TRATAMIENTO DIGITAL DE SEÑALES
Ingeniería de Telecomunicación (4º, 2º c)

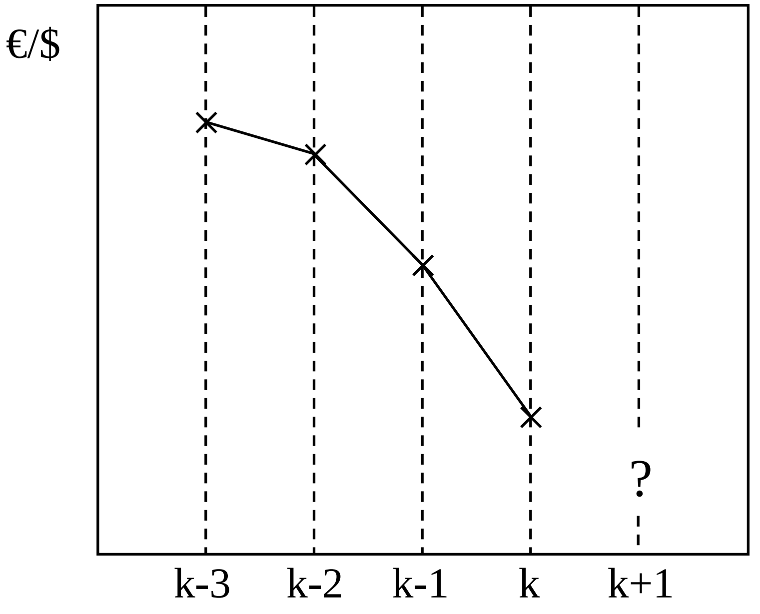
Unidad 3ª: Introducción a la estimación

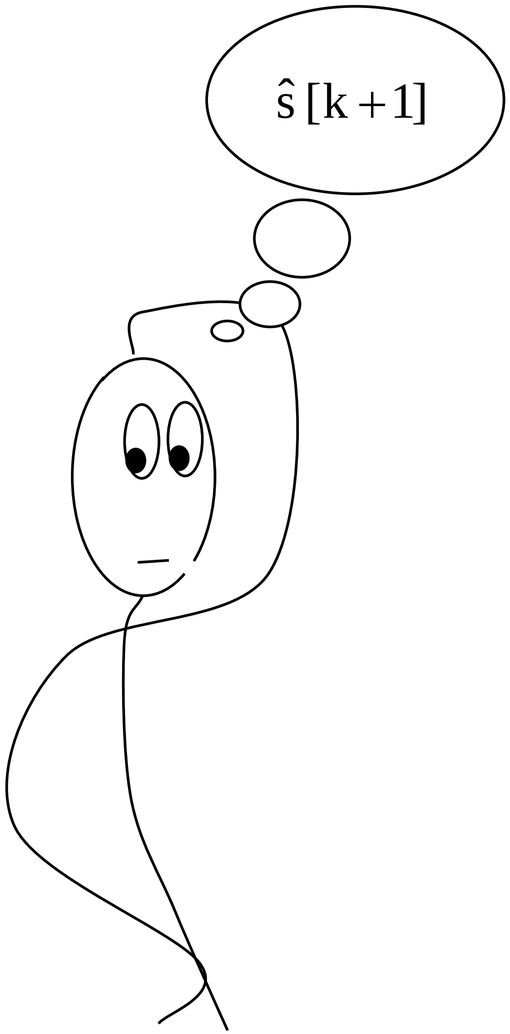
Aníbal R. Figueiras Vidal
Jesús Cid Sueiro
Ángel Navia Vázquez

Área de Teoría de la Señal y Comunicaciones
Universidad Carlos III de Madrid

Un problema de predicción

Predicción del tipo de cambio de divisas





- La utilidad es múltiple: desde operaciones especulativas hasta previsiones sobre la economía de los países.

- En cualquier caso, se pretende obtener $\hat{s}[k+1]$ <u>lo más parecido</u> posible a $s[k+1]$ (la aplicación concreta que se considere puede influir en la elección de la medida) utilizando los <u>datos</u> (y/o el <u>conocimiento</u>) disponibles.

