

Estimación y Predicción en Series Temporales

Práctico 3: Estimadores MLE y MAP

Departamento de Procesamiento de Señales

Instituto de Ingeniería Eléctrica
Facultad de Ingeniería

2022

1

Hallar el MLE del parámetro desconocido [7.3]
Consistencia del MLE [7.5]
Propiedades asintóticas del MLE [7.20]

2

Minimum Mean Square Error (MMSE)
Comparación MMSE vs MAP [11.3]
MMSE de una transformación lineal [11.8, 11.9]
Máximo a posteriori (MAP)
Hallar el estimador MAP [11.4]

Hallar el MLE del parámetro desconocido [7.3]

- Se tienen N muestras IID de las siguientes PDFs
 - a) Gaussiana

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

- b) Exponencial

$$p(x; \lambda) = \begin{cases} \lambda \exp(-\lambda x) & x > 0 \\ 0 & x < 0 \end{cases}$$

- Hallar el MLE y verificar que maximiza la función de verosimilitud.

Hallar el MLE del parámetro desconocido [7.3]

- a) Gaussiana

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

- Como son variables IID, la verosimilitud de las N muestras es:

$$p(\mathbf{x}; \mu) = \prod_{i=0}^{N-1} \frac{1}{\sqrt{2\pi}} \exp \left[-\frac{1}{2}(x_i - \mu)^2 \right]$$

- Como el logaritmo es una función monótona, maximizar $p(\mathbf{x}; \mu)$, es lo mismo que maximizar $\log p(\mathbf{x}; \mu)$

$$\log p(\mathbf{x}; \mu) = N \log \left(\frac{1}{\sqrt{2\pi}} \right) + \sum_{i=0}^{N-1} -\frac{1}{2}(x_i - \mu)^2$$

Hallar el MLE del parámetro desconocido [7.3]

$$\log p(\mathbf{x}; \mu) = -\frac{N}{2} \log(2\pi) - \frac{1}{2} \sum_{i=0}^{N-1} (x_i - \mu)^2$$

- Queremos maximizar según μ . La $\log p(\mathbf{x}; \mu)$ es un polinomio de grado 2 con concavidad negativa, tiene un único máximo global.

$$\frac{\partial \log p(\mathbf{x}; \mu)}{\partial \mu} = 0$$

$$\frac{\partial \log p(\mathbf{x}; \mu)}{\partial \mu} = \sum_{i=0}^{N-1} x_i - \mu = 0 \iff \mu = \frac{\sum_i x_i}{N}$$

- El MLE $\hat{\mu}$ es la media muestral.

Hallar el MLE del parámetro desconocido [7.3]

- b) Exponencial

$$p(x; \lambda) = \begin{cases} \lambda \exp(-\lambda x) & x > 0 \\ 0 & x < 0 \end{cases}$$

Planteamos la verosimilitud

$$p(\mathbf{x}; \lambda) = \prod_{i=0}^{N-1} \exp(-\lambda x_i)$$

Tomando logaritmo

$$\log p(\mathbf{x}; \lambda) = N \log(\lambda) - \sum_{i=0}^{N-1} \lambda x_i$$

Hallar el MLE del parámetro desconocido [7.3]

- Hallamos el máximo anulando la primera derivada

$$\frac{\partial \log p(\mathbf{x}; \lambda)}{\partial \lambda} = \frac{N}{\lambda} - \sum_{i=0}^{N-1} x_i$$

$$\frac{N}{\lambda} - \sum_{i=0}^{N-1} x_i = 0 \iff \lambda = \frac{N}{\sum_i x_i}$$

- El MLE $\hat{\lambda}$ es el inverso de la media muestral.

Hallar el MLE del parámetro desconocido [7.3]

- Solo falta comprobar que el valor hallado corresponde al máximo. Calculemos la derivada segunda y verifiquemos su valor negativo.

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

La segunda derivada es negativa $\forall \lambda$.

Consistencia del MLE [7.5]

- Un estimador $\hat{\theta}$ es consistente si, para cualquier $\epsilon > 0$,

$$\lim_{N \rightarrow \infty} \Pr\{|\hat{\theta} - \theta| > \epsilon\} = 0$$

- Demostrar que la media muestral es un estimador consistente para la estimación del nivel A de DC en AWGN. Sugerencia: Usar la desigualdad de Chebychev.

