2\right]$$

Expandiendo el cuadrado del binomio:

$$\mathrm{MSE}(\hat{\theta}) = \mathbb{E}_{X_1^n}\left[(\hat{\theta} - \mathbb{E}[\hat{\theta}])^2 + 2(\hat{\theta} - \mathbb{E}[\hat{\theta}])(\mathbb{E}[\hat{\theta}] - \theta) + (\mathbb{E}[\hat{\theta}] - \theta)^2\right]$$

Por la linealidad de la esperanza, podemos separar en tres términos:

$$\mathrm{MSE}(\hat{\theta}) = \mathbb{E}_{X_1^n}\left[(\hat{\theta} - \mathbb{E}[\hat{\theta}])^2\right] + \mathbb{E}_{X_1^n}\left[2(\hat{\theta} - \mathbb{E}[\hat{\theta}])(\mathbb{E}[\hat{\theta}] - \theta)\right] + \mathbb{E}_{X_1^n}\left[(\mathbb{E}[\hat{\theta}] - \theta)^2\right]$$

Analicemos cada término:

- **Primer término:** Este es la definición de la varianza del estimador:

$$\mathbb{E}_{X_1^n}\left[(\hat{\theta} - \mathbb{E}[\hat{\theta}])^2\right] = \mathrm{Var}(\hat{\theta})$$

- **Tercer término:** Como $\mathbb{E}[\hat{\theta}]$ y θ son constantes (no dependen de X_1^n), la esperanza no afecta este término:

$$\mathbb{E}_{X_1^n}\left[(\mathbb{E}[\hat{\theta}] - \theta)^2\right] = (\mathbb{E}[\hat{\theta}] - \theta)^2$$

  Este término representa el cuadrado de la diferencia entre el valor esperado del estimador y el valor verdadero del parámetro.

- **Segundo término:** Factorizamos las constantes fuera de la esperanza:

$$\mathbb{E}_{X_1^n}\left[2(\hat{\theta}-\mathbb{E}[\hat{\theta}])(\mathbb{E}[\hat{\theta}]-\theta)\right] = 2(\mathbb{E}[\hat{\theta}]-\theta)\cdot\mathbb{E}_{X_1^n}\left[\hat{\theta}-\mathbb{E}[\hat{\theta}]\right]$$

Pero sabemos que:

$$\mathbb{E}_{X_1^n}\left[\hat{\theta}-\mathbb{E}[\hat{\theta}]\right] = \mathbb{E}_{X_1^n}[\hat{\theta}]-\mathbb{E}[\hat{\theta}] = \mathbb{E}[\hat{\theta}]-\mathbb{E}[\hat{\theta}] = 0$$

Por lo tanto, el segundo término es cero:

$$2(\mathbb{E}[\hat{\theta}]-\theta)\cdot 0 = 0$$

Combinando todos los términos, obtenemos la descomposición deseada:

$$\mathrm{MSE}(\hat{\theta}) = \mathrm{Var}(\hat{\theta}) + (\mathbb{E}[\hat{\theta}]-\theta)^2$$

O equivalentemente:

$$\mathbb{E}_{X_1^n}\left[(\tau_n(X_1^n)-\theta)^2\right] = \mathrm{Var}(\tau_n(X_1^n)) + \left(\mathbb{E}_{X_1^n}(\tau_n(X_1^n))-\theta\right)^2$$

Esta descomposición es fundamental en teoría de estimación, pues muestra que el error cuadrático medio se puede descomponer en dos componentes: la varianza del estimador (que mide su dispersión) y el cuadrado de la diferencia entre su valor esperado y el valor verdadero del parámetro (que mide qué tan lejos está en promedio del valor real).

4. Suponga una variable aleatoria $X \sim \mathcal{N}(\mu,\sigma^2)$ y muestras de dicha variable a $X_1, X_2, \ldots, X_n$ i.i.d. Popularmente, se suelen estimar los dos parámetros de esta distribución como:

$$\hat{\mu} = \frac{\sum_{i=1}^n x_i}{n} \tag{6}$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n (x_i-\mu)^2 \tag{7}$$

(a) ¿Son los estimadores de ambos parámetros insesgados?

