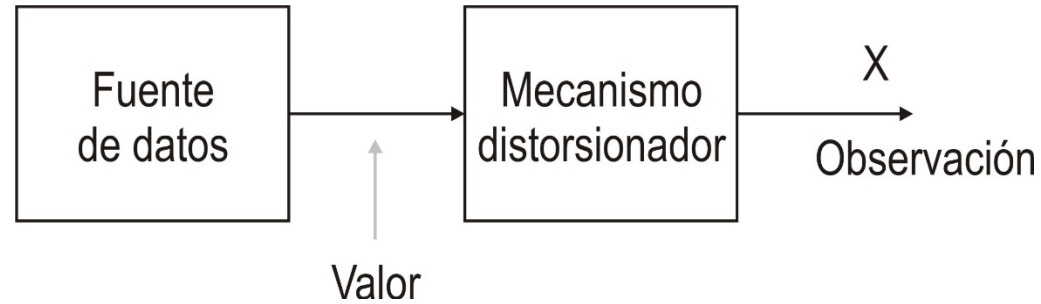


# Estimación

Estimadores Bayesianos

- Problema genérico



- La fuente puede generar un conjunto de valores o incluso cualquier valor.
- **Objetivo:** determinar una **estimación** del valor original a partir de la observación realizada de forma que se cometa el menor error
  - Basado en:
    - Conocimiento de las características de la fuente de datos
    - Conocimiento de las características del mecanismo distorsionador (ej.: canal, ruido, etc.)

- La **ESTIMACIÓN** se basa en encontrar el valor  $s$  de la variable aleatoria  $S$  que maximice  $f(s|x)$ 
  - Estimador MAP: tendrá en cuenta el conocimiento de la fuente:  $f(s)$
  - Estimador ML: no tendrá en cuenta o no conocerá la fuente
- Estos estimadores también se basarán en criterios bayesianos, tratando de minimizar una **función de coste** según un criterio de error preestablecido, y empleando herramientas estadísticas.

## Formalización

- Modelo del problema general de estimación
- Construiremos el modelo empleando un ejemplo como guía:

*“Queremos medir un voltaje  $s$  en un instante determinado. A partir de consideraciones físicas conocemos que el voltaje está entre  $-V$  y  $+V$  voltios. La medida está distorsionada por ruido que puede ser modelado como una variable aleatoria gaussiana de media nula, aditiva, e independiente ( $n$ ).”*

La observación realizada es

$$x = s + n$$

La fdp que gobierna el proceso de observación es

$$f(x|s) = \frac{1}{\sqrt{2\pi}\sigma_n} e^{-\frac{1}{2}\left(\frac{x-s}{\sigma_n}\right)^2}$$

El problema consiste en observar  $x$  y estimar  $s$

## Formalización

- El modelo del problema general de estimación tiene 4 componentes:

- **Espacio de parámetros**

La salida de la fuente es un parámetro o variable que podemos ver como un punto en el espacio de parámetros. En el caso de estimación de parámetros escalares, este espacio corresponderá a segmentos del eje real

En el ejemplo:  $[-V, V]$

- **Mapeo probabilístico del espacio de parámetros al espacio de observación**

Es la ley probabilística que gobierna el efecto de  $s$  en la observación realizada

