TRATAMIENTO DIGITAL DE SEÑALES

Ingeniería de Telecomunicación (4º, 2º c)

Unidad 2ª: Introducción a la decisión analítica

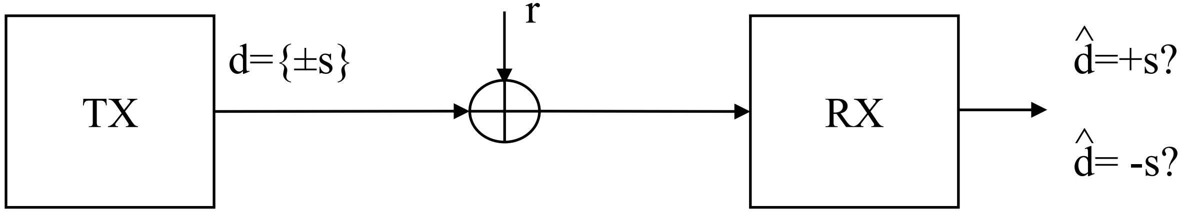
Aníbal R. Figueiras Vidal
Jesús Cid Sueiro
Ángel Navia Vázquez

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

Un problema de comunicaciones

Transmisión digital binaria (antipodal; canal ideal)

Modelo discreto:

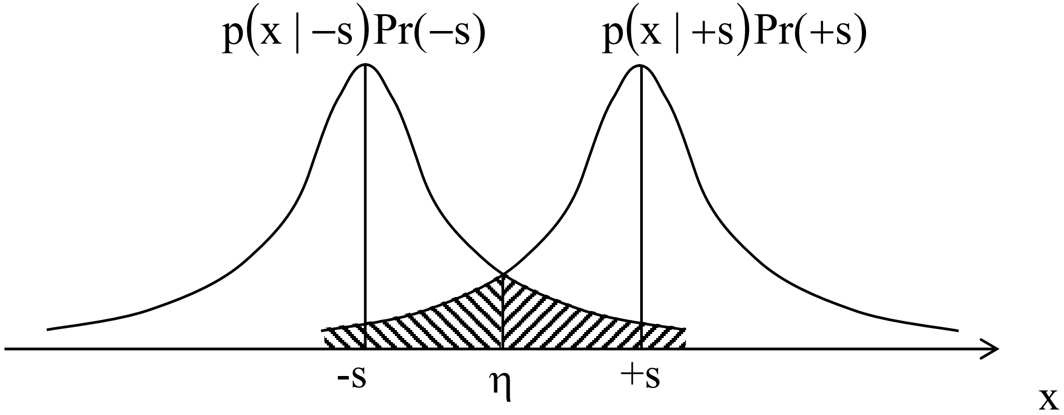


r es un ruido G(0, v) (y blanco)

(se conoce la "<u>física</u>" del problema)

Se observa:

$$x = +s + r, \quad \text{si} \quad +s$$
$$x = -s + r, \quad \text{si} \quad -s$$



y resulta "natural" decidir aplicando un umbral η a la observación x:

$$\hat{d} = \begin{cases} +s & \text{si } x > \eta \\ -s & \text{si } x < \eta \end{cases} \qquad x \underset{-s}{\overset{+s}{\gtrless}} \eta$$

Las probabilidades de error $\Pr(\hat{d} = +s, d = -s)$, $\Pr(\hat{d} = -s, d = +s)$, son las áreas rayadas de la figura: para iguales probabilidades de los símbolos transmitidos, la probabilidad total de error, $\Pr(\hat{d} = +s, d = -s) + \Pr(\hat{d} = -s, d = +s)$, se minimiza eligiendo η=0.

Complementos prácticos

- En radar, por cada celda de observación se tiene:

  $x = s + r$, si hay blanco

  $x = r$, si no

  y el umbral se determina para fijar Pr(s|0) en un cierto valor (P_{FA}, **probabilidad de falsa alarma**).

  (En realidad, se **integran** observaciones provinientes de repetidas iluminaciones del blanco).

