



Ingeniería Eléctrica

FACULTAD DE CIENCIAS
FÍSICAS Y MATEMÁTICAS
UNIVERSIDAD DE CHILE

Análisis de Sistemas Dinámicos y Estimación

EL3204-1

Pauta Auxiliar 13 - Repaso Examen

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 $\hat{\theta}(x_1, \dots, x_n)$ 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 $\hat{\theta}(x_1, \dots, x_n)$ es insesgado si:

$$\mathbb{E}[\hat{\theta}(X_1, \dots, X_n)] = \theta$$

Es decir, en promedio el estimador entrega el valor verdadero del parámetro.

2. **Estimador Asintóticamente Insesgado:** Si el sesgo tiende a cero cuando $n \rightarrow \infty$:

$$\lim_{n \rightarrow \infty} \mathbb{E}[\hat{\theta}(X_1, \dots, X_n)] = \theta$$

3. **Estimador Consistente:** Un estimador es consistente si converge en probabilidad al valor verdadero:

$$\lim_{n \rightarrow \infty} P_X(|\hat{\theta}(X_1, \dots, X_n) - \theta| > \epsilon) = 0 \quad \forall \epsilon > 0$$

4. **Observación:** Mediante la desigualdad de Chebyshev, si un estimador es insesgado y su varianza tiende a cero cuando $n \rightarrow \infty$, entonces es consistente:

$$\lim_{n \rightarrow \infty} \text{Var}(\hat{\theta}(X_1, \dots, X_n)) = 0 \implies \text{consistencia}$$

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:

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

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

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

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:

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

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:

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

Método de solución: Criterio de la primera derivada:

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

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

7. Estimación Bayesiana

En el enfoque bayesiano, el parámetro θ es considerado como una variable aleatoria con una distribución a priori $P_{\Theta}(\theta)$ o $f_{\Theta}(\theta)$. Se busca encontrar estimadores basados en la distribución a posteriori:

$$f_{\Theta|X}(\theta|x) = \frac{f_{X|\Theta}(x|\theta)f_{\Theta}(\theta)}{f_X(x)}$$

- **Estimador MMSE (Minimum Mean Square Error):**

El estimador que minimiza el error cuadrático medio es la esperanza condicional:

$$\hat{\theta}_{MMSE}(x) = \mathbb{E}[\Theta|X=x] = \int_{-\infty}^{\infty} \theta \cdot f_{\Theta|X}(\theta|x) d\theta$$

Para el caso discreto:

$$\hat{\theta}_{MMSE}(x) = \sum_{\theta} \theta \cdot P_{\Theta|X}(\theta|x)$$

- **Estimador MAP (Maximum A Posteriori):**

El estimador MAP maximiza la distribución a posteriori:

$$\hat{\theta}_{MAP}(x) = \arg \max_{\theta} f_{\Theta|X}(\theta|x) = \arg \max_{\theta} f_{X|\Theta}(x|\theta)f_{\Theta}(\theta)$$

Para el caso discreto:

$$\hat{\theta}_{MAP}(x) = \arg \max_{\theta} P_{\Theta|X}(\theta|x) = \arg \max_{\theta} P_{X|\Theta}(x|\theta)P_{\Theta}(\theta)$$

En la práctica se trabaja con el logaritmo:

$$\hat{\theta}_{MAP}(x) = \arg \max_{\theta} \ln f_{X|\Theta}(x|\theta) + \ln f_{\Theta}(\theta)$$

1. Considere el problema de estimar los parámetros $\theta > 0$ y $k > 0$ de la distribución Gamma. Para esto, considere que, si $X \sim \text{Gamma}(k, \theta)$ entonces

$$f_X(x) = \frac{x^{k-1} e^{-x/\theta}}{\Gamma(k) \theta^k} \mathbf{1}_{(0, \infty)}(x).$$

Suponga que usted tiene n realizaciones X_1, \dots, X_n i.i.d., donde $X_i \sim \text{Gamma}(k, \theta)$ para todo $i \in \{1, \dots, n\}$.

- (a) Asumiendo que k es conocido, encuentre un estimador de máxima verosimilitud para estimar θ a partir de X_1, \dots, X_n .

Solución:

Asumiendo que $k > 0$ es conocido, notemos que el estimador de máxima verosimilitud estará dado por

$$\hat{\theta}(X_1^n) = \arg \max_{\theta > 0} L(X_1^n | \theta) = \arg \max_{\theta > 0} \ln L(X_1^n | \theta).$$

Recordemos que es posible trabajar con la log-verosimilitud, ya que el logaritmo es una función monótona y estrictamente creciente, por lo que maximizar la verosimilitud es equivalente a maximizar la log-verosimilitud.

