Ingeniería Eléctrica
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS UNIVERSIDAD DE CHILE

Análisis de Sistemas Dinámicos y Estimación

EL3204-1

Auxiliar 12 - Estimación Bayesiana

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:

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

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. 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 $\theta \in \mathbb{R}$ dada una observación $X \in \mathbb{R}$, se sabe que la distribución condicional de Θ dado X está dotada de la siguiente función de densidad condicional definida como:

$$f_{\Theta|X}(\theta|x) = \begin{cases} e^{-(\theta - x)}, & \text{si } \theta > x \\ 0, & \text{si } x > \theta \end{cases}$$

   Encuentre el estimador de mínimo error cuadrático medio y MAP.

2. Sean $X_1, \ldots, X_n$ muestras i.i.d. de la variable aleatoria $X \sim \mathcal{N}(\mu, \sigma^2)$. A su vez, μ es una variable aleatoria continua tal que $\mu \sim \mathcal{N}(0, \sigma_0^2)$.

   1. Determine las densidades $f_{X_1,\ldots,X_n}(x_1, \ldots, x_n|\mu)$ y $f_M(\mu)$.
   2. Obtenga el estimador MAP de μ, $\hat{\mu}_{MAP}$.
   3. Determine los casos límite en que $\sigma_0^2 \longrightarrow \infty$ y $\sigma_0^2 \longrightarrow 0$. Interprete cada situación.

3. Considere que tiene un cuerpo radiactivo que, dentro de un cierto intervalo de tiempo, emite un cierto $\Theta \in \mathbb{N}$ partículas el cual usted quiere medir con un instrumento el cual le indica el total de partículas detectadas durante dicho intervalo. Sin embargo, su detector es imperfecto, por lo que hay una probabilidad p de que detecte correctamente cada partícula.

   Gracias a estudios anteriores, usted sabe que $\Theta \sim \text{Poisson}(\lambda)$, donde λ es la propiedad asociada a la tasa de emisión del cuerpo.

   1. Plantee el problema de estimar el número de partículas emitidas a partir del número de partículas detectadas, e indique las distribuciones asociadas. ¿Es un problema de estimación o detección?
   2. Encuentre el estimador MMSE.
   3. Encuentre el estimador MAP, relajando el espacio de decisión a los números reales. *Indicación:* Considere que $\Gamma(z+1) = z!$ es la continuación analítica de los factoriales, y que $\psi(z) = \frac{d \log \Gamma(z)}{dz}$ es la función digamma.