- En cualquier caso, ante canal ideal (como ocurre en radar, p. ej.), el modelo digital se deriva de observar durante un intervalo de duración T

  $x(t) = +s(t) + r(t)$, si +s

  $x(t) = -s(t) + r(t)$, si -s (análogamente si 0)

  con r(t) gaussiano y blanco: dada la incoherencia de r(t), un proceso de integración resulta razonable para pasar al modelo discreto.

Ejercicio de Ampliación

A: *La "integración" se puede realizar con un filtro lineal e invariante, muestreando su salida al concluir el intervalo de símbolo (T).*

Así, se obtendrán

$$s = \int_0^T s(T - t')\, h(t')\, dt'$$

$$r = \int_0^T r(T - t')\, h(t')\, dt'$$

y r mantendrá el carácter Gauss.

Resulta obvio, en la discusión anterior, que la probabilidad de error decrece con s^2/v (relación señal/ruido). Al ser el ruido blanco, su varianza depende de

$\int_0^T h^2(t')\, dt'$, *pero no de la forma de h(t).*

Elíjase dicha forma para maximizar la relación señal/ruido.

La energía de la salida debida a la señal es $s^2 = \left[\int_0^T h(t')\, s(T - t')\, dt'\right]^2$

y, aplicando la desigualdad de Cauchy-Schwartz

$$\left(\int uv\right)^2 \le \left(\int u^2\right)\left(\int v^2\right); \qquad = \text{ si y sólo si } u = kv$$

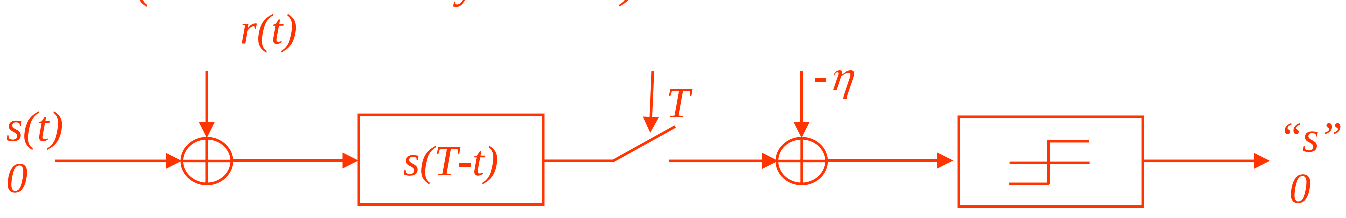
resulta

$$s^2 \leq \int_0^T h^2(t')\,dt' \int_0^T s^2(T - t')\,dt'$$

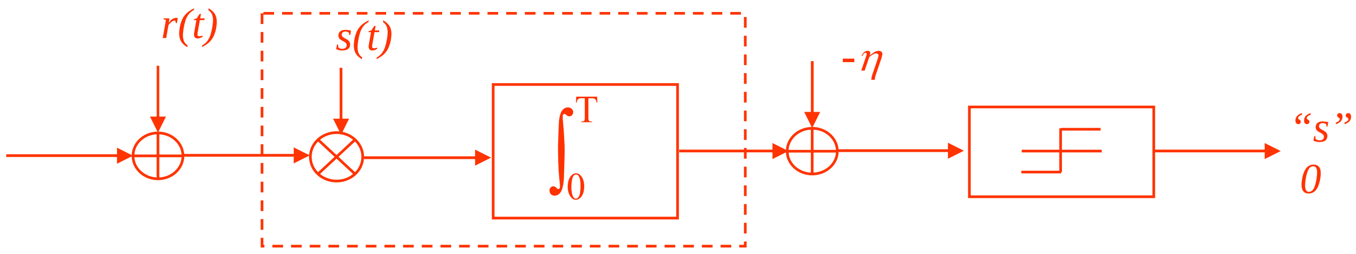
alcanzándose el máximo con

$$h(t) = s(T\text{-}t) \qquad \text{(la constante es irrelevante)}$$

*es decir, con un filtro lineal e invariante que tenga como respuesta al impulso el símbolo revertido (en el tiempo): es el **filtro adaptado**, óptimo (lineal e invariante) para detección (en ruido Gauss y blanco).*