- **Espacio de observación**

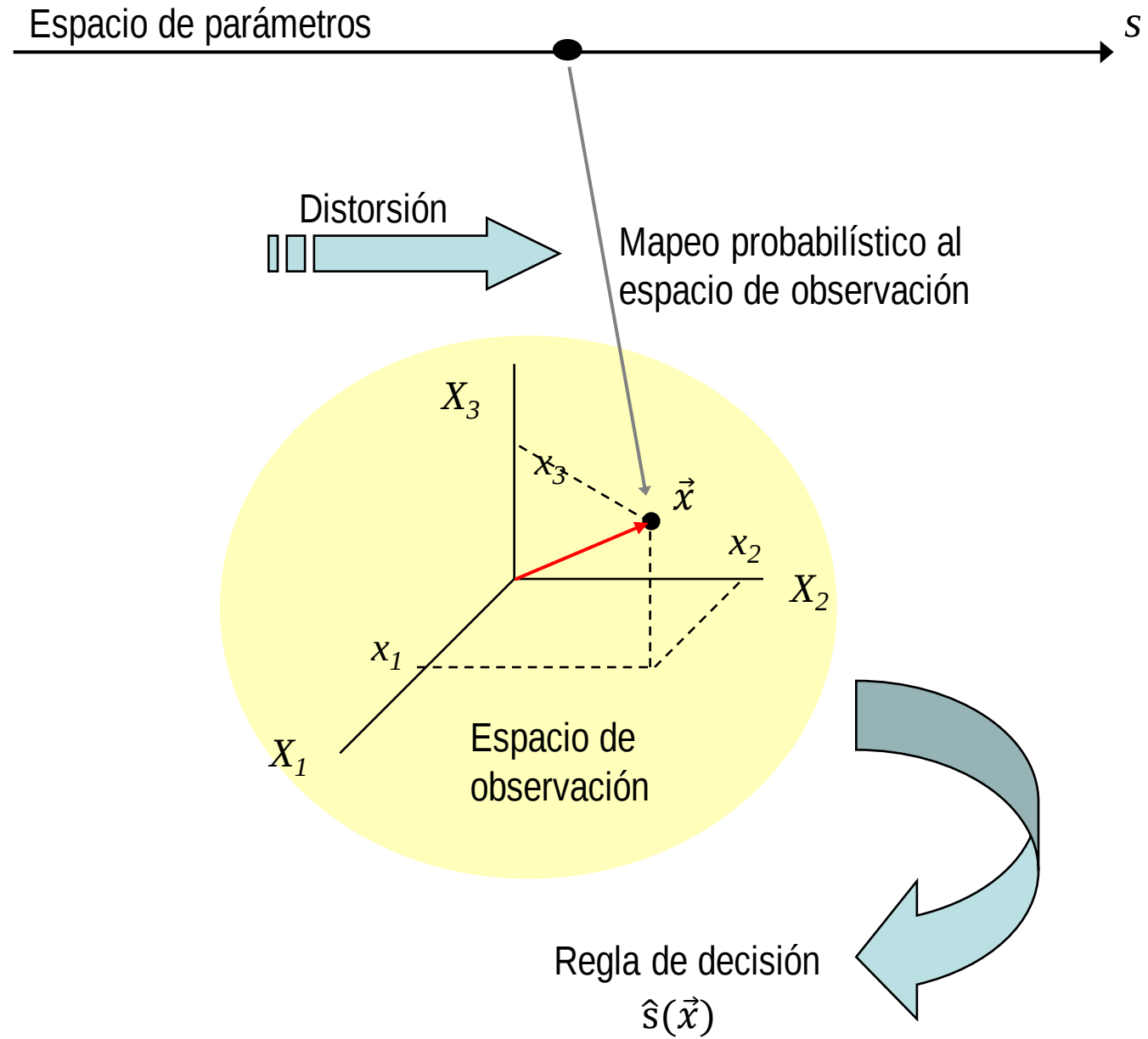
En el problema clásico, es el espacio de dimensión finita en el que se encuentra el vector observado

En el ejemplo, si únicamente se observa una vez la tensión, es de una dimensión. Si no, de tantas como muestras se tomen

- **Regla de estimación**

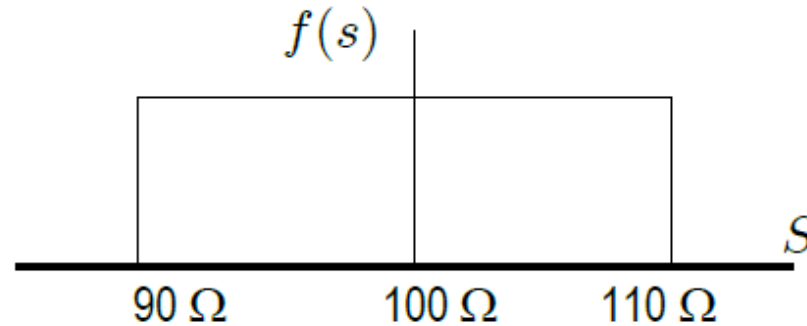
Mapeo de la observación en una estima:  $\hat{s}(\vec{x})$

# Formalización



## Planteamiento formal del problema

- Podemos partir del caso en el que conocemos la función densidad de probabilidad del fenómeno cuyo valor queremos estimar:  $f(s)$
- Ejemplo: valor de una resistencia de  $100\ \Omega$  con tolerancia 10%.



