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

Unidad 12ª: Calidad de estimadores y decisores

Aníbal R. Figueiras Vidal
Jesús Cid Sueiro
Ángel Navia Vázquez

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

A. Parámetros de calidad en estimación

A.1. De un parámetro determinista

Una caracterización completa la proporcionaría p ($\hat{s}$| s):

- la calidad sería mayor con la concentración en torno a $\hat{s} = s$
- no proporciona una indicación global, sino microscópica (para cada s)

Habría que:

- calcularlo (a partir del conocimiento de la física del problema);
- estimarlo, en casos “máquina”: nótese que se trata de estimar una ddp para cada valor de s.

En general, resulta complicado.

Dada la lectura que se hace de p ($\hat{s}$ | s), una medida útil será el <u>error cuadrático condicional</u>

$$E\left\{(\hat{s}-s)^2 \,|\, s\right\} = \int_{-\infty}^{\infty} (\hat{s}-s)^2 \, p(\mathbf{x}\,|\,s)\, d\mathbf{x}$$

que suele descomponerse en la conocida forma

$$E\left\{(\hat{s}-s)^2 \,|\, s\right\} = E^2\left\{\hat{s}-s\,|\,s\right\} + \mathrm{Var}\left\{\hat{s}-s\,|\,s\right\}$$

* $E\left\{\hat{s}-s\,|\,s\right\} = E\left\{\hat{s}\,|\,s\right\} - s$ es el **sesgo** del estimador

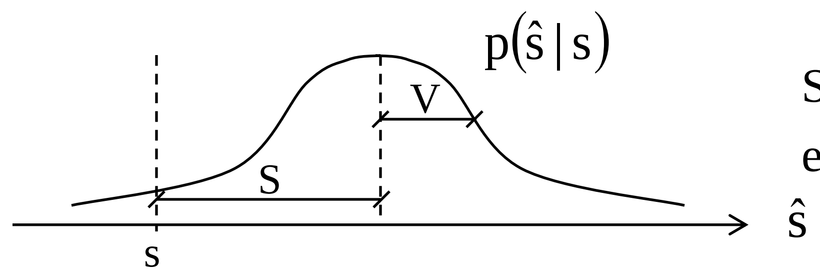
(tiene sentido de medida de error sistemático)

Si es nulo, el estimador se dice **insesgado**.

∗ $\mathrm{Var}\{\hat{s}-s \mid s\} = \mathrm{Var}\{\hat{s} \mid s\}$ es la **varianza** del estimador (del error) (tiene sentido de medida de dispersión, de error “aleatorio”)

Dado que nos movemos con información estadística, sólo puede ser nula cuando en la estimación se emplean $K \to \infty$ muestras: en cuyo caso el estimador se dice **consistente** en varianza

La (deseable) menor varianza se expresa como mayor **eficiencia**.



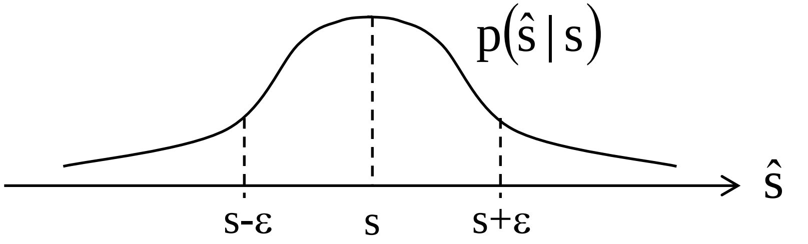
Sesgo(S) y varianza (V) en estimación

Como con la ddp condicional, esta(s) característica(s) puede(n), en principio, calcularse (en planteamientos analíticos) o estimarse (en planteamientos máquina); siendo cierto, en todo caso, que es conveniente proceder a cálculo y estimación para observar su concordancia.

Intervalos de confianza

Con la información estadística (o mediante estimación por frecuencias relativas) se puede manejar

$$\Pr\left\{\left|\hat{s}-s\right|<\varepsilon\right\}$$



que es (también) la probabilidad de que s se encuentre en el **intervalo** (aleatorio) $\left(\hat{s}-\varepsilon,\hat{s}+\varepsilon\right)$: asociado a esa (probabilidad de) **confianza**.