- Así son los problemas de **estimación** (**predicción**, si se estima un valor futuro): la diferencia fundamental con los de clasificación es que el resultado es un <u>valor estimado</u> contínuo $\hat{s}$ ($\hat{\mathbf{s}}$).

- Obviamente, el <u>estimador</u> es una función continua $f_{\mathbf{w}}(\mathbf{x})$.

- La predicción es una de las tareas típicas del **Análisis de Series** Temporales (y Espaciales), parte de la Estadística clásica. También del Tratamiento (Digital) de Señales.

- En ocasiones, no se realiza la predicción en sentido estricto: p. ej., para inversión bursátil interesa saber cuándo conviene comprar o vender, y se recurre a manejar "predictivamente" los datos para construir **indicadores** que marcan los correspondientes instantes (entre ellos, se mantiene la posición); en definitiva, se decide. Esto constituye lo que se llama **Análisis Técnico**: el **Fundamental** se apoya en consideraciones de teoría económica.

Discusión

Q: ¿Es razonable realizar una predicción a partir de únicamente el registro del valor a predecir?

Salvo que sea una variable "físicamente" aislada, importan:

- *variables "exógenas": cuyos valores se generan independientemente de la considerada, pero influyen sobre ésta;*
- *variables análogas: que evolucionan siguiendo mecanismos en parte comunes con el que rige la variable considerada.*

Q: ¿Es natural tomar los valores más recientes del registro de una variable como observaciones para su predicción? Si cree que no, dé un ejemplo.

Si un proceso tiene una fuerte componente cíclica, son (en principio) relevantes las observaciones separadas por un periodo. En general, el problema de selección de variables está abierto para cada caso que se considere.

Q: Señale qué diferencias existen entre las consideraciones destinadas a elegir f_w *respecto a las de elección de* F_w.

Conceptualmente, ninguna: sólo que se trata de una función que toma valores contínuos, y no discretos.

Q: ¿Se puede resolver un problema de estimación de una variable multidimensional **s** *mediante la resolución separada de la estimación de cada una de sus componentes?*

Sólo si la medida de similitud entre **s** *y* $\hat{\mathbf{s}}$ *se puede descomponer en suma de medidas de similitud de sus componentes; lo que se da, p. ej., para el error cuadrático.*

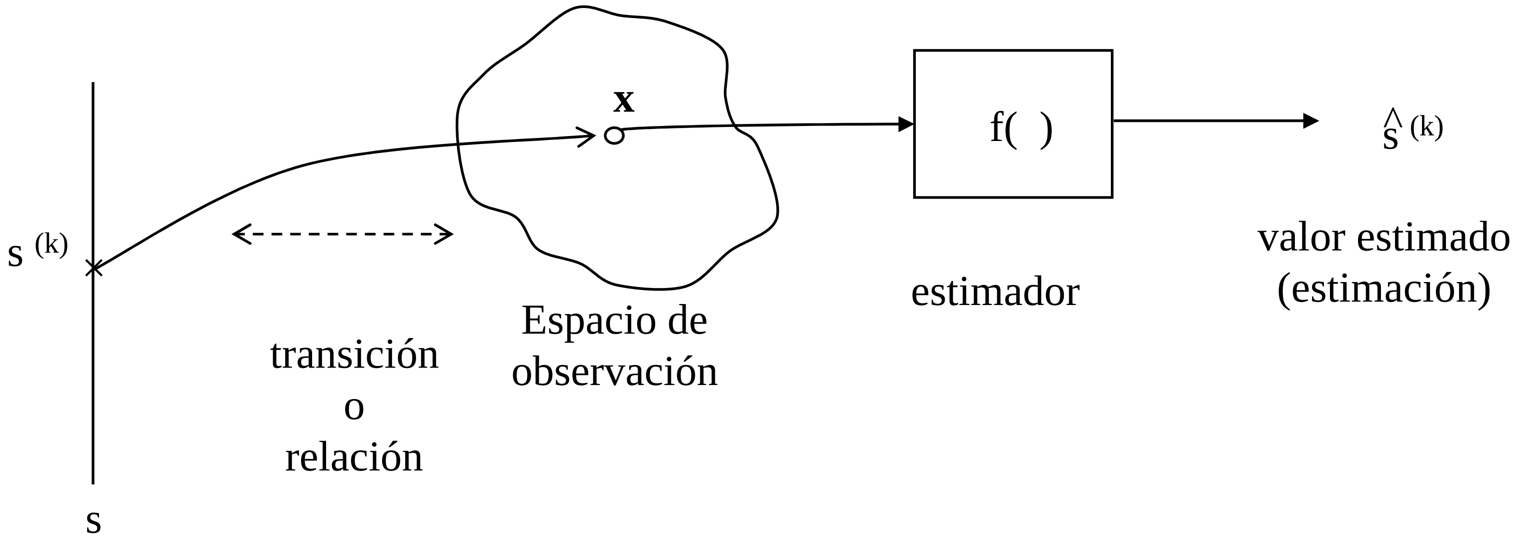
T: Comente el modo de funcionamiento de algún indicador para inversión bursátil.