*también realizable como **receptor de correlación***



<u>Discusión</u>

Q: *¿Qué ocurriría si el canal no fuese ideal?*

Aparecería el fenómeno de la ***Interferencia Inter-Símbolos*** *("ISI"), y su presencia en las observaciones haría perder el carácter óptimo al proceso descrito.*

Q: *¿Mediante qué dos procedimientos fundamentales se combate la ISI?*

- *Para reducir su efecto, el conjunto perfil del símbolo transmitido-filtro integrador se diseña de forma que el resultado dé un valor nulo al ser muestreado en tiempos (0, -T, ±2T, ...); como el coseno alzado.*

- *En el diseño anterior no suele incluirse el canal por ser desconocido: su efecto se compensa con un* ***igualador****, sistema (lineal e invariante) que, en cascada con el canal, pretende una respuesta ideal.*

  Nótese que <u>no</u> es un proceso <u>óptimo</u> desde el punto de vista de decisión: no se considera el efecto del ruido (es útil si la ISI <u>domina</u>).

Q: ¿Cómo se procedería al análisis en el caso de transmitir multiples símbolos?

*Para las dos componentes (en fase y en cuadratura) conjuntamente o por separado, bastaría proyectar los símbolos sobre los componentes de una **base** de formas de onda (correspondientes a posibles filtrados) para reducir el problema al modelo discreto.*

Nótese la ventaja del empleo de símbolos antipodales: maximizan la relación señal/ruido; o de la ortogonalidad en el plano complejo.

Q: Indique qué aspectos se tienen en cuenta para diseñar un receptor de comunicaciones.

*La <u>información "física</u>" disponible sobre el problema: características de señal y ruido; adicionalmente se busca sencillez (igualadores lineales) para posiblitar la **adaptatividad** ante variaciones del canal.*

Q: ¿Por qué se minimiza la tasa de error en las comunicaciones digitales?

El objetivo es preservar la <u>calidad</u> de la información recibida: dada la forma habitual de transmitir (digitalización de forma de onda o digitalización no uniforme para modelos, protección posiblemente selectiva y aleatorización de símbolos transmitidos), minimizar la tasa de error es una forma aproximada de conseguirlo.

Nomenclatura

- Los procesos del tipo discutido se dicen de **tratamiento (digital) de señales** ("Digital Signal Processing", DSP) porque interviene el tiempo; especialmente, cuando la solución ha de establecerse en tiempo real.

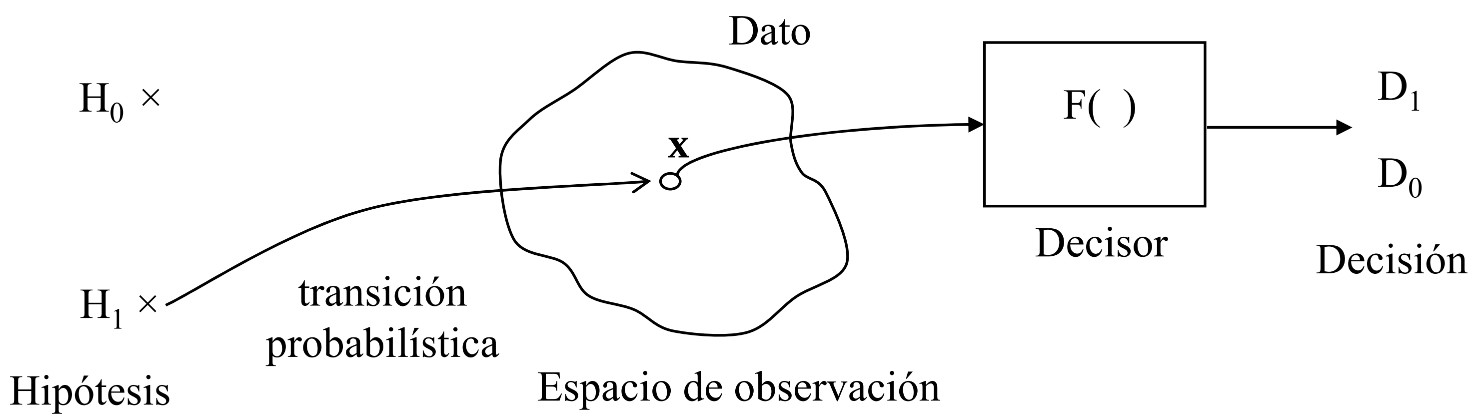
- Recuérdese que las imágenes se manejan vía exploración secuencial.