Esto tiene un sentido análogo al del error cuadrático condicional; en realidad, la **consistencia** en probabilidad

$$\Pr\left\{\left|\hat{s}-s\right|<\varepsilon\right\}\to 1$$

$$K\to\infty$$

está implicada por la consistencia en varianza, ya que la desigualdad de Tchebycheff implica

$$\Pr\left\{\left|\hat{s}-s\right|>\varepsilon\right\}\leq\frac{\mathrm{Var}\left\{\hat{s}\,|\,s\right\}}{\varepsilon^2}$$

Ejercicios

E: Calcúlense la media y la varianza condicionales del estimador muestral de la media de una va unidimensional.

$$E\left\{ \hat{m} \,|\, m \right\} = \frac{1}{K} \sum_{k=1}^{K} E\left\{ x^{(k)} \right\} = m$$

El estimador es insesgado.

Para muestras tomadas independientemente,

$$Var\left\{ \hat{m} \,|\, m \right\} = \frac{1}{K^2} \sum_{k=1}^{K} Var\left\{ x^{(k)} \right\} = \frac{v}{K}$$

El estimador es consistente en varianza.

Nótese que no hay dependencia de la distribución de las muestras.

E: Calcúlese la media del estimador muestral de la varianza de una va unidimensional.

(No emplearemos | m,v)

Como es

$$\hat{v} = \frac{1}{K}\sum_{k=1}^{K}\left(x^{(k)} - \hat{m}\right)^2 = \frac{1}{K}\left(\sum_{k=1}^{K} x^{(k)2} - 2\hat{m}\sum_{k=1}^{K} x^{(k)} + K\hat{m}^2\right) = \frac{1}{K}\sum_{k=1}^{K} x^{(k)2} - \hat{m}^2$$

suponiendo muestras independientes

$$E\{\hat{v}\} = \frac{1}{K}\sum_{k=1}^{K} E\left\{x^{(k)2}\right\} - E\left\{\hat{m}^2\right\} = Var\left\{x^{(k)}\right\} + m^2 - \left(Var\{\hat{m}\} + m^2\right) =$$

$$= v - \frac{v}{K} = \frac{K-1}{K}v$$

Es sesgado: pero asintóticamente insesgado

Compromiso sesgo-varianza

Se ha encontrado que, para el estimador muestral de la varianza,

$$E\{\hat{v}\} = \frac{K-1}{K} v$$

y puede pensarse en modificar el estimador para corregir el sesgo; lo que se consigue recurriendo al estimador muestral insesgado de v

$$\tilde{v} = \frac{K}{K-1}\hat{v} = \frac{1}{K-1}\sum_{k=1}^{K}\left(x^{(k)} - \hat{m}\right)^2$$

(téngase en cuenta que v no es conocida); pero así se tiene

$$\mathrm{Var}\{\tilde{v}\} = \left(\frac{K}{K-1}\right)^2 \mathrm{Var}\{\hat{v}\}$$

incrementándose la varianza.

(En realidad, no puede decirse que uno de estos estimadores sea mejor que el otro: están determinísticamente relacionados).

Lo otro que antecede es una manifestación del **compromiso sesgo-varianza** frecuentemente implícito en estimación.

A.2. <u>De un parámetro aleatorio</u>

El papel de $p(\hat{s}\,|\,s)$ lo haría aquí $p(\hat{s},\,s)$: que también debería concentrarse en torno a $\hat{s} = s$ (aunque es una ddp 2-D, y no una familia de ddp 1-D como $p(\hat{s}\,|\,s)$).

Error cuadrático, sesgo y varianza se dan promediados respecto a s: así,

$$E\left\{(\hat{s}-s)^2\right\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(\hat{s}-s)^2\,p(\mathbf{x}\,|\,s)\,p(s)\,d\mathbf{x}\,ds$$

que, como $E\{\hat{s}-s\} = E\{\hat{s}\} - E\{s\}$ y $\mathrm{Var}\{\hat{s}-s\}$ es un número y no una función de s.