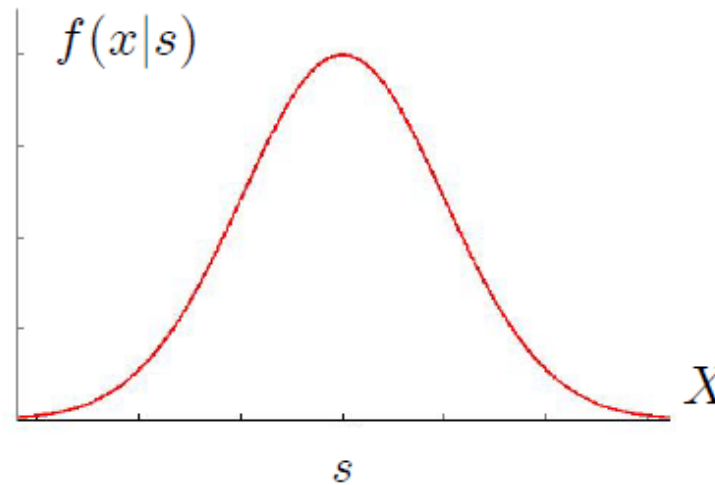
- La observación que realizamos es una variable aleatoria  $X$ , que, lógicamente, estará relacionada estadísticamente con  $S$ :

$$f(x|s)$$

- Dado que se ha producido un cierto valor  $s$  que quiero estimar, mis observaciones (variable aleatoria) seguirán una cierta distribución.

## Planteamiento formal del problema

- Ejemplo: si el ruido es gaussiano de media cero:



- Una vez que se ha producido un valor  $s$ , aunque no se conozca, tendremos una variable aleatoria gaussiana sumada a ese valor constante.
- Al igual que en el problema de decisión, mis observaciones tendrán las propiedades del ruido, excepto la media que será la suma de la media del ruido y el valor  $s$  que ha tomado la variable aleatoria  $S$ .

## Planteamiento formal del problema

- En este ejemplo, lo que pretendemos es encontrar el valor de  $s$ .
- Buscamos el valor  $s$  que hace máxima la fdp condicional  $f(s|x)$
- Aplicando el Teorema de Bayes:

$$f(s|x) = \frac{f(x|s)f(s)}{f(x)}$$

$f(x|s)$  : relación estadística entre  $X$  y  $S$

$f(s)$  : información a priori sobre  $S$

$f(x)$  : interviene exclusivamente como un factor de escala

- La forma de  $f(s|x)$ , que es la que nos interesa estudiar, dependerá del problema:

$$f(s|x) = K \cdot f(x|s) \cdot f(s)$$

$$f(s|x) = K \cdot f(x|s) \cdot f(s)$$

- Si el estimador que empleamos es consistente, al aumentar el número de observaciones  $f(s|x)$  debería tender a ser una delta.
- Si el estimador además es insesgado, la delta estará situada exactamente en el valor  $s$  buscado.
- El conocimiento de  $f(s|x)$  nos va a permitir determinar cuál es el valor de  $s$  más probable.
  - Si se toma el valor que hace máxima  $f(s|x)$  estamos aplicando un criterio de estimación MAP
    - Tiene en cuenta la información a priori del fenómeno estudiado
  - En el caso del criterio de estimación de mínimo error cuadrático medio (MSE), se toma el valor

$$\hat{s}_{MSE} = E[s|x]$$

$$f(s|x) = K \cdot f(x|s) \cdot f(s)$$

- Si se descarta la información a priori, el máximo de la fdp condicional estudiada,  $f(s|x)$ , coincide (se aproxima) con el de  $f(x|s)$
- Este es el estimador ML (máxima verosimilitud)