T: Otros problemas que se plantean en términos que permiten la decisión/clasificación analítica son:

- *la elección de una estrategia en un juego*
- *la detección de la presencia de un componente químico en una muestra*

Discuta los aspectos generales de dichos problemas.

Visión (analítica) de la decisión



Nótese que con el conocimiento estadístico:

* Probabidades "a priori": $\Pr(H_i)$
* Verosimilitudes: $p(\mathbf{x}|H_i)$
* Parámetros de coste: C_{ji} (se toman no negativos, y $C_{ji} > C_{ii}$, $\forall\, j \neq i$)

se diseña el decisor F

Recordatorio de relaciones

$$p(\mathbf{x} \mid H_i)\Pr(H_i) = p(\mathbf{x}, H_i)$$: ddp conjunta de **x** y H_i

$$\sum_i p(\mathbf{x} \mid H_i)\Pr(H_i) = p(\mathbf{x})$$: ddp de **x** (probabilidad total)

$$\Pr(H_i \mid \mathbf{x}) = \frac{p(\mathbf{x} \mid H_i)\Pr(H_i)}{p(\mathbf{x})}$$: probabilidad "a posteriori" de H_i (vista la observación)

$$= \frac{p(\mathbf{x} \mid H_i)\Pr(H_i)}{\sum_j p(\mathbf{x} \mid H_j)\Pr(H_j)}$$: forma de Bayes

$$\int_X \Pr(H_i \mid \mathbf{x})\, p(\mathbf{x})\, d\mathbf{x} = \Pr(H_i)$$

Con la información estadística disponible, se puede establecer que, en cualquier instancia, se da una cierta hipótesis H_i que origina una observación $\mathbf{x}$: decidir D_j llevará aparejado un coste

$$C(D_j, H_i) = C_{ji}$$

que se puede promediar respecto de cada elemento aleatorio ($\mathbf{x}$ o H_i) o respecto a ambos.

Es claro que, si se pretende un diseño para ser aplicado en un número elevado de casos, habrá que fijar como objetivo algún promedio de este coste.

Ejercicios de Ampliación

A: *Exprese analíticamente el coste medio de decidir D_j a la vista de* **x**. *Indique con qué decisor se minimiza.*

$$\overline{C}\left(D_j \mid \boldsymbol{x}\right) = E\left\{C\left(D_j, H\right) \mid \boldsymbol{x}\right\} = \sum_i C_{ji} Pr\left(H_i \mid \boldsymbol{x}\right)$$

$$\boldsymbol{x} \to D_{j^*} \quad si \quad j^* = \arg\left[\min_j\left\{\sum_i C_{ji} Pr\left(H_i \mid \boldsymbol{x}\right)\right\}\right]$$

A: *Exprese analíticamente el coste medio global. Indique con qué decisor se minimiza.*

$$\overline{C} = E\left\{C\left(D(\boldsymbol{x}), H\right)\right\} = \sum_j \sum_i C_{ji} Pr\left(D_j, H_i\right)$$