Teorema

Sea X una V.A con esperanza finita μ y varianza finita $\sigma^2 \neq 0$. Entonces para cualquier real $k > 0$

$$\Pr\{|X - \mu| \geq k\sigma\} \leq \frac{1}{k^2}$$

Consistencia del MLE [7.5]

Se tiene

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

- La esperanza del estimador $\hat{A} = \frac{\sum_i x_i}{N}$, es $\mathbb{E}[\hat{A}] = A$
- La varianza es

$$\text{var}(\hat{A}) = \text{var}\left(\frac{\sum_i x_i}{N}\right) = \frac{\sum_i \text{var}(x_i)}{N^2} = \frac{\sigma^2}{N}$$

- Usando $k = \frac{\epsilon}{\sigma}$ en la desigualdad

$$\Pr\{|\hat{A} - A| \geq \epsilon\} \stackrel{\text{Chebychev}}{\leq} \frac{\sigma^2}{N\epsilon^2} \xrightarrow{N \rightarrow \infty} 0$$

Propiedades asintóticas del MLE [7.20]

Se consideran N muestras de la señal

$$x[n] = s[n] + w[n]$$

donde $s[n]$ es desconocido y $w[n]$ es WGN de varianza σ^2 .
Hallar el MLE de $s[n]$ y analizar las propiedades asintóticas:
sesgo, eficiencia, distribución normal y consistencia.

- Las muestras son IID

$$p(\mathbf{x}; \mathbf{s}) = p_w(\mathbf{x} - \mathbf{s})$$

$$p(\mathbf{x}; \mathbf{s}) = \prod_{i=0}^{N-1} \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{1}{2} \left(\frac{x_i - s_i}{\sigma} \right)^2 \right]$$

Propiedades asintóticas del MLE [7.20]

- Aplicamos logaritmo

$$\log p(\mathbf{x}; \mathbf{s}) = -\frac{N}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=0}^{N-1} (x_i - s_i)^2$$

- Tomamos derivadas parciales respecto a cada s_i

$$\frac{\partial \log p(\mathbf{x}; \mathbf{s})}{\partial s_i} = \frac{x_i - s_i}{\sigma^2} = 0 \iff s_i = x_i$$

- El estimador MLE de cada $\hat{s}[n]$ es simplemente el valor de la muestra $x[n]$

Propiedades asintóticas del MLE [7.20]

- Sesgo:

El estimador es insesgado ($\mathbb{E}[\hat{s}[n]] = \mathbb{E}[x[n]] = \mathbb{E}[s[n]]$)

- Distribución normal (PDF):

$$(x[n] - s[n]) \sim \mathcal{N}(0, \sigma^2)$$

$$x[n] \sim \mathcal{N}(s[n], \sigma^2)$$

El estimador tiene distribución normal.

Propiedades asintóticas del MLE [7.20]

- Eficiencia:

Calculemos la CRLB para ver si la varianza del estimador la alcanza, de ser así, tenemos un estimador eficiente.

$$\text{var}(\hat{s}[n]) \geq \frac{1}{-\mathbb{E}_x \left[\frac{\partial^2 \log p(x[n]; s[n])}{\partial s[n]^2} \right]}$$

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

La varianza del estimador alcanza la CRLB, por lo tanto, es eficiente.

Propiedades asintóticas del MLE [7.20]

- Consistencia:
El estimador no es consistente pues cuando $N \rightarrow \infty$ la varianza del estimador se mantiene constante. (No hay ningún efecto de promediado, ya que se quieren estimar tantos parámetros como muestras se tienen)
- Para que la distribución asintótica del MLE sea gaussiana, se deben tener N muestras para estimar cada parámetro, de forma que cuando $N \rightarrow \infty$ se pueda aplicar el Teorema Central del Límite. (Ver Apéndice 2 [?])

Comparación MMSE vs MAP [11.3]

- Para la siguiente PDF a posteriori:

$$p(\theta|x) = \begin{cases} \exp[-(\theta - x)] & \theta > x \\ 0 & \theta < x \end{cases}$$

Hallar los estimadores MMSE y MAP.

Comparación MMSE vs MAP [11.3]

- Estimador MAP

Es el θ que maximiza la posteriori $\rightarrow \hat{\theta}_{MAP} = x$

- Estimador MMSE

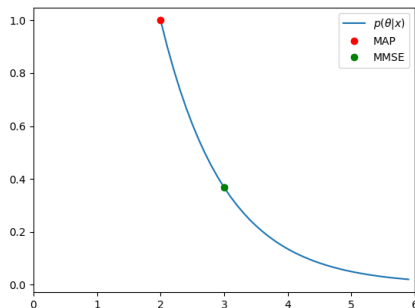
$$\mathbf{E}[\theta|x] = \int_x^\infty \theta \exp[-(\theta - x)] d\theta$$

$$\int_x^\infty \theta \exp[-(\theta - x)] d\theta = -e^x (\theta e^{-\theta} + e^{-\theta}) \Big|_x^\infty$$

$$= e^x (x e^{-x} + e^{-x}) = x + 1$$

Comparación MMSE vs MAP [11.3]

- $\hat{\theta}_{MAP} = x$
- $\hat{\theta}_{MMSE} = x + 1$



MMSE de una transformación lineal [11.8, 11.9]

- Considerar un parámetro θ que evoluciona en el tiempo con la siguiente ley determinística:

$$\theta[n] = A\theta[n - a]$$

Siendo A una matriz $p \times p$ conocida e invertible y $\theta[0]$ un vector aleatorio (realización del parámetro desconocido).