Dado que las realizaciones son i.i.d., la función de verosimilitud es el producto de las densidades individuales:

$$L(X_1^n | \theta) = \prod_{i=1}^n f_X(X_i | \theta) = \prod_{i=1}^n \frac{X_i^{k-1} e^{-\frac{X_i}{\theta}}}{\Gamma(k) \theta^k}.$$

Aplicando el logaritmo natural a ambos lados, obtenemos:

$$\begin{aligned} \ln L(X_1^n | \theta) &= \ln \left(\prod_{i=1}^n \frac{X_i^{k-1} e^{-\frac{X_i}{\theta}}}{\Gamma(k) \theta^k} \right) \\ &= \sum_{i=1}^n \ln \left(\frac{X_i^{k-1} e^{-\frac{X_i}{\theta}}}{\Gamma(k) \theta^k} \right) \\ &= \sum_{i=1}^n \left[\ln(X_i^{k-1}) + \ln \left(e^{-\frac{X_i}{\theta}} \right) - \ln(\Gamma(k)) - \ln(\theta^k) \right] \\ &= \sum_{i=1}^n \left[(k-1) \ln X_i - \frac{X_i}{\theta} - \ln \Gamma(k) - k \ln \theta \right]. \end{aligned}$$

Así, la log-verosimilitud queda expresada como:

$$\ln L(X_1^n | \theta) = \sum_{i=1}^n \left[(k-1) \ln X_i - \frac{X_i}{\theta} - \ln \Gamma(k) - k \ln \theta \right].$$

Luego, si queremos maximizarla, podemos utilizar la condición de primer orden, es decir, derivar e igualar a 0. Para calcular la derivada con respecto a θ , aplicamos el operador derivada término

a término:

$$\begin{aligned}\frac{\partial}{\partial \theta} (\ln L(X_1^n | \theta)) &= \frac{\partial}{\partial \theta} \left(\sum_{i=1}^n \left[(k-1) \ln X_i - \frac{X_i}{\theta} - \ln \Gamma(k) - k \ln \theta \right] \right) \\ &= \sum_{i=1}^n \frac{\partial}{\partial \theta} \left[(k-1) \ln X_i - \frac{X_i}{\theta} - \ln \Gamma(k) - k \ln \theta \right] \\ &= \sum_{i=1}^n \left[\frac{\partial}{\partial \theta} (k-1) \ln X_i - \frac{\partial}{\partial \theta} \left(\frac{X_i}{\theta} \right) - \frac{\partial}{\partial \theta} \ln \Gamma(k) - \frac{\partial}{\partial \theta} (k \ln \theta) \right].\end{aligned}$$

Notemos que $(k-1) \ln X_i$ y $\ln \Gamma(k)$ no dependen de θ , por lo que sus derivadas son cero. Para el término $\frac{X_i}{\theta}$, recordemos que $\frac{\partial}{\partial \theta} \left(\frac{X_i}{\theta} \right) = \frac{\partial}{\partial \theta} (X_i \theta^{-1}) = -X_i \theta^{-2} = -\frac{X_i}{\theta^2}$. Para el término $k \ln \theta$, tenemos $\frac{\partial}{\partial \theta} (k \ln \theta) = \frac{k}{\theta}$. Entonces:

$$\begin{aligned}\frac{\partial}{\partial \theta} (\ln L(X_1^n | \theta)) &= \sum_{i=1}^n \left[0 - \left(-\frac{X_i}{\theta^2} \right) - 0 - \frac{k}{\theta} \right] \\ &= \sum_{i=1}^n \left[\frac{X_i}{\theta^2} - \frac{k}{\theta} \right] \\ &= \frac{1}{\theta^2} \sum_{i=1}^n X_i - \sum_{i=1}^n \frac{k}{\theta} \\ &= \frac{1}{\theta^2} \sum_{i=1}^n X_i - \frac{nk}{\theta}.\end{aligned}$$

Igualando a 0 para encontrar el máximo, y despejando θ , tenemos

$$\frac{1}{\theta^2} \sum_{i=1}^n X_i - \frac{nk}{\theta} = 0 \quad \Leftrightarrow \quad \theta = \frac{1}{nk} \sum_{i=1}^n X_i.$$

Así, podemos ver que la solución al problema de optimización y, por lo tanto, la expresión para el estimador de máxima verosimilitud, corresponde a

$$\hat{\theta}(X_1^n) = \frac{1}{nk} \sum_{i=1}^n X_i.$$

(b) Verifique si el estimador es insesgado, y obtenga su varianza.