$$\hat{s}_{ML} = \operatorname{argmax}\{f(x|s)\}$$

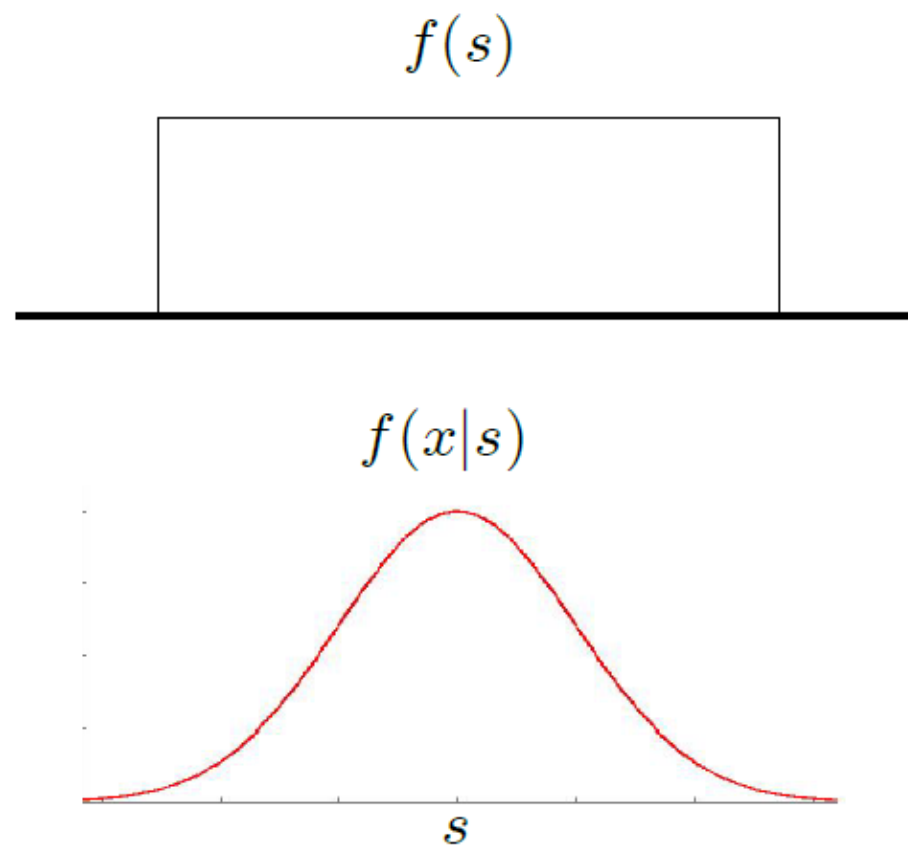
- Diferente del MAP:

$$\hat{s}_{MAP} = \operatorname{argmax}\{f(s|x)\}$$

- Se define como el valor de  $S$  que maximiza la función de verosimilitud. Esta maximización se realiza sobre todo el rango posible de valores de  $S$ .

## Planteamiento formal del problema

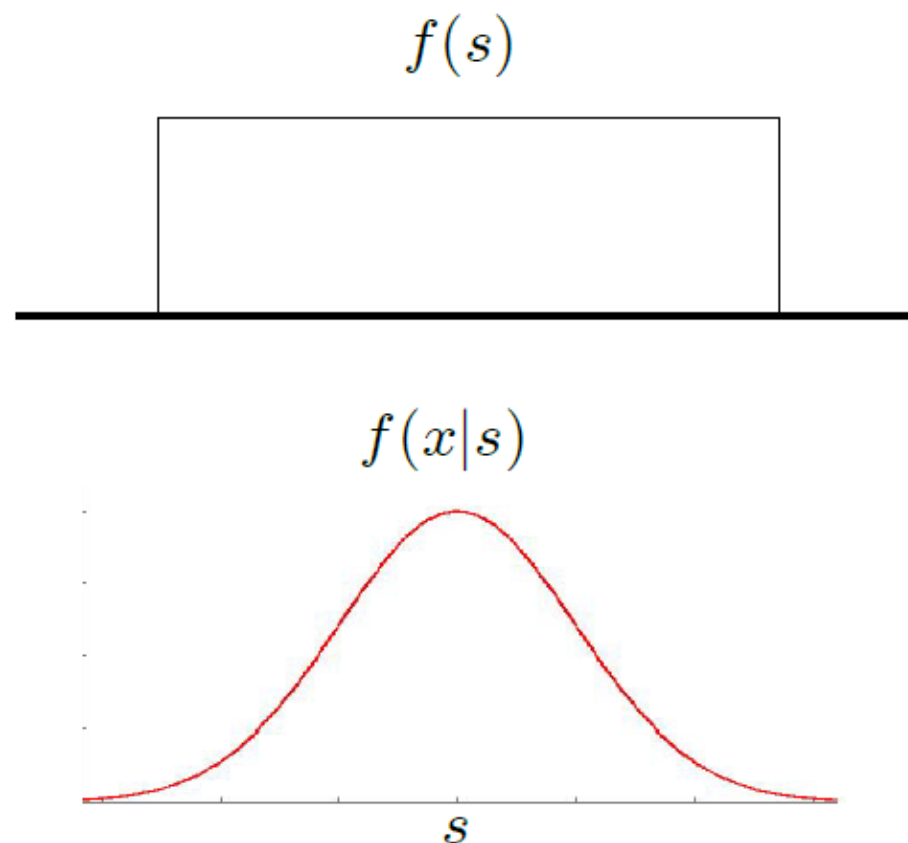
- Ejemplo MAP
  - Se desea estimar el valor que toma una variable aleatoria uniforme observada en ruido blanco gaussiano de media nula y varianza  $\sigma^2$