Solución:

1. **Para $\hat{\mu}$:**

   Calculamos la esperanza del estimador:

$$\mathbb{E}[\hat{\mu}] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n X_i\right] = \frac{1}{n}\sum_{i=1}^n \mathbb{E}[X_i] = \frac{1}{n}\sum_{i=1}^n \mu = \frac{n\mu}{n} = \mu$$

   Por lo tanto, $\hat{\mu}$ es un **estimador insesgado** de μ.

2. **Para $\hat{\sigma}^2$:**

   Calculamos la esperanza del estimador:

$$\mathbb{E}[\hat{\sigma}^2] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2\right] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[(X_i-\mu)^2]$$

   Sabemos que $\mathbb{E}[(X_i-\mu)^2] = \mathrm{Var}(X_i) = \sigma^2$, entonces:

$$\mathbb{E}[\hat{\sigma}^2] = \frac{1}{n}\sum_{i=1}^{n}\sigma^2 = \frac{n\sigma^2}{n} = \sigma^2$$

   Por lo tanto, $\hat{\sigma}^2$ también es un **estimador insesgado** de σ^2.

Por lo tanto ambos estimadores son insesgados.

(b) Determine si los estimadores son consistentes. *Indicación:* Utilice el hecho que si $Z \sim \mathcal{N}(0,1)$ entonces $Y = Z^2 \sim \chi^2_{(1)}$ con $\mathrm{Var}(Y) = 2 \cdot 1$.

Solución:
Un estimador $\hat{\theta}_n$ es consistente si $\lim_{n\to\infty} \mathbb{P}(|\hat{\theta}_n - \theta| > \epsilon) = 0$ para todo $\epsilon > 0$. Por la desigualdad de Chebyshev, si el sesgo tiende a cero y la varianza tiende a cero cuando $n \to \infty$, entonces el estimador es consistente.

- **Para $\hat{\mu}$:**

  Ya vimos que el sesgo es cero. Ahora calculamos la varianza:

$$\mathrm{Var}(\hat{\mu}) = \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}(X_i) = \frac{1}{n^2}\cdot n\sigma^2 = \frac{\sigma^2}{n}$$

  Como $\lim_{n\to\infty}\frac{\sigma^2}{n} = 0$, el estimador $\hat{\mu}$ es **consistente**.

- **Para $\hat{\sigma}^2$:**

  El sesgo es cero. Ahora calculamos la varianza. Definimos $Z_i = \frac{X_i-\mu}{\sigma} \sim \mathcal{N}(0,1)$, entonces:

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2 = \frac{\sigma^2}{n}\sum_{i=1}^{n}Z_i^2$$

  Como $Z_i^2 \sim \chi^2_{(1)}$, tenemos $\mathrm{Var}(Z_i^2) = 2$. Las variables son independientes, entonces:

$$\mathrm{Var}(\hat{\sigma}^2) = \mathrm{Var}\left(\frac{\sigma^2}{n}\sum_{i=1}^{n}Z_i^2\right) = \frac{\sigma^4}{n^2}\sum_{i=1}^{n}\mathrm{Var}(Z_i^2) = \frac{\sigma^4}{n^2}\cdot 2n = \frac{2\sigma^4}{n}$$

  Como $\lim_{n\to\infty}\frac{2\sigma^4}{n} = 0$, el estimador $\hat{\sigma}^2$ es **consistente**.

Ambos estimadores son consistentes.

(c) Obtenga la cota de Cramer-Rao de cada parámetro ¿Se esta cumpliendo el criterio de mínima varianza?