A´. **Las Cotas de Cramer – Rao**

Establecen cotas inferiores para las varianzas de los errores de estimación.

A´.1. **Caso de parámetro determinista**

$$\operatorname{Var}\{\hat{s}-s\,|\,s\} \geq \frac{\left(1+\dfrac{d\,E\{\hat{s}-s\,|\,s\}}{d\,s}\right)^2}{E\left\{\left[\dfrac{\partial \ln p(\mathbf{x}\,|\,s)}{\partial s}\right]^2 |\,s\right\}}$$

$$= -\frac{\left(1+\dfrac{d\,E\{\hat{s}-s\,|\,s\}}{d\,s}\right)^2}{E\left\{\dfrac{\partial^2 \ln p(\mathbf{x}\,|\,s)}{\partial s^2}\,|\,s\right\}}$$

Se usa una u otra forma según resulte más cómodo.

La prueba se incluye en el Apéndice.

<u>Discusión</u> (para estimadores insesgados)

La cota sólo se alcanza si

$$\frac{\partial \ln p(\mathbf{x}\,|\,s)}{\partial s} = k(s)(\hat{s} - s)$$

entonces $\hat{s}$ es un estimador <u>absolutamente eficiente</u>.

– Supuesto que existe, cumplirá

$$\left.\frac{\partial \ln p(\mathbf{x}\,|\,s)}{\partial s}\right|_{s=\hat{s}_{ml}} = 0 = k(\hat{s}_{ml})(\hat{s} - \hat{s}_{ml})$$

cumpliéndose la segunda igualdad para $\hat{s} = \hat{s}_{ml}$: <u>es el ML</u>

– Si no existe, no se puede decir que el ML tenga varianza mínima. Así, para eficiencia:
– se comprueba si el ML es ae (alcanza la cota);
– si no, habrá de buscarse un MVUE (Estimador Insesgado de Mínima Varianza)

Trabajo: Determinación de estimadores MVUEs.

Ejercicios

E: Discutir el carácter eficiente del estimador $\hat{m}_{ml}$ para el caso Gauss (media muestral).

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

de donde

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

y, al tratarse de un estimador insesgado, se ha de aplicar la cota de Cramer-Rao

$$Var\{\hat{m} \mid m\} \geq \frac{1}{E\left\{\frac{\partial^2 \ln p(\mathbf{x} \mid m)}{\partial m^2}\right\}} = \frac{v}{K}$$

que coincide con la varianza del ML: luego el ML es eficiente.

A´.2. Caso de parámetro aleatorio

$$\mathrm{Var}\left\{\hat{s}-s\right\} \geq \frac{1}{E\left\{\left[\frac{\partial \ln p(\mathbf{x}, s)}{\partial s}\right]^2\right\}}$$

$$= -\frac{1}{E\left\{\frac{\partial^2 \ln p(\mathbf{x}, s)}{\partial s^2}\right\}}$$

La demostración sigue una vía análoga a la anterior, partiendo de

$$E\left\{\hat{s}-s-b\right\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(\hat{s}-s-b)\,p(\mathbf{x},s)\,d\mathbf{x}\,ds = 0$$

y recordando que $E\{\hat{s} - s - b\}$ es una constante.

La discusión de eficiencia es paralela a la anterior, pero referida al estimador MAP (en lugar de al ML). Además, como $\hat{s}_{ms}(\mathbf{x})$ es insesgado, habrán de coincidir.

B. <u>Decisión</u>

Aparte del coste medio (cuando proceda), ya se sabe que son relevantes

– La probabilidad de **error tipo I** o **falsa alarma**

$$P_I = P_{FA} = \Pr(D_1 \mid H_0) = \int_{X_1} p(\mathbf{x}|H_0)\,d\mathbf{x} = \int_{\eta}^{\infty} p(\Lambda|H_0)\,d\Lambda$$

– La probabilidad de **error tipo II** o **pérdida**

$$P_{II} = P_M = \Pr(D_0 \mid H_1) = \int_{X_0} p(\mathbf{x}|H_1)\,d\mathbf{x} = \int_0^{\eta} p(\Lambda|H_1)\,d\Lambda$$

($P_D = 1 - P_M$: probabilidad de detección).