T: Otros problemas que se plantean en términos de estimación son:

- *la medida de una magnitud física*
- *la determinación de los parámetros de un modelo*
- *la recuperación de sincronismo (de portadora, de símbolo, ...) en una transmisión*
- *la predicción del consumo eléctrico en una cierta zona*
- *el trazado de un enlace ferroviario*

Discuta los aspectos generales de dichos problemas.

Visión de la estimación



Formulación analítica

Se suponen conocidos (habitualmente)

* un **coste** $C(s, \hat{s})$ (no negativo)
  típicamente: $C(s, \hat{s}) \rightarrow C(s - \hat{s}) = C(e)$;   e: **error** de estimación

* la **verosimilitud** $p(\mathbf{x}|s)$

* si s es una v. a.: su ddp "**a priori**", $p(s)$

De lo anterior (si s es v.a.): $p(\mathbf{x} \mid s)\, p(s) = p(\mathbf{x}, s)$

$$\int_{(s)} p(\mathbf{x} \mid s)\, p(s) ds = p(\mathbf{x})$$

$$p(s \mid \mathbf{x}) = \frac{p(\mathbf{x} \mid s) p(s)}{p(\mathbf{x})} = \frac{p(\mathbf{x} \mid s) p(s)}{\int_{(s)} p(\mathbf{x} \mid s)\, p(s) ds}$$

Ejercicios de Ampliación

A: Sea s una va.

Compare la elección de $\hat{s}$ mediante la minimización del coste medio global, $\overline{C}(\hat{s})$, y la del coste medio a la vista de la observación, $\overline{C}(\hat{s}\mid \boldsymbol{x})$

$$\overline{C}(\hat{s}) = E_{x,s}\{C(s,\hat{s}(\boldsymbol{x}))\} = \int\!\!\int C(s,\hat{s})\, p(\boldsymbol{x},s)\, d\boldsymbol{x}\, ds =$$

$$= \int p(\boldsymbol{x})\left[\int C(s,\hat{s})\, p(s\mid \boldsymbol{x})\, ds\right] d\boldsymbol{x}$$

*El paréntesis cuadrado es $\overline{C}(\hat{s}\mid \boldsymbol{x})$: si se minimiza para cada **x**, se minimiza $\overline{C}(\hat{s})$.*

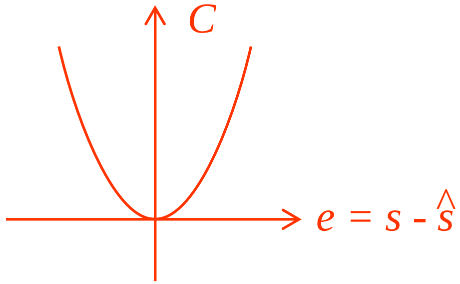
*Esta es la base de la **Teoría Bayesiana** de la Estimación.*

A: Sea s una va.