Solución:
Para obtener la cota de Cramér-Rao, necesitamos calcular la información de Fisher. La densidad de probabilidad de una variable $X \sim \mathcal{N}(\mu, \sigma^2)$ es:

$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

La log-verosimilitud para n observaciones i.i.d. es:

$$\ln L(\mu, \sigma^2) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

- **Para μ:**

  Calculamos la primera derivada:

  $$\frac{\partial \ln L}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)$$

  La segunda derivada:

  $$\frac{\partial^2 \ln L}{\partial \mu^2} = -\frac{n}{\sigma^2}$$

  La información de Fisher es:

  $$\mathcal{I}_n(\mu) = -\mathbb{E}\left[\frac{\partial^2 \ln L}{\partial \mu^2}\right] = \frac{n}{\sigma^2}$$

  La cota de Cramér-Rao para μ es:

  $$\mathrm{Var}(\hat{\mu}) \geq \frac{1}{\mathcal{I}_n(\mu)} = \frac{\sigma^2}{n}$$

  Como $\mathrm{Var}(\hat{\mu}) = \frac{\sigma^2}{n}$, el estimador $\hat{\mu}$ **alcanza la cota de Cramér-Rao** y es de **mínima varianza**.

- **Para σ^2:**

  Calculamos la primera derivada:

  $$\frac{\partial \ln L}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i-\mu)^2$$

  La segunda derivada:

  $$\frac{\partial^2 \ln L}{\partial (\sigma^2)^2} = \frac{n}{2\sigma^4} - \frac{1}{\sigma^6}\sum_{i=1}^{n}(x_i-\mu)^2$$

Tomando esperanza:

$$\mathbb{E}\left[\frac{\partial^2 \ln L}{\partial(\sigma^2)^2}\right] = \frac{n}{2\sigma^4} - \frac{1}{\sigma^6} \cdot n\sigma^2 = \frac{n}{2\sigma^4} - \frac{n}{\sigma^4} = -\frac{n}{2\sigma^4}$$

La información de Fisher es:

$$\mathcal{I}_n(\sigma^2) = -\mathbb{E}\left[\frac{\partial^2 \ln L}{\partial(\sigma^2)^2}\right] = \frac{n}{2\sigma^4}$$

La cota de Cramér-Rao para σ^2 es:

$$\text{Var}(\hat{\sigma}^2) \geq \frac{1}{\mathcal{I}_n(\sigma^2)} = \frac{2\sigma^4}{n}$$

Como $\text{Var}(\hat{\sigma}^2) = \frac{2\sigma^4}{n}$, el estimador $\hat{\sigma}^2$ también **alcanza la cota de Cramér-Rao** y es de **mínima varianza**.

Ambos estimadores cumplen el criterio de mínima varianza.

5. Considere un sistema de modulación AM que genera una señal discreta

$$s(k) = A \cdot \cos\left(\frac{2\pi}{T} \cdot k\right) \tag{8}$$

donde $k \in \{1, \ldots, n\}$, que depende del parámetro A y donde $T > 0$ es un número entero conocido. $(s(k))_{k=1,\ldots,n}$ no es observable directamente si no por medio de un ruido aditivo:

$$X_k(w) = s(k) + N_k(w) \tag{9}$$

Donde $N_1(w), N_2(w), \ldots, N_n(w)$ son variables aleatorias gaussianas independientes e idénticamente distribuidas con media cero y varianza $\sigma^2 > 0$.

(a) Notar que $X_1, \ldots, X_n$ son variables aleatorias gaussianas e independientes. Con ello determine su distribución conjunta $f_X(x_1, \ldots, x_k|A)$.

Solución:
Dado que $X_k = s(k) + N_k$ donde $N_k \sim \mathcal{N}(0, \sigma^2)$, entonces:

$$X_k \sim \mathcal{N}\left(s(k), \sigma^2\right) = \mathcal{N}\left(A\cos\left(\frac{2\pi k}{T}\right), \sigma^2\right)$$

La densidad de probabilidad de cada X_k es:

$$f_{X_k}(x_k|A) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\left(x_k - A\cos\left(\frac{2\pi k}{T}\right)\right)^2}{2\sigma^2}\right)$$

Como las variables son independientes, la distribución conjunta es el producto de las densidades marginales:

$$f_X(x_1,\ldots,x_n|A) = \prod_{k=1}^{n} f_{X_k}(x_k|A) = \prod_{k=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\left(x_k - A\cos\left(\frac{2\pi k}{T}\right)\right)^2}{2\sigma^2}\right)$$

Simplificando:

$$f_X(x_1,\ldots,x_n|A) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{k=1}^{n}\left(x_k - A\cos\left(\frac{2\pi k}{T}\right)\right)^2\right)$$

(b) Del punto anterior, determine la función de densidad condicional $f_X(x_1,\ldots,x_n|A)$ (también conocida como verosimilitud) y con ello el estimador de máxima verosimilitud de A dadas las observaciones $x_1,\ldots,x_n$, es decir, la solución de:

$$\hat{A}_{ML}(x_1,\ldots,x_n) = \arg\max_{A} \ln(f_X(x_1,\ldots,x_n|A)) \tag{10}$$

Donde $A \in \mathbb{R}^+$.