Se aprecia que existe un compromiso en su reducción: vía partición en X_0, X_1, o establecimiento de η.

No siempre se emplean directamente P_I y P_{II} ó P_{FA} y P_D; p. ej., en diagnóstico se suele tomar H_0 como hipótesis correspondiente a situación normal, y H_1 como hipótesis de defecto o enfermedad, y se emplean:

- Especificidad: $E = \Pr\left(D_0 \mid H_0\right) = 1 - P_{FA}$
- Sensibilidad: $S = \Pr\left(D_1 \mid H_1\right) = P_D$

(subsiste, naturalmente, el compromiso: para su maximización).

Como en estimación, estas características

- se pueden calcular, supuesta conocida la "física" del problema
- se pueden estimar, como frecuencias relativas

y conviene contrastar resultados.

Trabajo: la técnica del "Importance Sampling"

Ejercicio

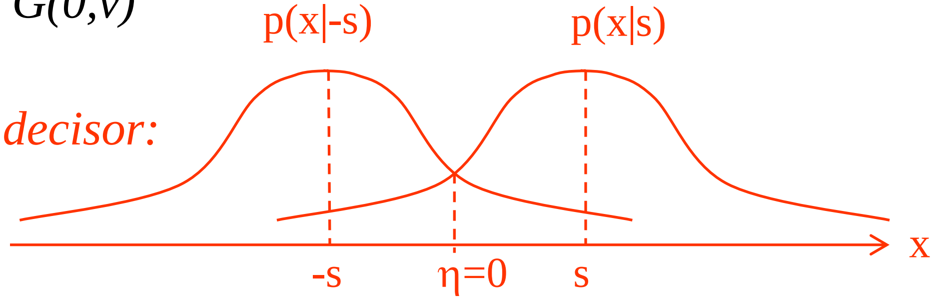
E: Calcúlense las características del decisor ML para

$$H_0 : x = -s + r$$

$$H_1 : x = s + r$$

siendo s una constante y r una va G(0,v)

Ya se ha visto la forma del decisor:



y es obvia la igualdad de ambas probabilidades de error:

$$P_{FA} = P_M = P_e = \frac{1}{\sqrt{2\pi v}} \int_0^{\infty} exp\left[-\frac{(x+s)^2}{2v} \right] dx$$

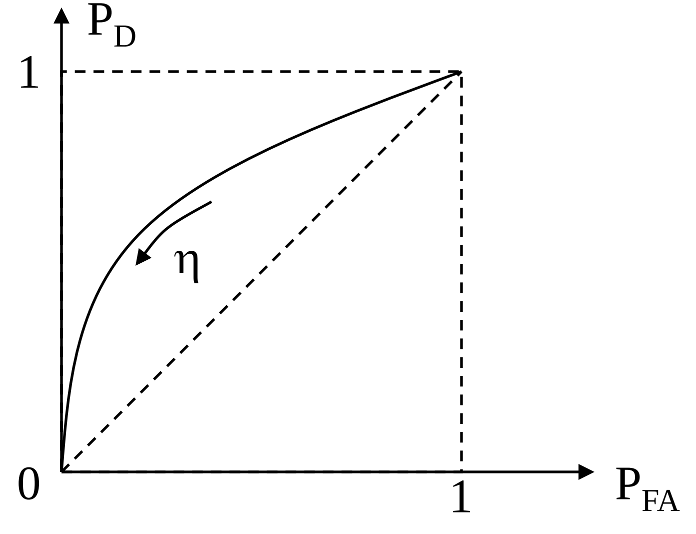
que se reduce inmediatamente a

$$\frac{1}{\sqrt{2\pi v}} \int_s^{\infty} exp\left[-\frac{x'^2}{2v} \right] dx' = \frac{1}{\sqrt{2\pi}} \int_{s/v}^{\infty} exp\left[-\frac{x''^2}{2} \right] dx'' = erf\left(\frac{s}{\sqrt{v}} \right)$$

Nótese la dependencia con la SNR (s^2/v).

