

23.554

LBS 1144639

043
269

BCA

UNIVERSIDAD DE SEVILLA
SECRETARIA CIENCIAS

22-6-76

ENTRADA N.º 561

"ALGUNAS CUESTIONES SOBRE DETECCION Y ESTIMACION
DE SEÑALES EN TEORIA DE LA INFORMACION"

UNIVERSIDAD DE SEVILLA
FACULTAD DE MATEMATICAS
BIBLIOTECA

V-B-EL DIRECTOR
Rafael Infante

MEMORIA que para optar el grado
de Doctor, presenta el Licenciado
do en Ciencias Matemáticas, don
Luis PARRAS GUIJOSA.



Director de la Tesis: Profesor Dr. D. RAFAEL INFANTE MACIAS.

UNIVERSIDAD DE SEVILLA. FACULTAD DE CIENCIAS. Junio de 1976.

Expreso mi profundo agradecimiento al Departamento de Estadística Matemática e Investigación Operativa y muy especialmente al Profesor Dr. D. Rafael Infante Macías, director de este trabajo, por su constante ayuda y estímulo en los años de preparación de esta memoria.

Sevilla, junio de 1976

INTRODUCCION

Los orígenes de la Teoría de la Información se remontan a unas ideas formuladas hace más de cien años. No obstante, desde un punto de vista práctico, uno de los primero que logró dar una medida cuantitativa de la información fué Hartley⁽¹⁾ en 1928, quien centrando su interés en los problemas de transmisión de señales en los sistemas de comunicación, definió la razón de transmisión de información para un sistema de comunicación determinado como el logaritmo decimal del número de mensajes posibles que una fuente puede enviar a través de dicho sistema. De su definición se deduce que suponía que todos los mensajes eran igualmente probables.

Durante la Segunda Guerra Mundial Norbert Wiener fué responsable, en gran parte, del desarrollo de una filosofía general de comunicación y control, llamada cibernética, y formuló el concepto de que tanto las señales deseables como las que no lo sean -ruido-, se podía definir en términos probabilísticos

(1) Hartley: "Transmissions of Information" (Bell System. Tech. - J. Vol. 7, 1928), que estuvo muy influenciado por una ley descubierta simultáneamente por Nyquist en los Estados Unidos y Kupfmuller en Alemania.

como procesos estocásticos. Su trabajo se orientó, principalmente, a problemas militares de control. por ejemplo el del control automático de fuego, subrayando la recuperación (interpolación) y predicción (extrapolación) de señales, después de haber sido perturbadas por ruido. Su trabajo era muy conocido a finales de la Segunda Guerra Mundial, pero no fué publicado hasta 1949.⁽¹⁾ Inspirándose en los conceptos de Wiener y teniendo en cuenta los efectos probabilísticos del mensaje y del ruido Claude E. Shannon realizó unos trabajos, ya clásicos, en 1949 ⁽²⁾, introduciendo los conceptos de entropía y capacidad del canal de un sistema de comunicación, relacionándolos mediante los teoremas de codificación.

No obstante, la formulación de Shannon difiere de la de Wiener tanto en la naturaleza de la señal transmitida, como en el tipo de la señal que se toma en el receptor. En el modelo de Shannon se le asocia una señal a cada posible mensaje que pueda emitir la fuente; esta señal pertenece a un conjunto específico (proceso de codificación). El mensaje codificado se envía a través del canal y, a la salida del mismo, se realiza la operación inversa (proceso de decodificación), es decir, se toma una decisión acerca de la señal que se había emitido.

Los objetivos que persigue son aumentar el tamaño del vocabulario, es decir, considerar el mayor número posible de mensajes, y, al mismo tiempo, hacer que la probabilidad de identificación correcta sea también lo mayor posible. Se observa que todo

(1) N.Wiener: "The extrapolation, interpolation, and smoothing of Stationary Time Series with Engineering applications" Ed. Wiley.

(2) C.E.Shannon: "A Mathematical Theory of Communication" Bell Syst. Tech. J. Journal, Vol. 27, 1949, pp. 379-423, 623-656.

ello depende esencialmente de las propiedades del canal y, un tema fundamental, será el análisis de diferentes tipos de canales. Otro problema básico será la elección de un alfabeto de entrada que se pueda emplear con una probabilidad de error pequeña.

Por otro lado, en el modelo de Wiener se envía una señal aleatoria a través del canal sin emplear el proceso de codificación. Se supone que el canal es fijo y se considera como un mecanismo que añade un ruido aleatorio a la señal que ha sido enviada a través de él. En este caso, el decodificador opera sobre la señal recibida para obtener una decisión acerca de la señal emitida o de alguna propiedad de la misma. Por ejemplo, en el problema de predicción, el decodificador estima el valor de la señal para un tiempo futuro.

El objetivo básico es diseñar un receptor que nos proporcione una decisión óptima bajo un determinado criterio.

En nuestro trabajo hemos seguido en parte por el camino emprendido por Wiener y, muy particularmente, el referente a la detección y extracción de señales en presencia de ruidos aunque sin perder de vista los trabajos de Shannon y sus continuadores.

El área de la detección y extracción presenta una combinación de las técnicas clásicas de la inferencia estadística y los procesos estocásticos en la caracterización de la comunicación, radar, sonar y otros modernos sistemas de proceso de datos.

El problema de la detección de una señal en presencia de ruido es equivalente al que, en terminología estadística, se conoce con el nombre de test de hipótesis. Contrastamos la hipóte-

sis de que solamente se presenta ruido durante el intervalo de observación frente a la (o las), hipótesis de que se presente una señal (o una de varias señales). En cambio, empleamos la palabra extracción para describir un proceso de decisión en que buscamos un estimador de la señal o de uno o más de sus parámetros. Es un problema totalmente equivalente al que en terminología estadística conocemos con el nombre de estimación paramétrica.

Sea S una señal emitida por la fuente, y V el valor observado; $\sigma(S)$ es la distribución a priori de las señales en el espacio Ω , y $F(V/S)$ es la función de probabilidad de los valores V para un S dado.

A partir de los datos observados podemos tomar una decisión referente a la señal emitida, siendo Δ el conjunto de todas las posibles decisiones γ y que llamamos espacio de decisiones.

Definimos una regla de decisión como una ley de probabilidad condicionada $S(\gamma / V)$ definida sobre Δ y tal que nos indica la probabilidad de tomar la decisión γ cuando hemos observado V . Si esta regla de decisión asigna el valor 1 a una decisión determinada, la llamaremos no aleatorizada, pero, si, por el contrario, dos o más decisiones tienen probabilidades positivas, la regla de decisión se dice aleatorizada.

Cuando tomamos la decisión γ y se encuentra presente la señal S , asignamos una función de costo $C(S, \gamma)$.

Para cada dato V observado, el coste esperado de una deci-

si \acute{o} n γ , es

$$C(\gamma / V) = \int_{\Omega} C(S, \gamma) P(S/V) dS$$

donde $P(S/V)$ es la probabilidad a posteriori de que se haya emitido la se \acute{n} al S dado que hemos recibido V .