Determine el estimador $\hat{s}_{ms}$ que minimiza el error cuadrático medio (MMSE) $C(e) = e^2$.

$$\min_{\hat{s}} \overline{C}(\hat{s} \mid \mathbf{x}) = \min_{\hat{s}} \int (s-\hat{s})^2 p(s \mid \mathbf{x}) ds$$

$$\left.\frac{\partial \overline{C}}{\partial \hat{s}}\right|_{\hat{s}=\hat{s}_{ms}} = -2\int (\hat{s}-\hat{s}_{ms}) p(s \mid \mathbf{x}) ds = 0$$

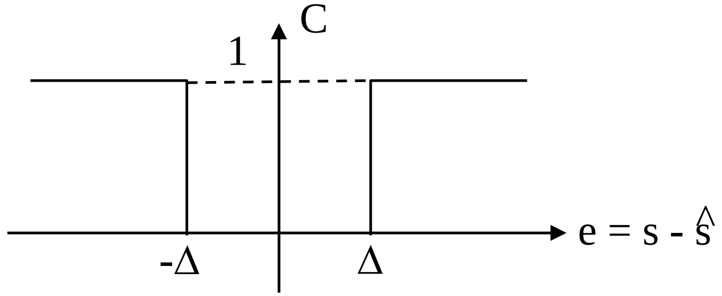


$$\hat{s}_{ms} \int p(s \mid \mathbf{x}) ds = \hat{s}_{ms} = \int s p(s \mid \mathbf{x}) ds = \underline{E\{s \mid \mathbf{x}\}} : \textbf{media a posteriori}$$

$$\frac{\partial^2 \overline{C}}{\partial \hat{s}^2} = 2\int p(s \mid \mathbf{x}) ds = 2 > 0 \quad : \quad \text{es un mínimo}$$

(la minimización del momento de segundo orden es una propiedad conocida de la media)

A: *Determine el estimador $\hat{s}_{map}$ de una va. s que minimiza la función de coste de la figura*

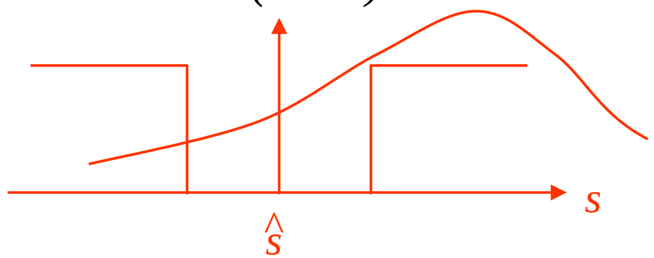


con $\Delta \to 0$ (suficientemente pequeño respecto al ritmo de variación de $p(s\,|\,\mathbf{x})$).

*Justifique su denominación: "**Máximo A Posteriori**" (MAP).*

Es claro que se trata de

$$\underset{\hat{s}}{\text{máx}} \int_{\hat{s}-\Delta}^{\hat{s}+\Delta} p(s\,|\,\mathbf{x})\,ds$$



que, con $\Delta \to 0$, conduce a: $\hat{s}_{map} = arg\left\{ \underset{s}{\text{máx}}\, p(s\,|\,\mathbf{x}) \right\} = mod(s\,|\,\mathbf{x})$

moda a posteriori

Suele ser útil: $\hat{s}_{map} = arg\left\{ \underset{s}{\text{máx}} \ln p(s\,|\,\mathbf{x}) \right\}$

A: *Se denomina estimador de* ***Máxima Verosimilitud*** *(ML) al que se obtiene como*

$$\hat{s}_{ml} = arg \left\{ \underset{s}{máx} (ln) p(\mathbf{x} \mid s) \right\}$$

a) Si s es una va., relacione $\hat{s}_{ml}$ *con* $\hat{s}_{map}$

b) ¿Puede emplearse $\hat{s}_{ml}$ *en algún otro caso?*

a) $ln\ p(s \mid \mathbf{x}) = ln\ p(\mathbf{x} \mid s) + ln\ p(s) - ln\ p(\mathbf{x})$

$arg \{ \underset{s}{máx}\ ln\ p(s \mid \mathbf{x})\} = arg \{ \underset{s}{máx}\ [ln\ p(\mathbf{x} \mid s) + ln\ p(s)]\}$

y, si se desconoce p(s), lo lógico es no hacer intervenir p(s): pasando así del $\hat{s}_{map}$ *al* $\hat{s}_{ml}$.

