Teoria da Estimação

Obter o melhor estimador de um ou mais parâmetros desconhecidos a partir de um conjunto de observações de um processo estocástico relacionado aos parâmetros desconhecidos.

Conjunto de amostras: $\underline{x} = [x(0) \ldots x(N-1)]^T$

Parâmetros desconhecidos: $\underline{\theta} = [\theta(0) \ldots \theta(L-1)]^T$

Estimador:

$$\underline{\hat{\theta}} = \varphi\{x(0) \ldots x(N-1)\} = \varphi\{\underline{x}\}$$

Como $\underline{x}$ é um processo aleatório, o estimador é uma variável aleatória.

Desafio: determinar o conjunto de funções $\varphi\{\cdot\}$ que proporciona a melhor estimativa de $\underline{\theta}$.

Propriedades dos Estimadores:

1) Polarização (Bias)

Um estimador é dito não-polarizado se

$$E\{\underline{\hat{\theta}}\} = E\{\underline{\theta}\}$$

Se $\underline{\theta}$ é determinístico, $E\{\underline{\hat{\theta}}\} = \underline{\theta}$.

$\text{Bias}(\hat{\theta}) = E\{\underline{\theta}\} - E\{\underline{\hat{\theta}}\}$

2) Eficiência:

No caso escalar, um estimador $\hat{\theta}$ é mais eficiente que um estimador $\tilde{\theta}$, considerando que ambos não sejam polarizados, se $Var(\hat{\theta}) \leq Var(\tilde{\theta})$.

No caso de múltiplos parâmetros, a condição é dada por

$$C_{\underline{\hat{\theta}}}(\underline{\hat{\theta}}) \leq C_{\underline{\tilde{\theta}}}(\underline{\tilde{\theta}}) \quad (C_{\underline{\tilde{\theta}}}(\underline{\tilde{\theta}}) - C_{\underline{\hat{\theta}}}(\underline{\hat{\theta}}) \text{ é semidefinida positiva})$$

onde $C_{\underline{x}} = E\{(\underline{x} - E[\underline{x}])(\underline{x} - E\{\underline{x}\})^H\}$ é a matriz de covariância de $\underline{x}$.

No contexto de estimadores não polarizados, é possível estabelecer um limite inferior de desempenho, o chamado limite de Cramér-Rao (Cramér-Rao Bound CRB)

$$Var(\hat{\theta}) \geq F^{-1}(\theta) \quad (\text{escalar})$$

ou

$$C(\underline{\hat{\theta}}) \geq F^{-1}(\underline{\theta}) \quad (\text{vetorial})$$

F é a matriz de informação de Fisher.

Exemplo: Constante mais ruído aditivo gaussiano.

$$x(n) = \theta + \eta(n) \qquad \eta(n) \sim N(0, \sigma^2)$$

$$\underline{x} = [x(0) \dots x(N-1)]^T$$

Dois possíveis estimadores:

$$\tilde{\theta} = x(0)$$

$$\hat{\theta} = \frac{1}{N} \sum_{n=0}^{N-1} x(n)$$

Polarização:

$$E\{\tilde{\theta}\} = E\{x(0)\} = \theta$$

$$E\{\hat{\theta}\} = E\left\{\frac{1}{N}\sum_{n=0}^{N-1} x(n)\right\} = \frac{1}{N}\sum_{n=0}^{N-1} E\{x(n)\} = \frac{1}{N}(N\theta) = \theta$$

Eficiência:

$$Var(\tilde{\theta}) = Var(x(0)) = \sigma^2$$

$$Var(\hat{\theta}) = Var\left(\frac{1}{N}\sum_{n=0}^{N-1} x(n)\right) = \frac{1}{N^2}\sum_{n=0}^{N-1} Var(x(n)) = \frac{1}{N^2}(N\sigma^2) = \frac{\sigma^2}{N}$$

$\therefore$ O estimador de média amostral é mais eficiente se $N > 1$

Podemos apontar duas famílias principais de métodos de estimação:

<u>1º Caso</u>: $\underline{\theta}$ é um vetor de parâmetros constantes e desconhecidos.