Para cualquier regla de decisi \acute{o} n $\delta(\gamma / V)$ el riesgo a posteriori para V fijo es

$$R(V, \delta) = \int_{\Delta} \delta(\gamma / V) C(\gamma / V) d\gamma$$

y el riesgo total para la regla de decisi \acute{o} n es

$$R_T(P, \delta) = \int_{\Gamma} R(V, \delta) P(V) dV$$

donde $P(V)$ es la probabilidad de que se presente el resultado V .

Como el coste esperado de una decisi \acute{o} n $C(\gamma / V)$ es un n \acute{u} mero real, tendr \acute{a} una cota inferior $\underline{C}(V)$ para todas las decisiones γ del espacio de decisiones Δ . Por su propia definici \acute{o} n se verifica que $\underline{C}(V) \leq R(V, \delta)$ y, por tanto,

$$R_T(P, \delta) \geq R^* = \int_{\Gamma} P(V) \underline{C}(V) dV$$

siendo R^* el riesgo de Bayes.

La regla de decisi \acute{o} n bayesiana \acute{o} ptima minimiza el riesgo total de decisi \acute{o} n; por tanto, su riesgo total es el riesgo de Bayes.

Hemos resumido los problemas que aparecen en la recepci \acute{o} n de un sistema de comunicaci \acute{o} n consider \acute{a} ndolos como un problema de decisi \acute{o} n. Tanto la detecci \acute{o} n de se \acute{n} ales como la extracci \acute{o} n de las mismas estar \acute{i} an incluidas en el breve res \acute{u} men anterior sin m \acute{a} s -

que considerar las reglas de decisión tales que para las señales recibidas nos conduzcan a determinados conjuntos del espacio de señales o bien a señales (o parámetros de las mismas), individuales respectivamente.

Una vez especificada la solución general de los problemas de detección y estimación de señales pasamos a indicar brevemente el trabajo desarrollado en esta memoria.

En el Capítulo I encontramos una cota inferior para el riesgo en la extracción en el caso general, pasando a continuación a estudiar dos casos particulares de funciones de pérdida, obteniendo para ellos una cota de la que se deduce, como caso particular, la cota de Shannon.

En el Capítulo II proseguimos este estudio, pero dirigiéndolo ahora al caso de la detección de señales, y obtenemos también algunos resultados interesantes, análogos a los del capítulo anterior; viendo, finalmente, que la cota de Fano es un caso particular de la que nosotros hemos obtenido.

Para poder realizar la estimación de señales necesitamos conocer con probabilidad uno el conjunto al que pertenece la señal presente durante el intervalo de observación. Nosotros hemos debilitado esta hipótesis de partida y ello nos ha conducido a estudiar en el Capítulo III el caso de detección y estimación conjunta para hipótesis múltiples, obteniendo al particularizar al caso binario resultados análogos a los dados por Middleton y Espósito.

En el Capítulo IV hemos estudiado el mismo proceso pero pa

ra funciones de pérdida especiales llegando a situaciones más -
concretas que anteriormente, y al particularizar tanto para el
caso binario, como para detección y estimación puras, llegamos
a los mismos resultados obtenidos ya anteriormente por otros -
autores al enfrentarse con estos problemas directamente.

CAPITULO I

COTA INFERIOR PARA EL RIESGO EN LA EXTRACCION

1.1. Cota para el riesgo a posteriori en la extracción

- Teorema

1.2. Función de pérdida del valor absoluto

- Teorema

1.2.1. Cota para el riesgo total

1.2.2. Ejemplos:

Caso 1: Ley de probabilidad a priori uniforme

Caso 2: Ley de probabilidad a priori $N(\mu, \sigma_s^2)$

1.2.3. Riesgo total normalizado

- Definición

- Teorema

1.3. Función de pérdida cuadrática

- Teorema

1.3.1. Cota inferior para el riesgo total

1.3.2. Ejemplos:

Caso 1: Ley de probabilidad a priori uniforme

Caso 2: Ley de probabilidad a priori $N(\mu, \sigma_s^2)$

1.3.3. Riesgo total relativo

- Definición

1.4. Relación con la Cota de Shannon

Bibliografía

COTA INFERIOR PARA EL RIESGO EN LA EXTRACCION

A veces es bastante más complicado calcular el rendimiento o calidad de un sistema de comunicación óptimo en un determinado sentido, que hallar una cota de dicha calidad sin necesidad de resolver el problema de optimización. Nosotros hemos encontrado una cota general que es válida para una gran cantidad de sistemas, viendo, posteriormente, su relación con la conocida cota de Shannon basada en la razón de distorsión para un determinado criterio de fiabilidad.

Señalemos, finalmente, que todo nuestro trabajo se apoya, fundamentalmente, en el teorema de convexidad de Kullback, así como en la expresión dada por Middleton de los problemas de extracción como sinónimos a los de estimación de la Teoría de la Decisión.

1.1. COTA PARA EL RIESGO A POSTERIORI EN LA EXTRACCION.

Siguiendo los trabajos de Middleton y Van Meter [7] (11) llamemos al espacio de señales emitidas, sobre el que hay de-

finida una ley de probabilidad a priori $P_S(s)$. Sea el espacio de señales recibidas, también llamado espacio de observaciones sobre el que también suponemos hay definida otra ley de probabilidad $P_R(r)$. Consideremos, asimismo, la ley de probabilidad $P_{S/r}(s/r)$ definida sobre \mathcal{R} .

A partir de los $n = 2WT$ puntos muestrales obtenemos una estimación de la señal enviada durante el intervalo $(0, T)$. Sea, $\hat{s}(r)$ un estimador de dicha señal (extractor, Middleton y Van Meter [7]), para cada r , la decisión $\hat{s}(r)$ es un elemento del espacio.

Sea ahora $C(s, \hat{s}(r))$ la función de pérdida asociada al par de variables $(s, \hat{s}(r))$, ambos elementos del espacio \mathcal{S} definimos el riesgo a posteriori, [4], como

$$R(r, s) = \int_{\mathcal{S}} C(s, \hat{s}(r)) p(s/r) d\tau \quad (1.1)$$

donde τ es una medida de volumen definida sobre \mathcal{S} y $p(s/r)$ es la derivada de Radon-Nicodym [8] de la función $P_{S/r}(s/r)$

El riesgo total [4], $R_T(P_R, s)$, será

$$R_T(P_R, s) = \int_{\mathcal{R}} R(r, s) dP_R \quad (1.2)$$

Enunciamos a continuación el teorema de convexidad de Kullback, alrededor del cual girará nuestro trabajo en este capítulo.