## Planteamiento formal del problema

## Planteamiento formal del problema

- Ejemplo MAP
  - Se desea estimar el valor que toma una variable aleatoria uniforme observada en ruido blanco gaussiano de media nula y varianza  $\sigma^2$



- ¿Qué sucedería si el ruido tuviese media no nula?

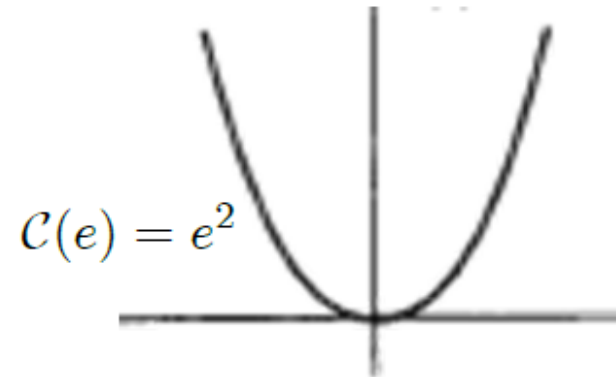
$$f(s|x) = K \cdot f(x|s) \cdot f(s)$$

- El conocimiento de  $f(s|x)$  nos va a permitir determinar cuál es el valor  $s$  óptimo según el criterio de error que se haya elegido.
- La estimación Bayesiana parte de la minimización de una función de coste (o riesgo) que depende del criterio de error.
- ¿Cómo es esa función de coste?

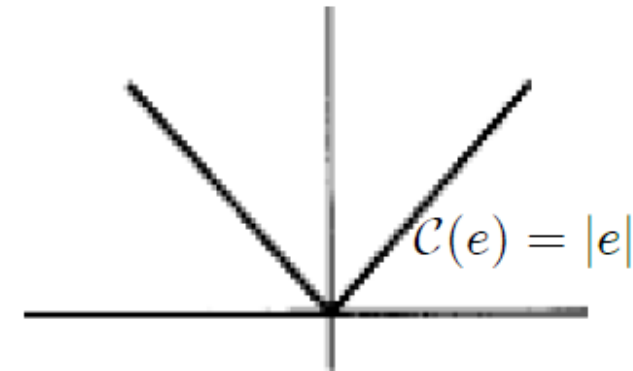
- Función de riesgo Bayesiana:

$$\mathcal{R} = E[\mathcal{C}(e)]$$

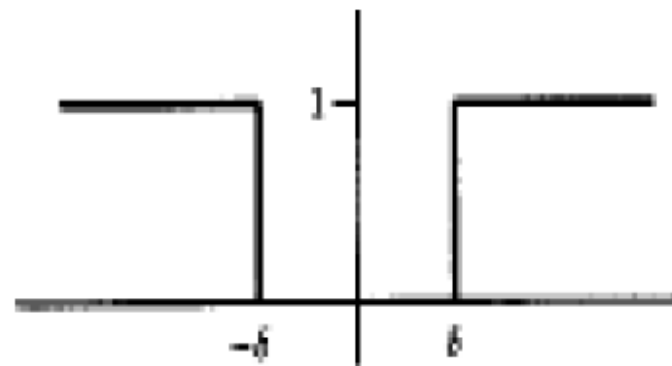
- Donde  $\mathcal{C}(e)$  es la función de coste empleada, dependiente del error  $e$ .