- Não há informação estatística sobre $\underline{\theta}$
- Principal representante: estimador de máxima verossimilhança (Maximum Likelihood, ML)

$$\hat{\underline{\theta}}_{ML} = \arg\max_{\underline{\theta}} f_X(\underline{X}|\underline{\theta}) \rightarrow$$

Para um conjunto de observações $\underline{X}$, buscamos os parâmetros $\underline{\theta}$ que levam à maior probabilidade $f_X(X|\theta)$ com a qual os dados foram gerados.

ou

$$\hat{\underline{\theta}}_{ML} = \arg\max_{\underline{\theta}} \left[\ln\left(f_X(\underline{X}|\underline{\theta})\right)\right]$$

Solução: $$\frac{\partial}{\partial \underline{\theta}} \ln\left[f_X(\underline{X}|\underline{\theta})\right] = 0$$

Frequentemente, o problema tem natureza não linear e apresenta mais do que uma solução. Nesse caso, $\hat{\underline{\theta}}_{ML}$ é o máximo global.

O estimador ML é assintoticamente eficiente, isto é,

$$Var(\hat{\theta}) = CRB \quad \text{qdo} \quad N \rightarrow \infty$$

<u>Exemplo</u>: Constante mais ruído WGN

$$x(n) = \theta + \eta(n) \qquad n = 0, 1, \ldots N-1 \qquad \eta(n) \sim N(0, \sigma^2)$$

$$f_{x|\theta}(x(n)|\theta) = f_\eta(x(n)-\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(x(n)-\theta)^2\right]$$

$$f_{\underline{x}|\theta}(\underline{x}|\theta) = \prod_{n=0}^{N-1} f_{x|\theta}(x(n)|\theta) = \prod_{n=0}^{N-1} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(x(n)-\theta)^2\right]$$

$$= \frac{1}{(2\pi\sigma^2)^{N/2}} \cdot \exp\left[-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x(n)-\theta)^2\right]$$

Aplicando o logaritmo natural e derivando,

$$\frac{\partial \ln\left[f_{\underline{x}|\theta}(\underline{x}|\theta)\right]}{\partial\theta} = \frac{\partial}{\partial\theta}\left[-\ln\left[(2\pi\sigma^2)^{N/2}\right] - \frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x(n)-\theta)^2\right]$$

$$= \frac{1}{\sigma^2}\sum_{n=0}^{N-1}(x(n)-\theta) = \frac{N}{\sigma^2}(\bar{x}-\theta)$$

$$\bar{x} = \frac{1}{N}\sum_{n=0}^{N-1} x(n)$$

Igualando a derivada a 0:

$$\frac{N}{\sigma^2}(\bar{x}-\hat{\theta}) = 0 \Rightarrow \boxed{\hat{\theta} = \bar{x}}$$

2º caso : $\underline{\theta}$ é um vetor aleatório.

- É necessário um conhecimento estatístico sobre as V.As em $\underline{\theta}$: $f_\Theta(\underline{\theta})$

- Abordagem Bayesiana de estimação.

Veremos dois representantes desta família de estimadores: MAP e MMSE

(1) Estimador de Máxima Probabilidade a Posteriori (MAP)

- Dado o conjunto de observações $\underline{x}$, buscamos o vetor de parâmetros θ com a maior probabilidade $f_{\Theta|X}(\underline{\theta}|\underline{x})$ de ter gerado o conjunto de amostras.