TEOREMA 1.1. Supongamos las medidas μ_1 y μ_2 , definidas sobre el espacio X y tales que μ_1 y μ_2 sean absolutamente contínuas con respecto a la medida λ . Notemos

$$f_1(x) = \frac{d\mu_1}{d\lambda} \quad \text{y} \quad f_2(x) = \frac{d\mu_2}{d\lambda}$$

las derivadas de Radon-Nicodým y definamos

$$D \left[f_1, f_2 \right] = \int_x - \ln \frac{f_2(x)}{f_1(x)} d\mu_1$$

En estas condiciones $D \left[f_1, f_2 \right] \geq 0$, con igualdad si sólo si $\mu_1 = \mu_1^2$

(Kullback, 1968)

Definamos ahora la función de densidad

$$p(s, \hat{s}, \alpha) = k(\alpha) \exp \left\{ -\alpha C(s, \hat{s}) \right\} \quad (1.3)$$

siendo α una constante y $k(\alpha)$ el factor de normalización, que ha de ser

$$k(\alpha) = \left[\int_{\Omega} \exp \left\{ -\alpha C(s, s) \right\} d\sigma \right]^{-1} \quad (1.4)$$

siempre que se verifique

$$\int_{\Omega} \exp \left\{ -\alpha C(s, s) \right\} d\sigma < +\infty \quad (1.5)$$

Definamos ahora

$$\xi(r, \alpha) = D \left[p(s/r), p(s, \hat{s}, \alpha) \right] \quad (1.6)$$

que, teniendo en cuenta el teorema 1.1. de convexidad se puede escribir como

$$\begin{aligned} \xi(r, \alpha) &= \int_{\Omega} - \ln \frac{p(s, \hat{s}, \alpha)}{p(s/r)} p(s/r) d\sigma = \\ &= - \int_{\Omega} \ln p(s, \hat{s}, \alpha) p(s/r) d\sigma + \int_{\Omega} \ln p(r/s) p(r/s) d\sigma = \\ &= - \int_{\Omega} \ln k(\alpha) p(s/r) d\sigma + \alpha \int_{\Omega} C(s, \hat{s}) p(s/r) d\sigma + \end{aligned}$$

$$\begin{aligned}
& + \int_{\Omega} \ln p(s/r) p(s/r) d\sigma = \\
& = - \ln k(\alpha) + \alpha R(r, s) - H(S/r) \quad (1.7)
\end{aligned}$$

de donde podemos obtener el riesgo a posteriori como

$$R(r, s) = \frac{1}{\alpha} [H(S/r) + \ln k(\alpha) + \epsilon(r, \alpha)] \quad (1.8)$$

siendo

$$H(S/r) = - \int_{\Omega} \ln p(s/r) p(s/r) d\sigma \quad (1.9)$$

la entropía condicional [9] [5] de la variable aleatoria S , definida sobre el espacio de señales Ω , al valor fijo $r \in \Gamma$.

Del teorema de convexidad se sigue que ha de ser $\epsilon(r, \alpha)$ no negativa y, por tanto, podemos afirmar que

$$R(r, s) \geq T[H(S/r), \alpha] \quad (1.10)$$

donde con $T[H(S/r), \alpha]$ hemos expresado

$$T[H(S/r), \alpha] = \frac{1}{\alpha} [H(S/r) + \ln k(\alpha)] \quad (1.11)$$

Como vemos, la cota obtenida depende del parámetro α , que ha de satisfacer la condición (1.5), para que no obstante, está indeterminado. Para obtener la cota óptima tendremos que maximizar la función $T[H, \alpha]$ con respecto a α , para lo cual sería ideal poseer $k(\alpha)$ en forma explícita.

1.2. FUNCION DE PERDIDA DEL VALOR ABSOLUTO

Supongamos ahora que Ω es un espacio unidimensional y sea

Sea la variable aleatoria definida sobre Ω , de la que conocemos su ley de probabilidad a priori. Así mismo, supongamos que $d^{\sigma} = -ds$, y que la función de pérdida es de la forma

$$C(s, \hat{s}) = |s - \hat{s}| \quad (1.12)$$

TEOREMA 2.1. En estas condiciones y si α_0 es el valor de α que maximiza $T[H, \alpha]$, tenemos

$$T[H(S/r), \alpha_0] = \frac{1}{2\alpha_0} \exp\{H(S/r)\}$$

En efecto, por ser $p(s, \hat{s}, \alpha)$ una función de densidad y ser $C(s, \hat{s}) = |s - \hat{s}|$, ha de ser⁽¹⁾

$$k(\alpha) = \frac{\alpha}{2} \quad (1.13)$$

Si tenemos en cuenta ahora (1.11) y (1.13), podemos escribir

$$\alpha T[H, \alpha] = H + \ln \alpha - \ln 2 \quad (1.14)$$

Pero si α_0 es el valor de α que maximiza $T[H, \alpha]$, para cada H fijo, deberá ser un cero de su primera derivada parcial con respecto a α ; luego derivando (1.14) con respecto a α , podemos escribir

$$T[H, \alpha] + \alpha \frac{\partial T[H, \alpha]}{\partial \alpha} = \frac{1}{\alpha} \quad (1.15)$$

que para $\alpha = \alpha_0$, quedaría

$$T[H, \alpha_0] = \frac{1}{\alpha_0} \quad (1.16)$$

(1) Bastaría expresar (2.3) para (1.12) y ver que sería la ley de Laplace (

que, por otro lado, tendría que ser igual a (1.11) para $\alpha = \alpha_0$; luego

$$\frac{1}{\alpha_0} = H(S/r) + \ln \frac{\alpha_0}{2} \quad (1.17)$$

de donde se sigue fácilmente que

$$\frac{\alpha_0}{2} = e \exp \{ H(S/r) \} \quad (1.18)$$

y de (1.18) obtenemos que

$$\frac{1}{\alpha_0} = \frac{1}{2e} \exp \{ H(S/r) \} \quad (1.19)$$

Si tenemos ahora en cuenta el valor que toma $T[H, \alpha]$ para $\alpha = \alpha_0$, que viene expresado por (1.16), comparando con (1.19) llegamos a que

$$T[H, \alpha_0] = \frac{1}{2e} \exp \{ H(S/r) \} \quad (1.20)$$

con lo que el teorema quedaría demostrado, siempre y cuando el valor $\alpha = \alpha_0$ haga menor que cero la segunda derivada de la función $T[H, \alpha]$. Para demostrar que ésto es cierto, derivaremos nuevamente con respecto a α en la expresión (1.15), obteniendo

$$2 \frac{\partial T}{\partial \alpha} + \alpha \frac{\partial^2 T}{\partial \alpha^2} = - \frac{1}{\alpha^2} \quad (1.21)$$

que para $\alpha = \alpha_0$ se convierte en

$$\left. \frac{\partial^2 T}{\partial \alpha^2} \right|_{\alpha = \alpha_0} = - \frac{1}{\alpha_0^2} \quad (1.22)$$

con lo que queda completada la demostración del teorema.

Este problema se podría enfocar de otra forma, llegando a

una cota del todo análoga a (1.20), como veremos a continuación.

De (1.8), y teniendo en cuenta el teorema de convexidad, se sigue que una condición necesaria para que el riesgo alcance la cota (1.11) es que sea $\xi(r, \alpha) = 0$, lo cual únicamente es posible para

$$p(s, \hat{s}, \alpha) = p(s/r) \quad (1.23)$$

salvo para conjuntos de probabilidad nula. Por lo tanto (1.23) se verificará si $p(s/r)$ tiene la forma

$$p_0(s/r) = k(\alpha) \exp\{-\alpha C(s, \hat{s})\} \quad (1.24)$$

en cuyo caso tendremos

$$\int_{\Omega} C(s, \hat{s}) p_0(s/r) ds = \int_{\Omega} C(s, \hat{s}) p(s, \hat{s}, \alpha') ds \quad (1.25)$$

y si $\alpha = \alpha_1$ es solución de la ecuación

$$\int_{\Omega} C(s, \hat{s}) p(s/r) ds = \int_{\Omega} C(s, \hat{s}) p(s, \hat{s}, \alpha) ds \quad (1.26)$$

existirá una función de densidad $p_0(s/r)$ para la que sea nulo $\xi(r, \alpha_1)$. Si tenemos ahora en cuenta (1.12), podremos resolver la ecuación (1.26) para α .