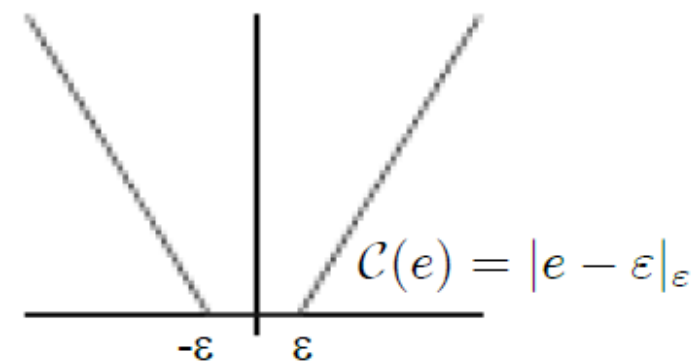
Error cuadrático



Error absoluto



Error de acierto o fallo  
(hit or miss)



Error  $\epsilon$ -insensitivo de Vapnik

## Estimadores Bayesianos

- El objetivo es encontrar la estima que minimiza el valor esperado del coste.
- El primer paso es escoger una función de coste adecuada, que permita tratar el problema
- Generalmente, es necesario un estudio del problema que se va a resolver para elegir bien la función de coste necesaria.
  - En la práctica, afortunadamente, una misma estima es óptima para muchas funciones de coste
- Del mismo modo que en el problema de decisión se disponía de las probabilidades a priori, ahora se dispone de un **densidad de probabilidad a priori**  $f(s)$ 
  - En decisión Bayesiana se asume conocida la probabilidad a priori
  - En estimación Bayesiana se asume conocida la densidad de probabilidad a priori
  - Si no fuese así, se trata de otro problema (ML)

- Definida una función de coste, el riesgo (coste) total viene dado por

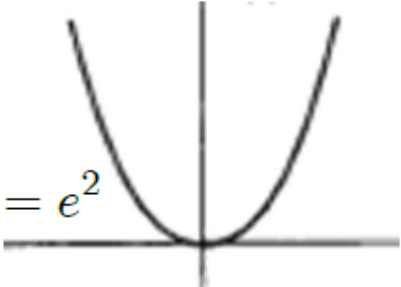
$$J \triangleq E[C(e)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} C(e) f(s, x) dx ds$$

donde  $e = s - \hat{s}(x)$

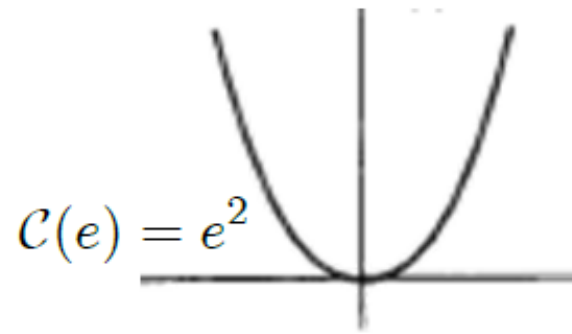
y la densidad de probabilidad conjunta es

$$f(s, x) = f(x) f(s|x)$$

- Si particularizamos estas expresiones para el caso MSE


$$J_{MSE} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^2 f(x) f(s|x) ds dx$$

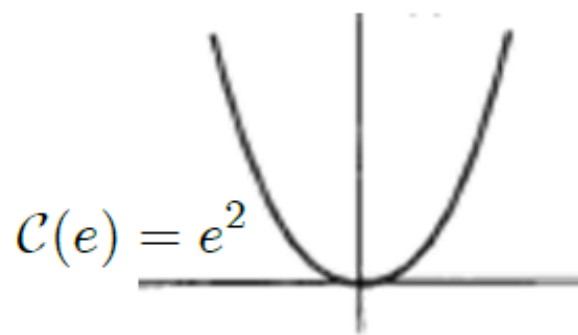
- Caso MSE



$$J_{MSE} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^2 f(x) f(s|x) ds dx$$