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

$\underline{\theta}$ uniforme

MAP = ML

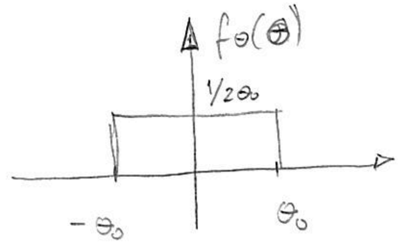
Solução obtida a partir de

$$\frac{\partial f_{X|\Theta}(\underline{x}|\underline{\theta})}{\partial \underline{\theta}} = 0$$

Exemplo: constante + WGN

$$f(\theta|\underline{x}) = \frac{f(\underline{x}|\theta)\, f(\theta)}{f(\underline{x})} = \frac{f(\underline{x}|\theta)\, f_\theta(\theta)}{\int_{-\infty}^{\infty} f(\underline{x}|\theta)\, f(\theta)\, d\theta}$$

$$f_\theta(\theta) = \begin{cases} \frac{1}{2\theta_0}, & |\theta| \le \theta_0 \\ 0, & |\theta| > \theta_0 \end{cases}$$



$$f(\theta|\underline{x}) = \begin{cases} \dfrac{\frac{1}{2\theta_0(2\pi\sigma^2)^{N/2}} \exp\left[-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x(n)-\theta)^2\right]}{\int_{-\theta_0}^{\theta_0} \frac{1}{2\theta_0(2\pi\sigma^2)^{N/2}} \exp\left[-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x(n)-\theta)^2\right] d\theta}, & |\theta| \le \theta_0 \\ 0, & |\theta| > \theta_0 \end{cases}$$

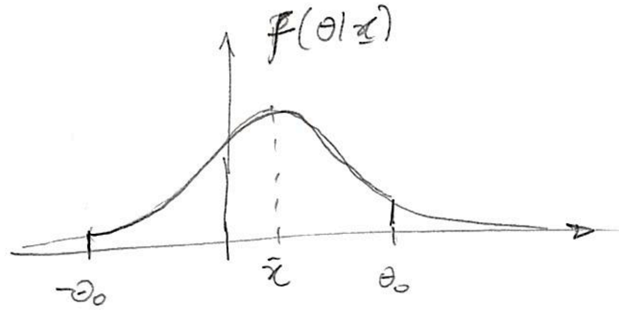
mas

$$\sum_{n=0}^{N-1}(x(n)-\theta)^2 = \sum_{n=0}^{N-1} x^2(n) - 2N\theta\bar{x} + N\theta^2 = N(\theta-\bar{x})^2 + \sum_{n=0}^{N-1} x^2(n) - N\bar{x}^2$$

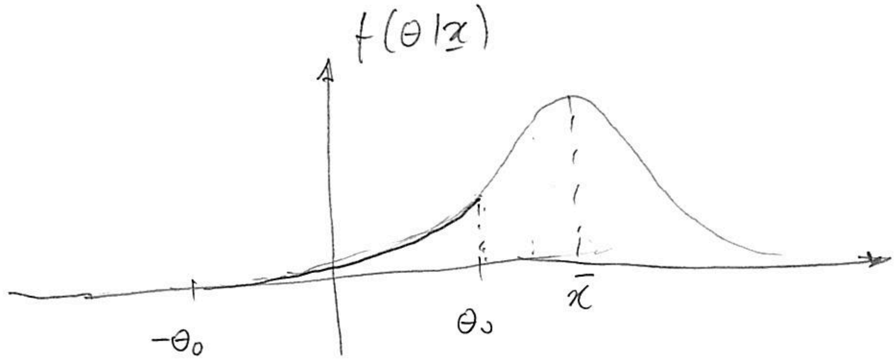
$$f(\theta|\underline{x}) \doteq \begin{cases} \frac{1}{c\sqrt{2\pi\frac{\sigma^2}{N}}} \exp\left[-\frac{1}{2\frac{\sigma^2}{N}}(\theta-\bar{x})^2\right], & |\theta| \le \theta_0 \\ 0, & |\theta| > \theta_0 \end{cases}$$

$$c = \int_{-\theta_0}^{\theta_0} \frac{1}{\sqrt{2\pi\frac{\sigma^2}{N}}} \exp\left[-\frac{1}{2\frac{\sigma^2}{N}}(\theta-\bar{x})^2\right] d\theta$$