Por ser $p(s, \hat{s}, \alpha)$ función de densidad, ha de ser

$$\int_{\Omega} k(\alpha) \exp\{-\alpha C(s, \hat{s})\} ds = 1 \quad (1.27)$$

Derivando ahora ambos miembros con respecto a α y teniendo en cuenta (1.13), llegamos a

$$\frac{1}{\alpha} \int_{\Omega} \exp\{-\alpha |s - \hat{s}|\} ds - \int_{\Omega} \frac{\alpha}{2} |s - \hat{s}| \exp\{-\alpha |s - \hat{s}|\} ds = 0$$

es decir

$$\frac{\alpha_1}{2} \int_{\hat{s}}^{\infty} |s - \hat{s}| \exp \left\{ -\alpha_1 |s - \hat{s}| \right\} ds = \frac{1}{\alpha_1} \quad (1.28)$$

y teniendo ahora en cuenta (1.1) para (1.12), podemos escribir

$$R(r, \hat{s}) = \frac{1}{\alpha_1} \quad (1.29)$$

Pero si tenemos en cuenta (1.8) e igualamos con (1.29), resulta que

$$1 = H(S/r) + \ln k(\alpha_1) + \xi(r, \alpha_1) \quad (1.30)$$

y operando de forma análoga a como lo hemos hecho en el teorema, sustituyendo al mismo tiempo $k(\alpha_1)$ por su valor expresado en (1.13), concluimos que

$$R(r, \hat{s}) = \frac{1}{2e} \exp \left\{ H(S/r) + \xi(r, \alpha_1) \right\} \quad (1.31)$$

y como $\xi(r, \alpha) > 0$, y para $\alpha = \alpha_1$ vale cero. al eliminarlo de (1.31) obtenemos una cota análoga a (1.20) del teorema 2.1. aunque debemos señalar que el valor α_0 que maximiza $T[H, \alpha]$ para H fijo no es, en general. igual al valor α_1 que es solución de la ecuación (1.26).

De (1.31) podemos concluir que, de todas las cotas inferiores del riesgo condicionado. la mejor será aquella que más se aproxime a dicho riesgo a posteriori. Dicha cota estará tanto más próxima a (1.1), para la función de pérdida (1.12), cuanto más se aproxime la ley de probabilidad $p(s/r)$ a una Ley de Laplace, pues tanto más pequeño sería $\xi(r, \alpha)$.

1.2.1. Cota para el riesgo total.

Teniendo en cuenta (1.2) y (1.31), podemos escribir el riesgo total para $\alpha = \alpha_1$, como

$$R_T(P_R, \hat{s}) = \int_{\Gamma} \frac{1}{2e} \exp \{ H(S/r) + \varepsilon(r, \alpha) \} dP_R \quad (1.32)$$

y si tenemos en cuenta la desigualdad de Jensen [2], queda

$$R_T(P_R, s) \geq \frac{1}{2e} \exp \{ H(S/R) + \varepsilon_m \} \quad (1.33)$$

siendo ε_m la media de $\varepsilon(r, \alpha)$ y $H(S/R)$ la media de $H(S/r)$, ambas con respecto a $r \in \Gamma$.

Pero como, por otro lado, la entropía a priori es [1]

$$H(S) = - \int_{\Omega} \ln p(s) p(s) ds \quad (1.34)$$

y la cantidad de información transmitida a través del canal [1]

$$I(S||R) = H(S) - H(S/R) \quad (1.35)$$

se puede escribir (1.33) como sigue

$$R_T(P_R, \hat{s}) \geq \frac{1}{2e} \exp \{ H(S) \} \exp \{ - [I(S||R) - \varepsilon_m] \} \quad (1.36)$$

De todo lo expuesto deducimos que es posible interpretar, tanto en (1.31) como en (1.36), que $\varepsilon(r, \alpha)$ y ε_m son, respectivamente, las pérdidas de información debidas a la extracción.

Podemos eliminar ε_m en (1.36), ya que es mayor ó igual - que cero, obteniendo así como cota para el riesgo total

$$R_T(P_R, \hat{s}) \geq F_S \cdot R_I \quad (1.37)$$

Donde con F_S hemos querido expresar

$$F_S = \frac{1}{2^e} \exp \{ H(S) \} \quad (1.38)$$

y con R_I el resto, es decir

$$R_I = \exp \{ - I(S \parallel R) \} \quad (1.39)$$

Vemos que F_S depende únicamente de la distribución a priori de las señales y que R_I depende de la información transmitida a través del canal.

A continuación consideraremos dos ejemplos con leyes de probabilidad a priori muy frecuentes en Teoría de la Información.

1.2.2. Ejemplos.

A continuación damos dos ejemplos para la función de pérdida (1.12), considerando como leyes de probabilidad a priori la u uniforme y normal respectivamente, llegando a la necesidad de introducir la noción de riesgo total normalizado.

Casoll. - Ley de probabilidad a priori uniforme.

Suponemos que la ley con que las señales se emiten y envían a través del canal es la uniforme en el intervalo $(\mu - \frac{w}{2}, \mu + \frac{w}{2})$, es decir

$$p(s) = \begin{cases} \frac{1}{w}, & \mu - \frac{w}{2} < s < \mu + \frac{w}{2} \\ 0, & s < \mu - \frac{w}{2}, s > \mu + \frac{w}{2} \end{cases} \quad (1.40)$$

de la que sabemos que [12]

$$E[S] = \mu; \quad \sigma_S^2 = \frac{w^2}{12} \quad (1.41)$$

donde con σ_S^2 expresamos la varianza de la ley a priori P_S .

Además, por la definición de la entropía a priori,

$$H(S) = - \int_{\mu - \frac{w}{2}}^{\mu + \frac{w}{2}} \frac{1}{w} \ln(1/w) ds = \frac{\ln w}{w} \left[s \right]_{\mu - \frac{w}{2}}^{\mu + \frac{w}{2}} = \ln w \quad (1.42)$$

por lo que, teniendo en cuenta (1.38), podemos escribir

$$F'_S = \frac{1}{2e} \cdot e^{-\ln w} = \frac{w}{2e} \quad (1.43)$$

y si tenemos en cuenta (1.41), podemos expresar F'_S como función de la desviación típica de la siguiente forma

$$F'_S = \frac{\sqrt{3}}{e} \sigma_S \quad (1.44)$$

Caso 2. - Ley de probabilidad a priori $N(\mu, \sigma_S^2)$.