Dado que cada **x** *se atribuye (determinísticamente) a D_j si está en cierta región X_j,*

$$Pr\left(D_j, H_i\right) = Pr\left(H_i\right) \oint_{X_j} p\left(\boldsymbol{x} \mid H_i\right) d\boldsymbol{x}$$

luego

$$\overline{C} = \sum_j \oint_{X_j} \sum_i C_{ji} Pr\left(H_i\right) p\left(\boldsymbol{x} \mid H_i\right) d\boldsymbol{x}$$

cuya minimización determina $\{X_j\}$; equivale a:

$$\boldsymbol{x} \to D_{j^*} \quad si \quad j^* = arg\left[\min_j\left\{\sum_i C_{ji} Pr\left(H_i\right) p\left(\boldsymbol{x} \mid H_i\right)\right\}\right]$$

A: *¿Cuál de los dos objetivos anteriores, min $\overline{C}(D_j \mid \mathbf{x})$ ó min $\overline{C}$, elegiría usted en una aplicación práctica?*

El primero minimiza el coste medio ante una observación; el segundo, el coste medio global...

Como las observaciones aparecen según ciertas estadísticas, han de estar relacionados.

En efecto:

$$\sum_i C_{ji}\, Pr(H_i \mid \boldsymbol{x}) = \frac{1}{p(\boldsymbol{x})} \sum_i C_{ji}\, p(\boldsymbol{x}, H_i)$$

$$\sum_i C_{ji}\, p(\boldsymbol{x} \mid H_i) Pr(H_i) = \sum_i C_{ji}\, p(\boldsymbol{x}, H_i)$$

que difieren en un factor que no depende de j: luego ambos objetivos dan <u>igual decisión</u>.

*Lo resuelto constituye la base de la **Teoría Bayesiana** de la Decisión.*

A: De lo anterior, derive un test de umbral para el caso binario

$$D_0 : \qquad C_{00}\, Pr(H_0)\, p(\mathbf{x} \mid H_0) + C_{01}\, Pr(H_1)\, p(\mathbf{x} \mid H_1)$$

$$D_1 : \qquad C_{10}\, Pr(H_0)\, p(\mathbf{x} \mid H_0) + C_{11}\, Pr(H_1)\, p(\mathbf{x} \mid H_1)$$

$$C_{00}\, Pr(H_0)\, p(\mathbf{x} \mid H_0) + C_{01}\, Pr(H_1)\, p(\mathbf{x} \mid H_1) \underset{D_0}{\overset{D_1}{\gtrless}} C_{10}\, Pr(H_0)\, p(\mathbf{x} \mid H_0) + C_{11}\, Pr(H_1)\, p(\mathbf{x} \mid H_1)$$

$$(C_{01} - C_{11}) Pr(H_1) p(\mathbf{x} \mid H_1) \underset{D_0}{\overset{D_1}{\gtrless}} (C_{10} - C_{00}) Pr(H_0) p(\mathbf{x} \mid H_0)$$

$$\Lambda(\mathbf{x}) = \frac{p(\mathbf{x} \mid H_1)}{p(\mathbf{x} \mid H_0)} \underset{D_0}{\overset{D_1}{\gtrless}} \frac{(C_{10} - C_{00}) Pr(H_0)}{(C_{01} - C_{11}) Pr(H_1)} \triangleq \eta$$

test de razón de verosimilitudes *("Likelihood Ratio Test", LRT)*

Suele ser cómodo:

$$\ln \Lambda(\mathbf{x}) = \ln p(\mathbf{x} \mid H_1) - \ln p(\mathbf{x} \mid H_0) \underset{D_0}{\overset{D_1}{\gtrless}} \ln \eta$$

Obviamente, también hay un ***test de probabilidades a posteriori****:*

$$\frac{Pr(H_1 \mid \mathbf{x})}{Pr(H_0 \mid \mathbf{x})} \underset{D_0}{\overset{D_1}{\gtrless}} \frac{C_{10} - C_{00}}{C_{01} - C_{11}} = \varphi$$

A: Justifique las denominaciones de los siguientes tests para casos particulares:

- ***máximo a posteriori*** *(MAP):* $C_{00} = C_{11} = 0,\ C_{10} = C_{01} = C$
- ***máxima verosimilitud*** *(ML):* $C_{00} = C_{11} = 0,\ C_{10} = C_{01} = C,\ Pr(H_0) = Pr(H_1)(=1/2)$