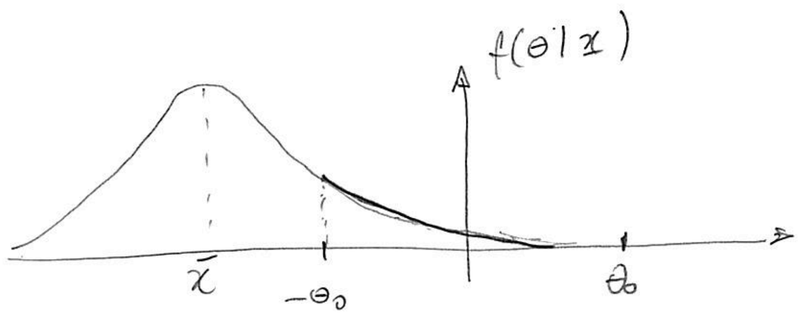
- $-\theta_0 \le \bar{x} \le \theta_0$:



- $\bar{x} > \theta_0$



- $\bar{x} < -\theta_0$



$$\hat{\theta}_{MAP} = \begin{cases} -\theta_0, & \bar{x} < -\theta_0 \\ \bar{x}, & -\theta_0 \le \bar{x} \le \theta_0 \\ \theta_0, & \bar{x} > \theta_0 \end{cases}$$

(7)L

(2) Estimador de mínimo erro quadrático médio (Minimum Mean-Squared Error, MMSE)

$$J_{MSE}(\underline{\hat{\theta}}) = E\left\{ \|\underline{\theta} - \underline{\hat{\theta}}\|^2 \right\}$$

Como $\underline{\theta}$ é um vetor de V.A.s, o operador de esperança leva em conta a p.d.f. conjunta $f_{X,\theta}(\underline{x},\underline{\theta})$

$$J_{MSE}(\underline{\hat{\theta}}) = \iint \|\underline{\theta} - \underline{\hat{\theta}}\|^2 f_{X,\theta}(\underline{x},\underline{\theta})\, d\underline{x}\, d\underline{\theta}$$
$$= \int \left\{ \int \|\underline{\theta} - \underline{\hat{\theta}}\|^2 f_{\theta|X}(\underline{\theta}|\underline{x})\, d\underline{\theta} \right\} f_X(\underline{x})\, d\underline{x}$$

Como $f_X(\underline{x}) \geq 0$, $J_{MSE}(\underline{\hat{\theta}})$ será minimizado se o termo entre $\{\cdot\}$ for minimizado para cada valor de $\underline{x}$:

$$\frac{\partial}{\partial \hat{\theta}} \int \|\theta - \hat{\theta}\|^2 f_{\theta|X}(\underline{\theta}|\underline{x})\, d\underline{\theta} = -2\int \theta\, f_{\theta|X}(\underline{\theta}|\underline{x})\, d\underline{\theta} + 2\hat{\theta} \underbrace{\int f_{\theta|X}(\underline{\theta}|\underline{x})\, d\underline{\theta}}_{=1}$$

Igualando a derivada a zero, temos:

$$\hat{\theta}_{MSE} = \int_{-\infty}^{\infty} \underline{\theta}\, f_{\theta|X}(\underline{\theta}|\underline{x})\, d\underline{\theta} = E\left\{ \underline{\theta} | \underline{x} \right\}$$

Em geral, o estimador MMSE tem natureza não-linear e é de difícil obtenção.

Alternativa: supor que $\underline{\hat{\theta}}$ é função linear das observações → Filtragem linear ótima

Exemplo :

$$\hat{\theta}_{MMSE} = E\{\theta \mid \underline{x}\}$$

$$= \int_{-\infty}^{\infty} \theta \, f(\theta \mid \underline{x}) \, d\theta$$

$$= \frac{\displaystyle\int_{-\theta_0}^{\theta_0} \theta \cdot \frac{1}{\sqrt{2\pi \frac{\sigma^2}{N}}} \exp\left[-\frac{1}{2\frac{\sigma^2}{N}} (\theta - \bar{x})^2\right] d\theta}{\displaystyle\int_{-\theta_0}^{\theta_0} \frac{1}{\sqrt{2\pi \frac{\sigma^2}{N}}} \exp\left[-\frac{1}{2\frac{\sigma^2}{N}} (\theta - \bar{x})^2\right] d\theta}$$