En este caso también vamos a tratar de expresar F'_S en función de la desviación típica. Para ello vamos a calcular la entropía a priori

$$\begin{aligned} H(S) &= \int_{-\infty}^{+\infty} - \ln \left[\frac{1}{\sigma_S \sqrt{2\pi}} \exp \left\{ - \frac{(s - \mu)^2}{2\sigma_S^2} \right\} \right] \frac{1}{\sigma_S \sqrt{2\pi}} : \\ &\quad \cdot \exp \left\{ - \frac{(s - \mu)^2}{2\sigma_S^2} \right\} ds = \\ &= - \ln(\sigma_S \sqrt{2\pi}) + \frac{1}{2\sigma_S^2} \int_{-\infty}^{+\infty} \frac{(s - \mu)^2}{\sigma_S \sqrt{2\pi}} \exp \left\{ - \frac{(s - \mu)^2}{2\sigma_S^2} \right\} ds = \\ &= - \ln(\sigma_S \sqrt{2\pi}) + \frac{1}{2} \quad (1.45) \end{aligned}$$

por lo que podemos escribir (1.38) como sigue

$$F'_S = \frac{1}{2e} \exp \left\{ \ln(\sigma_S \sqrt{2\pi}) + \frac{1}{2} \right\} = \frac{1}{2e} \sqrt{e} \cdot \sigma_S \sqrt{2\pi} =$$

$$-\sqrt{\frac{\pi}{2e}} \sigma_S \quad (1.45)$$

1.2.3. Riesgo total normalizado.

Los dos casos considerados anteriormente nos sugieren, para el caso de una función de pérdida del tipo (1.12), la definición de un nuevo concepto, que nosotros hemos denominado riesgo total normalizado, de forma que la cota para él sea totalmente universal.

DEFINICION 2.1. Definimos riesgo medio normalizado para la función de pérdida del valor absoluto, y lo notamos por R_N , a

$$R_N = \frac{R_T(P_R, \hat{s})}{\sigma_S} \quad (1.46)$$

donde σ_S es la desviación típica de la ley de probabilidad a priori.

Teniendo ahora en cuenta la definición 2.1. y (1.37), podemos escribir (1.46) como sigue

$$R_N \geq g_1 \cdot R_1 \quad (1.47)$$

donde con g_1 hemos expresado

$$g_1 = \frac{F'_S}{\sigma_S}$$

que es un parámetro totalmente independiente de la ley de probabilidad a priori que se considere.

Por otro lado, de la definición de R_1 , dada en (1.39), se deduce que cada cota superior de $I(S||R)$ nos proporciona una cota inferior de R_1 ; por lo tanto, si tenemos en cuenta la defi

nición de capacidad de un canal

$$C = \max_{\mathbb{S}} I(\mathbb{S} \| R) \quad (1.49)$$

Y es, evidentemente, C una cota superior de $I(\mathbb{S} \| R)$, nos proporcionará, por tanto, una cota inferior para R_1 , o lo que es lo mismo, para R_N , por lo que podremos escribir (1.48) en la forma

$$R_N \geq \sum_i c_i e^{-C} \quad (1.50)$$

donde todos los coeficientes que aparecen en la cota son parámetros constantes, independientes de las leyes de probabilidad que rigen el sistema de comunicación.

1.3. FUNCION DE PERDIDA CUADRATICA

Supongamos que el espacio Ω es un espacio euclídeo de dimensión M , con elementos $(s_1, s_2, \dots, s_M)^T = \bar{s}$, donde todos los s_i , $1 \leq i \leq M$, son números reales; es decir, $\Omega = \mathbb{R}^M$. Supongamos, así mismo, que $d\sigma = ds_1 ds_2 ds_3 \dots ds_M$ y que la función de pérdida $Q(\bar{s}, s)$ es de la forma

$$Q(\bar{s}, s) = (\bar{s} - s)^T D (\bar{s} - s) = \sum_{i=1}^M \sum_{j=1}^M d_{ij} (s_i - \bar{s}_i)(s_j - \bar{s}_j) \quad (1.51)$$

donde d_{ij} son los elementos de la matriz definida positiva D , para $1 \leq i, j \leq M$.

Para que se satisfaga (1.5) para todo $\alpha > 0$, es para lo que es preciso que sea D definida positiva [3].

TEOREMA 3.1. En las condiciones anteriores y si α_0 es el valor de α que maximiza $T[H, \alpha]$, tenemos

$$T[H(S/r), \alpha] = \frac{M}{2\alpha} |D| \exp\left\{\frac{2}{M} H(S/r)\right\}$$

donde con $|D|$ hemos expresado el determinante de la matriz D .

En efecto; por ser $p(s, s, \alpha)$, definida en (1.3), una función de densidad y ser $C(s, s)$ una forma cuadrática definida positiva, ha de ser

$$k(\alpha) = \frac{|D|^{1/2} \alpha^{M/2}}{\pi^{M/2}} \quad (1.52)$$

y si tenemos en cuenta (1.11) y la expresión anterior (1.52), podemos escribir

$$\begin{aligned} \alpha T[H, \alpha] &= H + \ln k(\alpha) = \\ &= \frac{1}{2} \ln |D| + \frac{M}{2} \ln \alpha - \frac{M}{2} \ln \pi \end{aligned} \quad (1.53)$$

Ahora bien, si α_0 es el valor de α que hace máxima la función $T[H, \alpha]$ para cada H fijo, ha de ser

$$\left. \frac{\partial T[H, \alpha]}{\partial \alpha} \right|_{\alpha = \alpha_0} = 0 \quad (1.54)$$

por lo que si derivamos en (1.53) con respecto a α obtenemos

$$T[H, \alpha] + \alpha \frac{\partial T[H, \alpha]}{\partial \alpha} = \frac{M}{2\alpha} \quad (1.55)$$

y sustituimos en (1.55) el valor α_0 , teniendo en cuenta (1.54), obtenemos

$$T[H, \alpha_0] = \frac{M}{2\alpha_0} \quad (1.56)$$

Recordando ahora la expresión (1.11) para $\alpha = \alpha_0$ e igualamos los segundos términos de dicha expresión y de (1.56)

$$\frac{M}{2} = H + \ln k(\alpha_0) \quad (1.57)$$

de donde se sigue que

$$\ln k(\alpha_0) = \frac{M}{2} - H \quad (1.58)$$

y si tenemos en cuenta (1.52)

$$\frac{M}{2} - H = \frac{1}{2} \ln |D| + \frac{M}{2} \ln \alpha_0 - \frac{M}{2} \ln \pi \quad (1.59)$$

de donde obtenemos que

$$\ln \left(\frac{1}{\alpha_0} \right)^{M/2} = \ln \left\{ \frac{|D|^{1/2}}{\pi^{M/2}} \cdot e^H \cdot e^{-M/2} \right\} \quad (1.60)$$

y de aquí deducimos

$$\left(\frac{1}{\alpha_0} \right)^{M/2} = \frac{|D|^{1/2}}{\pi^{M/2}} \cdot e^H \cdot e^{-M/2} \quad (1.61)$$

que se puede escribir también como

$$\frac{1}{\alpha_0} = \frac{|D|^{1/M}}{\pi} \cdot \frac{2H(S/r)}{M} \quad (1.62)$$

Si tenemos ahora en cuenta (1.56), podemos escribir

$$T(H, \alpha_0) = \frac{M |D|^{1/M}}{2e^\pi} \exp \left\{ \frac{2}{M} H(S/r) \right\} \quad (1.63)$$

con lo que el teorema estaría demostrado, a salvo, claro está, de demostrar que el valor $\alpha = \alpha_0$ hace negativa la segunda derivada parcial de la función $T[H, \alpha]$ con respecto a α , para cada H fijo. Para ello derivaremos nuevamente con respecto a α (1.55), obteniendo así

$$2 \frac{\partial T[H, \alpha]}{\partial \alpha} + \alpha \frac{\partial^2 T[H, \alpha]}{\partial \alpha^2} + - \frac{M}{2\alpha} \quad (1.64)$$

que para $\alpha = \alpha_0$, y teniendo en cuenta (1.54),

$$\frac{\partial^2 T[H, \alpha]}{\partial \alpha^2} \Big|_{\alpha = \alpha_0} = - \frac{M}{2\alpha^3} < 0 \quad (1.65)$$

ya que tanto α como M son positivos, De esta forma queda completa la demostración del teorema.