Solución:

Para verificar si es insesgado, queremos ver si $\mathbb{E}_{X_1^n} \{\hat{\theta}(X_1^n)\} = \theta$. Tomando la esperanza, comenzaremos notando que, al tener una suma, podemos distribuir el operador de esperanza, lo cual nos simplificará el desarrollo.

Recordemos que si $X \sim \text{Gamma}(k, \theta)$ entonces $\mathbb{E}\{X\} = k\theta$ y $\text{Var}(X) = k\theta^2$, tenemos

$$\begin{aligned}\mathbb{E}_{X_1^n} \left\{ \hat{\theta}(X_1^n) \right\} &= \mathbb{E}_{X_1^n} \left\{ \frac{1}{nk} \sum_{i=1}^n X_i \right\} \\ &= \frac{1}{nk} \sum_{i=1}^n \mathbb{E}_{X_i} \{X_i\} \\ &= \frac{1}{nk} \sum_{i=1}^n k\theta \\ &= \theta,\end{aligned}$$

de donde podemos ver que el estimador es insesgado. Por otra parte, para la varianza notemos que solamente podemos distribuir la varianza en la suma si las variables son independientes. En este caso, dado que X_1, \dots, X_n son i.i.d., podemos distribuirla, de modo que tenemos

$$\begin{aligned}\text{Var} \left(\hat{\theta}(X_1^n) \right) &= \text{Var} \left(\frac{1}{nk} \sum_{i=1}^n X_i \right) \\ &= \frac{1}{n^2 k^2} \sum_{i=1}^n \text{Var}(X_i) \\ &= \frac{1}{n^2 k^2} \sum_{i=1}^n k\theta^2 \\ &= \frac{\theta^2}{nk}.\end{aligned}$$

(c) Determine si el estimador que obtuvo es consistente.

Solución:

Recordemos que un estimador es consistente si y solo si converge en probabilidad al valor verdadero del parámetro, es decir, $\hat{\theta}(X_1^n) \xrightarrow{P} \theta$. Demostrar esta convergencia directamente puede ser complejo, por lo que utilizaremos un resultado alternativo: **un estimador es consistente si es asintóticamente insesgado y su varianza tiende a cero cuando el tamaño de la muestra tiende a infinito.**

Aplicando este criterio a nuestro caso:

1. **Insesgamiento asintótico:** Del inciso anterior, demostramos que el estimador es insesgado, es decir, $\mathbb{E}[\hat{\theta}(X_1^n)] = \theta$ para todo n . Por lo tanto, es también asintóticamente insesgado.
2. **Varianza tiende a cero:** La varianza del estimador está dada por

$$\text{Var} \left(\hat{\theta}(X_1^n) \right) = \frac{\theta^2}{nk}.$$

Calculando el límite cuando $n \rightarrow \infty$:

$$\lim_{n \rightarrow \infty} \text{Var} \left(\hat{\theta}(X_1^n) \right) = \lim_{n \rightarrow \infty} \frac{\theta^2}{nk} = \theta^2 \cdot \lim_{n \rightarrow \infty} \frac{1}{nk} = 0.$$

Dado que ambas condiciones se cumplen, concluimos que el estimador $\hat{\theta}(X_1^n) = \frac{1}{nk} \sum_{i=1}^n X_i$ es **consistente**.

- (d) Obtenga la información de Fisher. ¿Qué le permite decir sobre el rendimiento del estimador que obtuvo?

Solución:

Recordemos que la información de Fisher corresponde a la cantidad de información que una variable aleatoria contiene sobre un parámetro desconocido. Se define como

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

Para una muestra i.i.d. de tamaño n , la información de Fisher total es $\mathcal{I}_n(\theta) = n \cdot \mathcal{I}_1(\theta)$, donde $\mathcal{I}_1(\theta)$ es la información de Fisher de una sola observación. Partimos de la log-densidad de una observación:

$$\ln f_X(x|\theta) = (k-1) \ln x - \frac{x}{\theta} - \ln \Gamma(k) - k \ln \theta.$$

Calculamos la primera derivada con respecto a θ . Notemos que los términos $(k-1) \ln x$ y $\ln \Gamma(k)$ no dependen de θ , por lo que sus derivadas son cero:

$$\begin{aligned} \frac{\partial}{\partial \theta} (\ln f_X(x|\theta)) &= \frac{\partial}{\partial \theta} \left(-\frac{x}{\theta} \right) + \frac{\partial}{\partial \theta} (-k \ln \theta) \\ &= -\frac{\partial}{\partial \theta} (x\theta^{-1}) - k \frac{1}{\theta} \\ &= -x(-\theta^{-2}) - \frac{k}{\theta} \\ &= \frac{x}{\theta^2} - \frac{k}{\theta}. \end{aligned}$$