# Estimadores Bayesianos

- Caso MSE



$$J_{MSE} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^2 f(x) f(s|x) ds dx$$

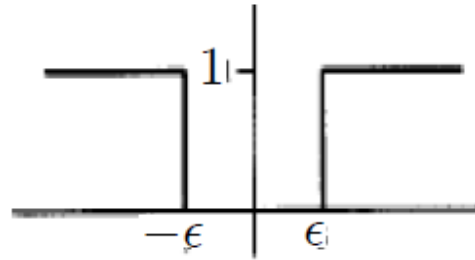
Minimizando el coste se obtiene

$$\hat{s}_{MSE}(x) = \int_{-\infty}^{\infty} s f(s|x) ds$$

que es lo mismo que

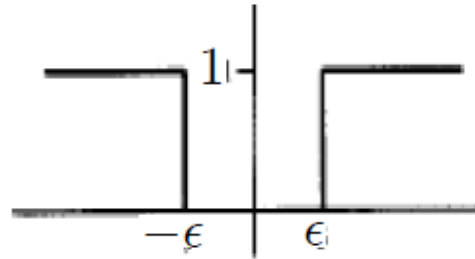
$$\hat{s}_{MSE}(x) = E[s|x]$$

- Caso “acierto o error”



# Estimadores Bayesianos

- Caso “acierto o error”



$$J = \int_{-\infty}^{\infty} f(x) \left[ 1 - \int_{\hat{s}(x) - \epsilon}^{\hat{s}(x) + \epsilon} f(s|x) ds \right] dx$$

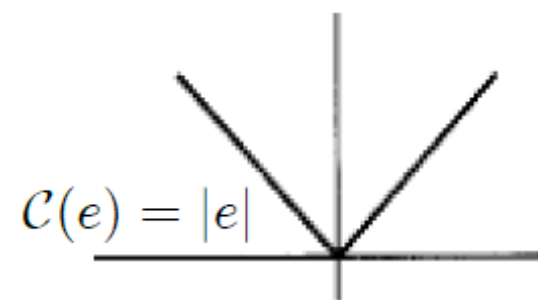
Para minimizar el coste se maximiza la integral interior.

Para ello:

$$\hat{s}_{MAP}(x) = \operatorname{argmax}\{f(s|x)\}$$

MAP = Maximum a Posteriori

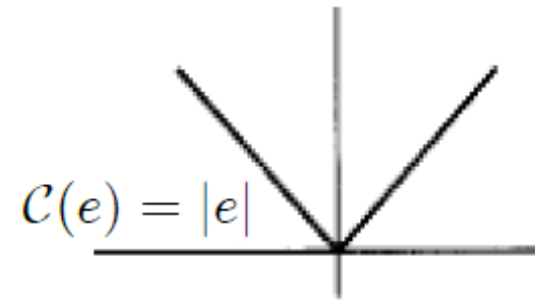
- Caso de valor absoluto del error



$$J_{MSE} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |e| f(x) f(s|x) ds dx$$

# Estimadores Bayesianos

- Caso de valor absoluto del error



$$J_{MSE} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |e| f(x) f(s|x) ds dx$$

Al minimizar se obtiene que

$$\hat{s}_{abs}(x) = \text{mediana}\{s|x\}$$

- En resumen:
  - El empleo de una función de coste cuadrática (mínimo error cuadrático medio) conduce a un estimador óptimo que es el valor medio de la fdp:

$$\hat{s}_{MSE}(x) = E[s|x]$$

- Una función de coste proporcional al error (error absoluto) resulta en un estimador óptimo que es la mediana de la fdp.