Al igual que para el caso de la función de pérdida del valor absoluto, podemos llegar a una cota del todo análoga a (1.63) por otro camino. Operando de forma análoga a dicho caso, pretendemos obtener una solución α_1 de la ecuación (1.26); para lo cual, por ser $p(s, \hat{s}, \alpha)$ función de densidad se verificaría (1.27), para la función de pérdida (1.51). Derivando ambos miembros con respecto a α , llegamos a

$$0 = \int_{\Omega} \frac{|D|^{1/2}}{\pi^{M/2}} \frac{M}{2} \alpha^{M/2 - 1} \exp\{-\alpha C(s, \hat{s})\} d\sigma - \\ - \int_{\Omega} \frac{|D|^{1/2}}{\pi^{M/2}} \alpha^{M/2} C(s, \hat{s}) \exp\{-\alpha C(s, \hat{s})\} d\sigma \quad (1.66)$$

de donde obtenemos, teniendo en cuenta (1.1) para la función de pérdida cuadrática, que

$$R(r, \hat{s}) = \frac{M}{2\alpha_1} \quad (1.67)$$

ysi, al igual que antes, tenemos en cuenta (1.8) y la igualamos con (1.67), resulta que

$$\frac{M}{2\alpha_1} = \frac{1}{\alpha_1} [H(S/r) + \ln k(\alpha_1) + \varepsilon(r, \alpha)] \quad (1.68)$$

y operando de forma análoga a como lo hemos hecho en el teorema 311, obtenemos

$$R(r, \hat{s}) = \frac{M}{2e^{\pi}} |D|^{1/M} \exp \left\{ \frac{2}{M} H(S/r) + \varepsilon(r, \alpha) \right\} \quad (1.69)$$

y como, según el teorema 1.1., es $\varepsilon(r, \alpha) \geq 0$, y para $\alpha = \alpha_1$ vale $\varepsilon(r, \alpha)$, al eliminarlo de (1.69) obtenemos una cota que es análoga a (1.63), pero hemos de señalar, como en el caso anterior, que aunque las cotas sean iguales, el α_0 que maximiza $T[H, \alpha]$ para cada H fijo no es, en general, igual al valor α_1 que es solución de (1.26) para la función de pérdida cuadrática.

Como para el caso de la función de pérdida del valor absoluto, de (1.69) concluimos que de todas las cotas inferiores del riesgo a posteriori, la mejor estará tanto más próxima al mismo cuanto más se aproxime la función de densidad $p(s/r)$ a una ley normal en el sentido en que $\varepsilon(r, \alpha)$ se hace pequeño.

1.3.1. Cota inferior para el riesgo total

Podemos escribir el riesgo total $R_T(P_R, \hat{s})$, como sigue

$$R_T(P_R, \hat{s}) = \int_r \frac{M}{2e^{\pi}} |D|^{1/M} \exp \left\{ \frac{2}{M} H(S/r) + \varepsilon(r, \alpha_1) \right\} dP_R \quad (1.70)$$

sin más que tener en cuenta (1.2) y (1.69), para $\alpha = \alpha_1$

Si tenemos en cuenta la desigualdad de Jensen⁽¹⁾, queda

$$R_T(P_R, \hat{s}) \geq \frac{M}{2e^{\pi}} |D|^{1/M} \exp \left\{ \frac{2}{M} H(S/R) + \varepsilon_m \right\} \quad (1.71)$$

donde, al igual que antes, ε_m es la media de $\varepsilon(r, \alpha)$ y $H(S/R)$

(1) Obsérvese que es del todo posible la aplicación de la desigualdad de Jensen, pues la función exponencial es una función convexa y (1.70) nos expresa su media ponderada.

la de $H(S/r)$ con respecto a $r \in \Gamma$.

Teniendo ahora en cuenta (1.34) y (1.35), podemos escribir el riesgo total de (1.71) en la forma

$$R_T(P_R, \hat{s}) \geq \left[\frac{M}{2e\pi} |D|^{1/M} \exp \left\{ \frac{2}{M} H(S) \right\} \exp \left\{ - \frac{2}{M} I(S \parallel R) \right\} \cdot \exp \left\{ + \frac{2}{M} \varepsilon_m \right\} \right] \quad (1.72)$$

donde, al igual que para el caso de la función de pérdida del valor absoluto, (1.12), podemos interpretar, tanto en (1.69) como en (1.72), que $\varepsilon(r, \alpha)$ y ε_m se consideren como la pérdida de información debida al tomar un proceso de decisión puntual.

Eliminando ε_m en (1.72) obtenemos la cota, análoga a (1.37), siguiente

$$R_T(P_R, \hat{s}) \geq F'_S R_1 \quad (1.73)$$

siendo ahora

$$F'_S = \frac{M}{2e\pi} |D|^{1/M} \exp \left\{ \frac{2}{M} H(S) \right\} \quad (1.74)$$

y R_1 la parte restante, es decir

$$R_1 = \exp \left\{ - \frac{2}{M} I(S \parallel R) \right\} \quad (1.75)$$

donde F'_S depende únicamente de la ley a priori que consideremos y R_1 de la información transmitida a través del canal. Vemos, además, que R_1 de (1.75) coincide con el correspondiente a la función de pérdida (1.12), es decir, con (1.39).

Calculamos seguidamente F'_S para los dos ejemplos ya cita

dos en la subsección 1.2.2., pero para la función de coste cuadrático. Consideraremos el caso unidimensional ($M = 1$) y $D \equiv 1$.

1.3.2. Ejemplos.

Comprobaremos en este caso que \mathcal{F}_S va a depender, directamente, de la varianza de la ley a priori, por lo que introduciremos ahora el concepto de riesgo total relativo.

Caso 1. - Ley de probabilidad a priori uniforme.

Al igual que en el caso 1. de la subsección 1.2.2. supondremos que la ley con que se emiten las señales para ser enviadas a través del canales la ley uniforme en el intervalo $(\mu - \frac{w}{2}, \mu + \frac{w}{2})$

De ella sabemos que⁽¹⁾

$$E[S] = \mu ; \quad \sigma_S^2 = \frac{w^2}{12} \quad \text{y} \quad H(S) = \ln w \quad (1.76)$$

por lo que, de (1.74), concluimos que

$$\mathcal{F}'_S = \frac{1}{2e\pi} \cdot 2 \ln w = \frac{w^2}{2\pi e} \quad (1.77)$$

y, teniendo en cuenta (1.76), quedaría

$$\mathcal{F}'_S = \frac{6}{\pi e} \sigma_S^2 \quad (1.78)$$

Caso 2. - Ley de probabilidad a priori $N(\mu, \sigma_S^2)$.