Ahora calculamos la segunda derivada:

$$\begin{aligned} \frac{\partial^2}{\partial \theta^2} (\ln f_X(x|\theta)) &= \frac{\partial}{\partial \theta} \left(\frac{x}{\theta^2} - \frac{k}{\theta} \right) \\ &= \frac{\partial}{\partial \theta} (x\theta^{-2}) - \frac{\partial}{\partial \theta} (k\theta^{-1}) \\ &= x(-2\theta^{-3}) - k(-\theta^{-2}) \\ &= -\frac{2x}{\theta^3} + \frac{k}{\theta^2}. \end{aligned}$$

Usando la segunda definición de información de Fisher, tenemos:

$$\begin{aligned}\mathcal{I}_1(\theta) &= -\mathbb{E} \left[\frac{\partial^2 \ln f_X(X|\theta)}{\partial \theta^2} \right] \\ &= -\mathbb{E} \left[-\frac{2X}{\theta^3} + \frac{k}{\theta^2} \right] \\ &= \mathbb{E} \left[\frac{2X}{\theta^3} \right] - \mathbb{E} \left[\frac{k}{\theta^2} \right] \\ &= \frac{2}{\theta^3} \mathbb{E}[X] - \frac{k}{\theta^2}.\end{aligned}$$

Recordando que para $X \sim \text{Gamma}(k, \theta)$, tenemos $\mathbb{E}[X] = k\theta$, sustituimos:

$$\mathcal{I}_1(\theta) = \frac{2}{\theta^3}(k\theta) - \frac{k}{\theta^2} = \frac{2k}{\theta^2} - \frac{k}{\theta^2} = \frac{k}{\theta^2}.$$

Usando la propiedad aditiva de la información de Fisher para muestras i.i.d.:

$$\mathcal{I}_n(\theta) = n \cdot \mathcal{I}_1(\theta) = n \cdot \frac{k}{\theta^2} = \frac{nk}{\theta^2}.$$

La cota de Cramér-Rao establece que para cualquier estimador insesgado $\hat{\theta}$ de θ , su varianza está acotada inferiormente por el inverso de la información de Fisher:

$$\text{Var}(\hat{\theta}) \geq \frac{1}{\mathcal{I}_n(\theta)} = \frac{1}{\frac{nk}{\theta^2}} = \frac{\theta^2}{nk}.$$

Del inciso anterior, obtuvimos que la varianza de nuestro estimador $\hat{\theta}(X_1^n) = \frac{1}{nk} \sum_{i=1}^n X_i$ es:

$$\text{Var}(\hat{\theta}(X_1^n)) = \frac{\theta^2}{nk}.$$

Observemos que:

$$\text{Var}(\hat{\theta}(X_1^n)) = \frac{\theta^2}{nk} = \frac{1}{\mathcal{I}_n(\theta)}.$$

Dado que la varianza del estimador es exactamente igual a la cota inferior de Cramér-Rao, el estimador **alcanza la cota de Cramér-Rao**. Esto significa que entre todos los estimadores insesgados posibles para θ , este tiene la menor varianza. No existe ningún otro estimador insesgado que pueda tener menor varianza.

- (e) Ahora considere que θ es conocido. ¿Es posible expresar, de forma cerrada, el estimador de máxima verosimilitud para estimar k a partir de X_1, \dots, X_n ?

Solución:

Cuando θ es conocido y queremos estimar k , la log-verosimilitud es:

$$\log L(X_1^n | k) = \sum_{i=1}^n \left[(k-1) \log X_i - \frac{X_i}{\theta} - \log \Gamma(k) - k \log \theta \right].$$

Derivando con respecto a k :

$$\frac{\partial \log L(X_1^n | k)}{\partial k} = \sum_{i=1}^n [\log X_i - \psi(k) - \log \theta],$$

donde $\psi(k) = \frac{\Gamma'(k)}{\Gamma(k)}$ es la función digamma.

Igualando a cero:

$$\sum_{i=1}^n \log X_i - n\psi(k) - n \log \theta = 0,$$

lo que implica:

$$\psi(k) = \frac{1}{n} \sum_{i=1}^n \log X_i - \log \theta = \log \left(\frac{\bar{X}_g}{\theta} \right),$$

donde $\bar{X}_g = \exp \left(\frac{1}{n} \sum_{i=1}^n \log X_i \right)$ es la media geométrica.

No es posible expresar \hat{k} de forma cerrada, ya que la ecuación $\psi(k) = c$ (donde c es una constante conocida) no tiene solución analítica explícita. La función digamma no tiene inversa en forma cerrada.