Indicación: Use el criterio de la primera derivada.

Hint: se debe llegar a una expresión cerrada función de $x_1,\ldots,x_n$ y parámetros conocidos del problema.

Solución:

Tomamos el logaritmo de la función de verosimilitud:

$$\ln f_X(x_1,\ldots,x_n|A) = -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{k=1}^{n}\left(x_k - A\cos\left(\frac{2\pi k}{T}\right)\right)^2$$

Para maximizar, derivamos respecto a A e igualamos a cero:

$$\frac{\partial}{\partial A}\ln f_X = -\frac{1}{2\sigma^2}\sum_{k=1}^{n} 2\left(x_k - A\cos\left(\frac{2\pi k}{T}\right)\right)\left(-\cos\left(\frac{2\pi k}{T}\right)\right) = 0$$

Simplificando:

$$\frac{1}{\sigma^2}\sum_{k=1}^{n}\left(x_k - A\cos\left(\frac{2\pi k}{T}\right)\right)\cos\left(\frac{2\pi k}{T}\right) = 0$$

Expandiendo:

$$\sum_{k=1}^{n} x_k\cos\left(\frac{2\pi k}{T}\right) - A\sum_{k=1}^{n}\cos^2\left(\frac{2\pi k}{T}\right) = 0$$

Despejando A:

$$\hat{A}_{ML} = \frac{\sum_{k=1}^{n} x_k \cos\left(\frac{2\pi k}{T}\right)}{\sum_{k=1}^{n} \cos^2\left(\frac{2\pi k}{T}\right)}$$

Para verificar que es un máximo, calculamos la segunda derivada:

$$\frac{\partial^2}{\partial A^2} \ln f_X = -\frac{1}{\sigma^2} \sum_{k=1}^{n} \cos^2\left(\frac{2\pi k}{T}\right) < 0$$

Como la segunda derivada es negativa, confirmamos que es un máximo.

Por lo tanto, el estimador de máxima verosimilitud es:

$$\boxed{\hat{A}_{ML}(x_1, \ldots, x_n) = \frac{\sum_{k=1}^{n} x_k \cos\left(\frac{2\pi k}{T}\right)}{\sum_{k=1}^{n} \cos^2\left(\frac{2\pi k}{T}\right)}}$$

(c) Verifique que $\hat{A}_{ML}(x_1, \ldots, x_n)$ es insesgado, determine su varianza y con ello muestre que es consistente.

Solución:

Para ver si es insesgado, calculamos la esperanza del estimador.

$$\mathbb{E}[\hat{A}_{ML}] = \mathbb{E}\left[\frac{\sum_{k=1}^{n} X_k \cos\left(\frac{2\pi k}{T}\right)}{\sum_{k=1}^{n} \cos^2\left(\frac{2\pi k}{T}\right)}\right] = \frac{\sum_{k=1}^{n} \mathbb{E}[X_k] \cos\left(\frac{2\pi k}{T}\right)}{\sum_{k=1}^{n} \cos^2\left(\frac{2\pi k}{T}\right)}$$

Como $\mathbb{E}[X_k] = A\cos\left(\frac{2\pi k}{T}\right)$:

$$\mathbb{E}[\hat{A}_{ML}] = \frac{\sum_{k=1}^{n} A\cos\left(\frac{2\pi k}{T}\right)\cos\left(\frac{2\pi k}{T}\right)}{\sum_{k=1}^{n} \cos^2\left(\frac{2\pi k}{T}\right)} = \frac{A\sum_{k=1}^{n} \cos^2\left(\frac{2\pi k}{T}\right)}{\sum_{k=1}^{n} \cos^2\left(\frac{2\pi k}{T}\right)} = A$$

Por lo tanto, $\hat{A}_{ML}$ es **insesgado**.