  - *MAP: llevando al test de probabilidades a posteriori*

$$\frac{Pr(H_1 \mid \mathbf{x})}{Pr(H_0 \mid \mathbf{x})} \underset{D_0}{\overset{D_1}{\gtrless}} 1; \qquad Pr(H_1 \mid \mathbf{x}) \underset{D_0}{\overset{D_1}{\gtrless}} Pr(H_0 \mid \mathbf{x})$$

  - *ML: llevando al LRT*

$$\frac{p(\mathbf{x} \mid H_1)}{p(\mathbf{x} \mid H_0)} \underset{D_0}{\overset{D_1}{\gtrless}} 1; \qquad p(\mathbf{x} \mid H_1) \underset{D_0}{\overset{D_1}{\gtrless}} p(\mathbf{x} \mid H_0)$$

Análogamente ocurre para hipótesis múltiples: se decide según la máxima $Pr(H_i|\mathbf{x})$ *o la máxima* $p(\mathbf{x}|H_i)$

La estrategia minimax

Si en una situación sólo son conocidos los costes C_{ji} y no las $Pr(H_i)$ ni las verosimilitudes, lo "razonable" es decidir aquella hipótesis que minimice el coste máximo que se puede producir

$$D_{j^*}: \quad j^* = \arg\left[\min_j \left\{ \max_i C_{ji} \right\}\right]$$

(lo que equivale a admitir "uniformidad" en la información estadística no disponible).

La adopción de esta estrategia es frecuente en los juegos: en los que la situación depende de la actuación de los otros jugadores.

También puede extenderse este modo de proceder a casos en que se desconoce sólo parte de la información: como las probabilidades a priori, p.ej.

Características de los decisores

– Probabilidad de **error tipo I** o de **falsa alarma**

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

– Probabilidad de **error tipo II** o de **pérdida**

$$P_{II} = P_M = \int_{X_0} p(\mathbf{x} \mid H_1)\, d\mathbf{x} = \int_{0}^{\eta} p(\Lambda \mid H_1)\, d\Lambda$$

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

Existe un obvio compromiso en la reducción de P_{FA} y P_M.
Pueden estimarse como frecuencias relativas.

El Criterio de Neyman-Pearson

En algunas situaciones, como en Radar, se busca maximizar P_D manteniendo acotada P_{FA}: $$P_{FA} = \alpha' \le \alpha$$

Supuestas conocidas las características estadísticas habituales, basta acudir a los multiplicadores de Lagrange y minimizar

$$F_\lambda = P_M + \lambda\left(P_{FA} - \alpha'\right) = \int_{X_0} p(\mathbf{x} \mid H_1)\, d\mathbf{x} + \lambda\left[\int_{X_1} p(\mathbf{x} \mid H_0)\, d\mathbf{x} - \alpha'\right] =$$
$$= \lambda\left(1 - \alpha'\right) + \int_{X_0}\left[p(\mathbf{x} \mid H_1) - \lambda\, p(\mathbf{x} \mid H_0)\right] d\mathbf{x}$$

que, obviamente, se minimiza si X_0 es la región en que el integrando es negativo; de modo que resulta

$$\Lambda(\mathbf{x}) \underset{D_0}{\overset{D_1}{\gtrless}} \lambda$$

que es un LRT. El valor de λ se obtiene de

$$P_{FA} = \int_\lambda^\infty p(\Lambda \mid H_0)\, d\Lambda = \alpha' \le \alpha$$

Es gran ventaja de este planteamiento el que permita tratar casos en que no es de aplicación directa la Teoría Bayesiana: p.ej., si **x** depende de un parámetro determinista (como una amplitud o una fase, p.ej.) bajo H_1, no existe $p(\mathbf{x} \mid H_1)$, y no se puede acudir directamente a un LRT: pero, como se verá, sí cabe aplicar el Criterio de N-P.