Característica de Operación (OC)

Recibe este nombre la curva (familia de curvas en caso de que exista una parametrización, como la anterior SNR) que relaciona dos características en compromiso en una decisión: típicamente P_D vs. P_{FA}; dicha curva se recorre si se permite la variación del umbral η.



La forma de la curva, y que se encuentre por encima de la diagonal principal, se comprende fácilmente.

La calidad de un diseño práctico (con η seleccionable) se aprecia por la cercanía de la curva a los lados izquierdo y superior del cuadrado que la limita.

En comunicaciones y radar, se habla de la ROC ("Receiver OC")

Ejercicio: Demostrar que η es la pendiente de la tangente a la curva en cualquier punto.

Ejercicio de ampliación

A: En **Recuperación de Información** *(IR, "Information Retrieval") se seleccionan ítems de una base (no estructurada) por su relevancia para un cierto uso (o usuario); los ítems pueden ser relevantes (H_1) o no (H_0); las características de calidad utilizadas son*

- *la precisión P: $Pr(H_1|D_1)$*
  (probabilidad de que un ítem seleccionado sea relevante)
- *la recuperación R: $Pr(D_1|H_1)$*
  (probabilidad de recuperar ítems relevantes)

a) Discútase el compromiso entre estas características.

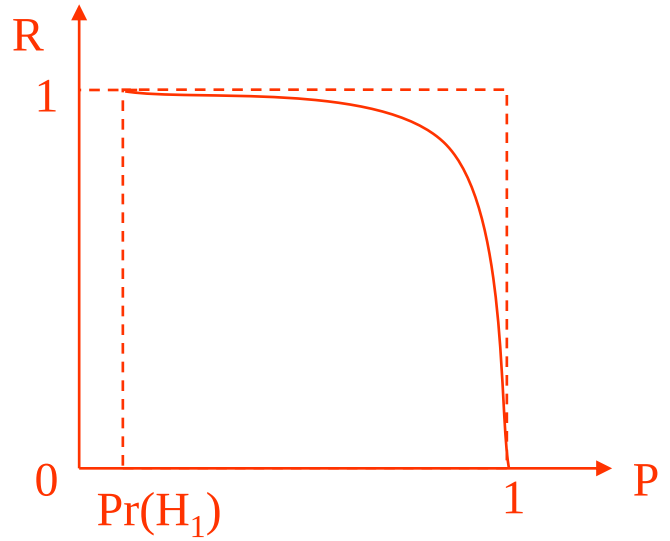
b) Indíquese la forma de la OC R vs. P.

a) $$Pr(H_1 \mid D_1) = \frac{Pr(D_1 \mid H_1)Pr(H_1)}{Pr(D_1 \mid H_1)Pr(H_1) + Pr(D_1 \mid H_0)Pr(H_0)} =$$

$$= \frac{1}{1 + \dfrac{Pr(D_1 \mid H_0)Pr(H_0)}{Pr(D_1 \mid H_1)Pr(H_1)}}$$

para que tienda a 1, ha de bajar $Pr(D_1|H_0)$ (P_{FA}) *y subir* $R = Pr(D_1|H_1)$ (P_D)*: cuyo compromiso ya es conocido.*

b) *La forma de la OC será:*



Apéndice: Cota C-R para parámetro determinista

$$E\{s-\hat{s}-b(s)\,|\,s\}=\int_{-\infty}^{\infty}\left[\hat{s}-s-b(s)\right]p(\mathbf{x}\,|\,s)\,d\mathbf{x}=0$$

$$\frac{\partial E}{\partial s}=\frac{\partial}{\partial s}\int \quad =-\int_{-\infty}^{\infty}\left[1+\frac{d\,b(s)}{d\,s}\right]p(\mathbf{x}\,|\,s)\,d\mathbf{x}\;+\;\int_{-\infty}^{\infty}\left[\hat{s}-s-b(s)\right]\frac{\partial p(\mathbf{x}\,|\,s)}{\partial s}\,d\mathbf{x}=0$$