Realizamos el calculo de su Varianza. Denotemos $c_k = \cos\left(\frac{2\pi k}{T}\right)$ y $C = \sum_{k=1}^{n} c_k^2$. Entonces:

$$\hat{A}_{ML} = \frac{1}{C}\sum_{k=1}^{n} X_k c_k$$

La varianza es:

$$\text{Var}(\hat{A}_{ML}) = \text{Var}\left(\frac{1}{C}\sum_{k=1}^{n} X_k c_k\right) = \frac{1}{C^2}\sum_{k=1}^{n} c_k^2 \,\text{Var}(X_k) = \frac{1}{C^2}\sum_{k=1}^{n} c_k^2\sigma^2 = \frac{\sigma^2 C}{C^2} = \frac{\sigma^2}{C}$$

Por lo tanto:

$$\boxed{\text{Var}(\hat{A}_{ML}) = \frac{\sigma^2}{\sum_{k=1}^{n} \cos^2\left(\frac{2\pi k}{T}\right)}}$$

Para mostrar consistencia, observamos que:

$$\sum_{k=1}^{n} \cos^2\left(\frac{2\pi k}{T}\right) \approx \frac{n}{2} \quad \text{cuando } n \to \infty$$

Por lo tanto:

$$\lim_{n\to\infty} \text{Var}(\hat{A}_{ML}) = \lim_{n\to\infty} \frac{\sigma^2}{\sum_{k=1}^{n} \cos^2\left(\frac{2\pi k}{T}\right)} = \lim_{n\to\infty} \frac{2\sigma^2}{n} = 0$$

Como el sesgo es cero y la varianza tiende a cero cuando $n \to \infty$, el estimador $\hat{A}_{ML}$ es **consistente**.

(d) Demuestre que $\hat{A}_{ML}(x_1, \ldots, x_n)$ es estimador insesgado de A de mínima varianza.
Hint: Utilizar la cota de Cramer-Rao y concluya su análisis.

Solución:
Para demostrar que es de mínima varianza, calcularemos la información de Fisher y verificaremos si el estimador alcanza la cota de Cramér-Rao.
La log-verosimilitud es:

$$\ln L(A) = -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{k=1}^{n}\left(x_k - A\cos\left(\frac{2\pi k}{T}\right)\right)^2$$

La primera derivada:

$$\frac{\partial \ln L}{\partial A} = \frac{1}{\sigma^2}\sum_{k=1}^{n}\left(x_k - A\cos\left(\frac{2\pi k}{T}\right)\right)\cos\left(\frac{2\pi k}{T}\right)$$

La segunda derivada:

$$\frac{\partial^2 \ln L}{\partial A^2} = -\frac{1}{\sigma^2}\sum_{k=1}^{n}\cos^2\left(\frac{2\pi k}{T}\right)$$

La información de Fisher es:

$$\mathcal{I}_n(A) = -\mathbb{E}\left[\frac{\partial^2 \ln L}{\partial A^2}\right] = \frac{1}{\sigma^2}\sum_{k=1}^{n}\cos^2\left(\frac{2\pi k}{T}\right)$$

La cota de Cramér-Rao para A es:

$$\text{Var}(\hat{A}) \geq \frac{1}{\mathcal{I}_n(A)} = \frac{\sigma^2}{\sum_{k=1}^{n}\cos^2\left(\frac{2\pi k}{T}\right)}$$

Como calculamos en la parte anterior:

$$\mathrm{Var}(\hat{A}_{ML}) = \frac{\sigma^2}{\sum_{k=1}^{n} \cos^2\left(\frac{2\pi k}{T}\right)}$$

Por lo tanto:

$$\mathrm{Var}(\hat{A}_{ML}) = \frac{1}{\mathcal{I}_n(A)}$$

Esto demuestra que $\hat{A}_{ML}$ **alcanza la cota de Cramér-Rao**. El estimador de máxima verosimilitud $\hat{A}_{ML}$ para el sistema de modulación AM es óptimo en el sentido de que minimiza la varianza entre todos los estimadores insesgados de A.