b) *Es obvio que* $p(\mathbf{x} \mid s)$ *existe aún si s es una* <u>*variable determinista*</u>*: por lo que puede emplearse para su estimación (ML).*

A: *Determine las estimaciones ML de la media y la varianza de una va. x con distribución G(m, v) de la que se hacen K observaciones independientes, $\{x^{(k)}\}$.*

(Nótese que m y v son deterministas; hacen el papel de s)

$$G(m,v): \qquad \frac{1}{\sqrt{2\pi v}} exp\left[-\frac{(x-m)^2}{2v}\right]$$

por ser las observaciones independientes

$$p(\mathbf{x}\,|\,s) = \prod_{k=1}^{K} \frac{1}{\sqrt{2\pi v}} exp\left[-\frac{\left(x^{(k)}-m\right)^2}{2v}\right]$$

$$ln\ p(\mathbf{x}\,|\,s) = \sum_{k=1}^{K}\left[-\frac{1}{2} ln\, 2\pi v - \frac{\left(x^{(k)}-m\right)^2}{2v}\right]$$

$$\frac{\partial ln}{\partial m} = \sum_{k=1}^{K}\frac{1}{v}\left(x^{(k)}-m\right); \qquad \left.\frac{\partial ln}{\partial m}\right|_{m=\hat{m}_{ml}} = 0 = \sum_{k=1}^{K}\left(x^{(k)}-\hat{m}_{ml}\right) \qquad \left(\frac{\partial^2 ln}{\partial m^2} = -\frac{K}{v} < 0\right)$$

$$\underline{\hat{m}_{ml} = \frac{1}{K}\sum_{k=1}^{K} x^{(k)}} \quad : \quad \textbf{media muestral} \quad \textit{(nótese su linealidad)}$$

$$\frac{\partial \ln}{\partial v} = \sum_{k=1}^{K}\left[-\frac{1}{2v} + \frac{\left(x^{(k)}-m\right)^2}{2v^2}\right] \quad ; \quad \left.\frac{\partial \ln}{\partial v}\right|_{\substack{v=\hat{v}_{ml} \\ m=\hat{m}_{ml}}} = 0 = -K + \frac{1}{\hat{v}_{ml}}\sum_{k=1}^{K}\left(x^{(k)}-\hat{m}_{ml}\right)^2$$

$$\left(\frac{\partial^2 \ln}{\partial v^2} = -\frac{1}{v^2}\sum_{k=1}^{K}\left(x^{(k)}-m\right)^2 < 0\right)$$

$$\hat{v}_{ml} = \frac{1}{K}\sum_{k=1}^{K}\left(x^{(k)}-\hat{m}_{ml}\right)^2 : \qquad \textbf{varianza muestral}$$

(si m estuviese dado, aparecería en $\hat{v}_{ml}$ *en el lugar de* $\hat{m}_{ml}$*).*

** Nótese que la estimación de la media equivale a*

$$\min_{m}\sum_{k=1}^{K}\left(x^{(k)}-m\right)^2$$

minimizar el error cuadrático muestral: de este hecho se deriva un conjunto de métodos de estimación de parámetros deterministas que se conocen como ***métodos de mínimos cuadrados*** *("Least Squares", LS) (¡no "medios"!)*

* *Nótese que para calcular $\hat{m}_{ml}$ y $\hat{v}_{ml}$ basta con conocer $\sum x^{(k)}$ y $\sum x^{(k)2}$; sin que sea necesario conocer por separado los valores de $\{x^{(k)}\}$. A este tipo de estadísticos que pueden sustituir a las observaciones se los conoce como **estadísticos suficientes**: su uso puede simplificar grandemente problemas de estimación y decisión.*

* *En el caso multidimensional, se obtendrían:*

$$\hat{\mathbf{m}}_{ml} = \frac{1}{K}\sum_{k=1}^{K} \mathbf{x}^{(k)}$$

$$\hat{\mathbf{V}}_{ml} = \frac{1}{K}\sum_{k=1}^{K} \left(\mathbf{x}^{(k)} - \hat{\mathbf{m}}_{ml}\right)\left(\mathbf{x}^{(k)} - \hat{\mathbf{m}}_{ml}\right)^T$$