$$\int_{-\infty}^{\infty}\left[\hat{s}-s-b(s)\right]\frac{\partial p(\mathbf{x}\,|\,s)}{\partial s}\,d\mathbf{x}=1+\frac{d\,b(s)}{d\,s}$$

$$\int_{-\infty}^{\infty}\left[\hat{s}-s-b(s)\right]\frac{\partial \ln p(\mathbf{x}\,|\,s)}{\partial s}\,p(\mathbf{x}\,|\,s)\,d\mathbf{x}=1+\frac{d\,b(s)}{d\,s}$$

Se aplica la desigualdad de Cauchy-Schwartz

$$\left(\int uv\right)^2\le\int u^2\int v^2$$: (= si y sólo si u=kv),

con $u=\frac{\partial \ln p(\mathbf{x}\,|\,s)}{\partial s}\sqrt{p(\mathbf{x}\,|\,s)}$, $v=\sqrt{p(\mathbf{x}\,|\,s)}\left[\hat{s}-s-b(s)\right]$

$$\int_{-\infty}^{\infty}\left[\frac{\partial \ln p(\mathbf{x}\,|\,s)}{\partial s}\right]^2 p(\mathbf{x}\,|\,s)\,d\mathbf{x}\int_{-\infty}^{\infty}\left[\hat{s}-s-b(s)\right]^2 p(\mathbf{x}\,|\,s)\,d\mathbf{x} \geq \left(1+\frac{d\,b(s)}{d\,s}\right)^2$$

de donde

$$\int_{-\infty}^{\infty}\left[\hat{s}-s-b(s)\right]^2 p(\mathbf{x}\,|\,s)\,d\mathbf{x} = \mathrm{Var}\left\{\hat{s}-s\,|\,s\right\} = \mathrm{Var}\left\{\hat{s}\,|\,s\right\} \geq \frac{\left[1+\dfrac{d\,b(s)}{d\,s}\right]^2}{E\left\{\left[\dfrac{\partial p(\mathbf{x}\,|\,s)}{\partial s}\right]^2 \middle|\, s\right\}}$$

Alternativamente,

$$\mathrm{Var}\left\{\hat{s}\,|\,s\right\} \geq -\frac{\left[1+\dfrac{d\,b(s)}{d\,s}\right]^2}{E\left\{\dfrac{\partial^2 \ln p(\mathbf{x}\,|\,s)}{\partial s^2} \middle|\, s\right\}}$$

(Nota: si se alcanza, no implica mínimo error cuadrático medio: es para un b(s) dado).

La forma alternativa nace de

$$\frac{\partial}{\partial s}\int_{-\infty}^{\infty} p(\mathbf{x}\,|\,s)\,d\mathbf{x} = \int_{-\infty}^{\infty} \frac{\partial p(\mathbf{x}\,|\,s)}{\partial s}\,d\mathbf{x} = \int_{-\infty}^{\infty} \frac{\partial \ln p(\mathbf{x}\,|\,s)}{\partial s}\,p(\mathbf{x}\,|\,s)\,d\mathbf{x} = 0$$

y derivando nuevamente la última integral

$$\int_{-\infty}^{\infty} \frac{\partial^2 \ln p(\mathbf{x}\,|\,s)}{\partial s^2}\,p(\mathbf{x}\,|\,s)\,d\mathbf{x} + \int_{-\infty}^{\infty}\left[\frac{\partial \ln p(\mathbf{x}\,|\,s)}{\partial s}\right]^2 p(\mathbf{x}\,|\,s)\,d\mathbf{x} = 0$$

que es

$$E\left\{\frac{\partial^2 \ln p(\mathbf{x}\,|\,s)}{\partial s^2}\,|\,s\right\} + E\left\{\left[\frac{\partial \ln p(\mathbf{x}\,|\,s)}{\partial s}\right]^2 |\,s\right\} = 0$$