Seguimos en el caso unidimensional y con la función de pérdida del caso anterior, es decir $C(s, \hat{s}) = (s - \hat{s})^2$, en este caso, y teniendo en cuenta (1.45), así como (1.74) para la función de pérdida indicada anteriormente, deducimos que

 (1) Bastaría considerar S.Wilks, 1.962) y (1.42).

$$\begin{aligned}
 \mathcal{F}_s &= \frac{1}{2e\pi} \exp \left\{ 2 \left[\ln \sigma_s \sqrt{2\pi} + \frac{1}{2} \right] \right\} = \\
 &= \frac{1}{2e\pi} (\sigma_s \sqrt{2\pi})^2 \cdot e = \sigma_s^2 \quad (1.79)
 \end{aligned}$$

1.3.3. Riesgo total relativo.

Estos dos ejemplos nos han sugerido la introducción de un nuevo concepto que, para funciones de pérdida del tipo (1.51), nos va a permitir dar una cota, análoga a (1.47), en la que todos los factores que intervienen en la misma sean constantes para cada sistema de comunicación y no dependan para nada de las leyes de probabilidad que rigen el proceso de emisión de señales.

DEFINICION 3.1.- Definimos riesgo total relativo para la función de pérdida cuadrática, y notamos R_{Tr} a la cantidad

$$R_{Tr} = \frac{R_P(P_R, \hat{s})}{\sigma_s^2} \quad (1.80)$$

donde con σ_s^2 notamos la varianza de la ley de probabilidad a priori.

Teniendo en cuenta (1.73), así como la definición 3.1, podemos escribir el riesgo total relativo de la siguiente forma

$$R_{Tr} = g_2 R_s \quad (1.81)$$

donde con g_2 hemos expresado

$$g_2 = \frac{\mathcal{F}_s}{\sigma_s^2} \quad (1.82)$$

que es un parámetro independiente de la ley de probabilidad a

priori. Es de señalar, por otro lado, que para las funciones de probabilidad a priori de los dos ejemplos anteriores, que son las que aparecen con mayor frecuencia en Teoría de la Información, el coeficiente ξ_2 es del orden de la unidad.

Siguiendo un razonamiento análogo al realizado para la función de pérdida cuadrática absoluto concluimos que la capacidad del canal dada en (1.49), que es una cota superior para $I(S||R)$, nos proporcionaría una cota inferior para R_I , y, por tanto, para el riesgo total relativo R_{Tr} . Tenemos entonces

$$R_{Tr} > \xi_2 \exp \left\{ - \frac{2C}{M} \right\} \quad (1.83)$$

siendo todas las constantes de la parte derecha parámetros independientes de las leyes de probabilidad que rigen el sistema de comunicación.

1.4. RELACION CON LA COTA DE SHANNON

Aunque nosotros hemos obtenido (1.47) y (1.73) a partir de la cota general (1.10), para diferentes funciones de pérdida, creemos que es posible obtener la cota (1.47) ó (1.73) al minimizar el riesgo total sujeto a la restricción de que la entropía de la ley de probabilidad a priori es fija, ya que la solución de este problema de minimización ha de ser la misma que la del problema de maximizar la entropía⁽¹⁾ a priori bajo la condición de que el valor medio de las funciones de pérdida (1.12) ó (1.51) sea fijo, problema que ya fué considerado por Shannon [10] en

(1) Problemas de este tipo han sido estudiados por Shannon [9], Ash [4], Miron Tribus: "Decisions rationnels en l'incertain", Gil y Melendreras: "Un algoritmo para el problema de máxima entropía en el caso discreto,

1.949 considerando las funciones que él llamó razones de distorsión para medir la fidelidad de un canal.

Por esto es por lo que podemos observar que existe una cierta similitud entre (1.50), (1.83) y la cota de Shannon que, para un criterio de fiabilidad cualquiera tiene la forma

$$R_T \hat{P}_R(\hat{s}) \geq \Delta^{-1}(C) \quad (1.84)$$

donde con Δ^{-1} queremos expresar la inversa de la razón de distorsión de Shannon. No obstante, la cota de Shannon (1.84) depende de la capacidad del canal C a través de la función Δ^{-1} que, a su vez, depende de la fuente de señales, mientras que la dependencia de C en nuestra cota (1.47) ó (1.73) tiene un carácter universal.

BIBLIOGRAFIA

1. ASH, R. : "Information Theory" Interscience Publishers 1967.
2. BECKMAN : "Convexity" Ed, Springer Verlag.
3. BELLMAN, R.: "Introducción al Análisis Matricial". Ed. Reverté.
4. FOURGBAUD-FUSCH : "Statistique" Ed. Duñod.
5. KINCHIN, A.I.: "Mathematical Foundations of Information Theory"
Ed. Dover Publications, 1957.
6. KULLBACK, S.: "Information Theory and Statistics" Ed. Dover Pu
blications, 1968.
7. MIDDLETON-VAN METER : "Detection and Extraction of Signals in
Noise from the Point of View of Statistical Decision
Theory". J. Soc. Ind. Appl. Math. Vol,3 Nº 4, pp 192
a 253, 1955; vol. 4 Nº 2, pp 86-119, 1956.
8. MUNROE, M. E.: "Measure an Integration" Ed. Addison Wesley 1971.
9. SHANNON, C.E.: "A Mathematical Theory of Communication"
Bell Sys. Tch. J. Vol. 27. 1949 pp 379-423; 623-656.
10. SHANNON C.E.: "Coding Theorems for a discrete source with a
Fidelity Criterion" IRE Nat. Conv. Rec. pt. 4. Mar.
1959. pp 147-163.
11. VAN METER y MIDDLETON, D. : "Modern Statistical Approaches to
Reception in Communication Theory". Trans. Prof. -
Group Information Theory, Sept, 1954. pp119-145 -
se. 35.
12. WILKS, S.: "Mathematical Statistics". Ed. John Wiley ands Sons
1962.

CAPITULO II

COTA INFERIOR PARA EL RIESGO EN LA DETECCION

2.1. Cota para el riesgo a posteriori

- Teorema 1
- Teorema 2
- Corolario
- Teorema 3
- Teorema 4
- Teorema 5

2.2. Cotas para el riesgo total

2.3. Relación con la Cota de Fano

Bibliografía

COTA INFERIOR PARA EL RIESGO EN LA DETECCION

Continuamos aquí el problema de hallar una cota inferior - para el riesgo, ya comenzado en el capítulo anterior, pero ahora para la detección (dominio de decisiones⁽¹⁾), Middleton y Van Meter, 1955). Hallamos, por caminos similares a los del capítulo anterior, una cota para la detección y, finalmente, vemos su relación con la desigualdad de Fano,

2.1. COTA PARA EL RIESGO A POSTERIORI EN LA DETECCION.

Vamos a probar que es posible emplear el proceso utilizado en el capítulo anterior para la extracción, a este nuevo caso, - sin más que exigir que

$$\sigma(\Omega) < +\infty \quad (2.1)$$

Supondremos, para ello, el espacio de señales Ω dividido en dos subconjuntos Ω_0 y su complementario, correspondientes a ruidos solamente y al envío de señal. Para cada valor recibido, r , sea $\hat{d}(r)$ la función de decisión que nos va a conducir a Ω_0 .