- Demostrar que el MMSE de $\theta[n]$ es:

$$\hat{\theta}[n] = A^n \hat{\theta}[0]$$

Siendo $\hat{\theta}[0]$ el MMSE de $\theta[0]$

MMSE de una transformación lineal [11.8, 11.9]

•

$$\hat{\boldsymbol{\theta}}[n] = \mathbf{E}[\boldsymbol{\theta}[n]|\mathbf{x}] = \mathbf{E}[A\boldsymbol{\theta}[n-1]|\mathbf{x}]$$

$$= A\mathbf{E}[\boldsymbol{\theta}[n-1]|\mathbf{x}] = A\hat{\boldsymbol{\theta}}[n-1]$$

...

$$\hat{\boldsymbol{\theta}}[n] = A^n \hat{\boldsymbol{\theta}}[0]$$

- Como el modelo es lineal y la esperanza también lo es, basta con estimar el parámetro en un tiempo para tener un estimado para todos los demás instantes

MMSE de una transformación lineal [11.8, 11.9]

- Un vehículo que arranca en una posición desconocida y se mueve con velocidad desconocida pero constante, puede modelarse de la siguiente forma:

$$x[n] = x[0] + v_x n$$

$$y[n] = y[0] + v_y n$$

Se quieren estimar todos los parámetros para todo tiempo y se modela la posición inicial y la velocidad como un vector aleatorio $\theta[n] = [x[n], y[n], v_x, v_y]^T$

MMSE de una transformación lineal [11.8, 11.9]

- Apliquemos lo hallado anteriormente. Primero escribamos la evolución de los parámetros en forma lineal:

$$\boldsymbol{\theta}[n] = A\boldsymbol{\theta}[n-1] = A^n\boldsymbol{\theta}[0]$$

Por ejemplo

$$x[n] - x[n-1] = vx$$

$$\rightarrow x[n] = x[n-1] + vx$$

Lo mismo sucede para $y[n]$ y las velocidades permanecen constantes.

MMSE de una transformación lineal [11.8, 11.9]

- Podemos escribirlo en forma matricial

$$A = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

- Aplicando el resultado anterior, podemos estimar la posición para cualquier tiempo con la matriz A y el estimador MMSE en tiempo 0

$$\hat{\theta}_{MMSE}[n] = A^n \hat{\theta}_{MMSE}[0]$$

Hallar el estimador MAP [11.4]

- Se observan N muestras de la siguiente señal:

$$x[n] = A + w[n]$$

A es desconocido, $\lambda > 0$ y $w[n]$ WGN de varianza σ^2 independiente de A . Hallar el estimador MAP de A , asumiendo que su distribución a priori es:

$$p(A) = \begin{cases} \lambda e^{-\lambda A} & A > 0 \\ 0 & A < 0 \end{cases}$$

Hallar el estimador MAP [11.4]

- El estimador MAP se halla de la siguiente forma:

$$\begin{aligned}\hat{A} &= \arg \max_A p(A|x) = \arg \max_A \frac{p(x|A)p(A)}{p(x)} \\ &= \arg \max_A \log p(\mathbf{x}|A) + \log p(A)\end{aligned}$$

- Utilizando las PDFs del problema:

$$\log p(\mathbf{x}|A) = -\frac{N}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=0}^{N-1} (x_i - A)^2$$

$$\log p(A) = \log(\lambda) - \lambda A$$

con $A > 0$

Hallar el estimador MAP [11.4]

- Tomamos derivada e igualamos a 0 para hallar el máximo

$$\frac{\partial \log p(\mathbf{x}|A) + \log p(A)}{\partial a} = \frac{\sum_{i=0}^{N-1} (x_i - A)^2}{\sigma^2} - \lambda A = 0$$

$$\iff \hat{A} = \frac{\sum_i x_i}{N} - \frac{\sigma^2 \lambda}{N}$$

- Si $N \rightarrow \infty$, \hat{A}_{MAP} tiende a la media muestral (ignoramos el prior). El prior también se debilita si las muestras tienen poco ruido.
- Si λ es grande (prior fuerte), el estimador se aleja de la media muestral. Lo mismo sucede si las muestras son ruidosas.
- Es un estimador sesgado por el prior.