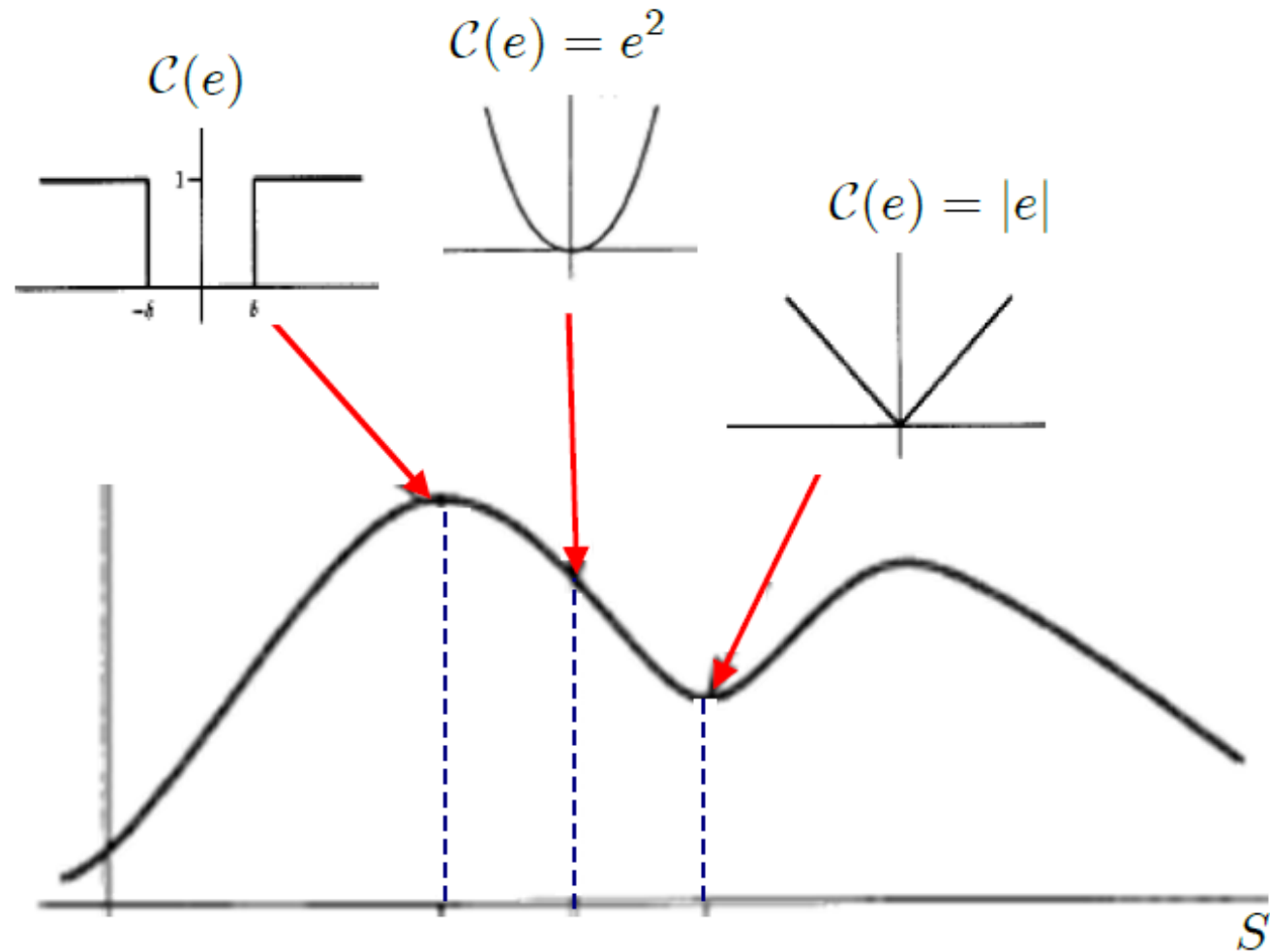
$$\hat{s}_{abs}(x) = \text{mediana}\{s|x\}$$

- Para la función de coste de “acierto o error”, el estimador óptimo es la moda (el valor máximo de la fdp).

$$\hat{s}_{MAP}(x) = \operatorname{argmax}\{f(s|x)\}$$

- Este caso se denomina MAP (Maximum a Posteriori)

- Si se dispone de una cierta fdp  $f(s|x)$ :



## Estimadores Bayesianos

- Los dos estimadores más empleados son el de mínima MSE y el MAP
- Cuando se emplea el criterio de mínima MSE se propone el valor que garantiza que el error cometido por la estima sea, en media, pequeño (supuestas varias estimaciones).
- Cuando se emplea el criterio MAP se pretende proponer como estima el valor que sea más probable, aún a riesgo de cometer un error grande si la estimación es errónea.

# Estimación

Estimadores Bayesianos