(1) Los dominios de decisiones son una generalización del conjunto de decisiones en el caso de mensajes discretos y de intervalos de decisiones en el caso de mensajes unidimensionales.

o a su complementario; notaremos, por tanto, con \hat{d} a un subespacio de Ω .

La función de pérdida, que notaremos $C(s, \hat{d})$, será

$$C(s, \hat{d}) = \begin{cases} 0, & \text{si } s \in \hat{d}(r) \\ 1, & \text{si } s \notin \hat{d}(r) \end{cases} \quad (2.2)$$

y, para ella, el riesgo a posteriori, que definimos de una forma análoga a (1.1), y notamos $R(r, \hat{d})$, será [2]

$$\begin{aligned} R(r, \hat{d}) &= \int_{\Omega} C(s, \hat{d}) p(s/r) d\sigma = P \{ s \notin \hat{d}(r)/r \} = \\ &= 1 - P_{S/r}(\hat{d}(r)) = P_e(r) \end{aligned} \quad (2.3)$$

donde con $P_e(r)$ hemos notado la probabilidad de error⁽¹⁾ al adoptar la decisión incorrecta.

El riesgo total, según (1.2), sería

$$R_T(P_R, \hat{d}) = \int_T P_e(r) dP_R = \bar{P}_e \quad (2.4)$$

Al igual que en (1.3), definimos la función de densidad $p(s, \hat{d}, \alpha)$ de la siguiente forma

$$p(s, \hat{d}, \alpha) = k(\alpha) \exp \{ -\alpha C(s, \hat{d}) \} \quad (2.5)$$

donde α es una constante y $k(\alpha)$ sería el factor de normalización,

 (1) Si r es tal que $\hat{d}(r) = \Omega_0$, sería $P_e(r)$ la probabilidad de elegir la hipótesis de que no se había emitido ninguna señal cuando sí había ocurrido; sería el error de tipo II. Si r es tal que $\hat{d}(r) = \Omega - \Omega_0$, $P_e(r)$ sería la probabilidad de elegir la hipótesis de que se había emitido señal cuando no era así. Sería el error de tipo I.

que calculamos como sigue

$$\begin{aligned}
 1 &= \int_{\Omega} p(s, \hat{d}, \alpha) d\sigma = \int_{\Omega} k(\alpha) \exp \{-\alpha c(s, \hat{d})\} d\sigma = \\
 &= k(\alpha) \left[\int_{\Omega - \hat{d}} e^{-\alpha} d\sigma + \int_{\hat{d}} d\sigma \right] = k(\alpha) \left[e^{-\alpha} [\sigma(\Omega - \hat{d}) + \sigma(\hat{d})] \right]
 \end{aligned}
 \tag{2.6}$$

de donde concluimos que

$$k(\alpha) = \left[e^{-\alpha} [\sigma(\Omega - \hat{d}) + \sigma(\hat{d})] \right]^{-1}
 \tag{2.7}$$

Sea, de nuevo, $\xi(r, \alpha)$ el valor definido en (1.6). Operando de forma análoga a como lo hicimos en el anterior capítulo, llegamos a la identidad

$$P_e(r) = \frac{1}{\alpha} \left[H(S/r) + \ln k(\alpha) + \xi(r, \alpha) \right]
 \tag{2.8}$$

de donde podríamos obtener para $P_e(r)$ una cota del todo análoga a (1.10). No obstante, para la función de pérdidas definida en (2.2), podemos actuar de dos formas, al igual que en la extracción, bien buscando aquél valor α_0 para el que obtengamos el mayor valor de (1.11), bien tomando el valor α_1 que sea solución de la ecuación (1.26), ambas para la función de pérdidas indicada anteriormente y para la función de densidad (2.5). Este será el camino que seguiremos.

Por ser $p(s, \hat{d}, \alpha)$ función de densidad, es evidente que

$$\int_{\Omega} p(s, \hat{d}, \alpha) d\sigma = 1
 \tag{2.9}$$

y teniendo en cuenta (2.7) y la definición de $p(s, \hat{d}, \alpha)$, resulta que (2.9) se puede escribir

$$1 = \frac{1}{e^{-\alpha} \sigma(\Omega - \hat{d}) + \sigma(\hat{d})} \int_{\Omega} e^{-\alpha} c(s, \hat{d}) d\sigma \quad (2.10)$$

Derivando ahora en ambos miembros con respecto a α , se sigue

$$0 = - \int_{\Omega} \frac{c(s, \hat{d})}{e^{-\alpha} \sigma(\Omega - \hat{d}) + \sigma(\hat{d})} e^{-\alpha} c(s, \hat{d}) d\sigma + \\ + \frac{\sigma(\Omega - \hat{d}) e^{-\alpha}}{[e^{-\alpha} \sigma(\Omega - \hat{d}) + \sigma(\hat{d})]^2} \int_{\Omega} e^{-\alpha} c(s, \hat{d}) d\sigma \quad (2.11)$$

de donde

$$\int_{\Omega} c(s, \hat{d}) p(s, \hat{d}, \alpha) d\sigma = \frac{\sigma(\Omega - \hat{d}) e^{-\alpha}}{e^{-\alpha} \sigma(\Omega - \hat{d}) + \sigma(\hat{d})} \quad (2.12)$$

y si tenemos en cuenta (2.3) y (1.26) para la función de pérdida (2.2), entonces (2.12) se puede escribir

$$P_e(r) = \frac{e^{-\alpha_1} \sigma(\Omega - \hat{d})}{e^{-\alpha_1} \sigma(\Omega - \hat{d}) + \sigma(\hat{d})} \quad (2.13)$$

de donde obtenemos

$$e^{-\alpha_1} (1 - P_e(r)) = \frac{\sigma(\hat{d}) P_e(r)}{\sigma(\Omega - \hat{d})} \quad (2.14)$$

es decir

$$e^{-\alpha_1} \left(\frac{P_e(r)}{1 - P_e(r)} \right) = \frac{\sigma(\hat{d})}{\sigma(\Omega) - \sigma(\hat{d})} \quad (2.15)$$

y despejando α_1 , de (2.15), obtenemos

$$\alpha_1 = - \ln \frac{P_e(r)}{1 - P_e(r)} \cdot \frac{\rho}{1 - \rho} \quad (2.16)$$

donde con ρ hemos expresado

$$\rho = \frac{\sigma(\hat{d})}{\sigma(\Omega)} \quad (2.17)$$

que, en cierto sentido, nos mide la amplitud del dominio que elijamos, para cada $r \in \Gamma$, dada.

Si tenemos en cuenta (2.7) y sustituimos el valor de α por el obtenido en (2.16), resulta

$$k(\alpha_1) = \frac{1}{\frac{P_e(r)}{1 - P_e(r)} \cdot \frac{1}{1 - \rho} \cdot \frac{\sigma(\Omega - \hat{d}) - \sigma(\hat{d})}{\sigma(\Omega)}} \quad (2.18)$$

pero dividiendo numerador y denominador por $\sigma(\Omega)$ y teniendo en cuenta (2.17), de (2.18) se concluye que

$$k(\alpha_1) = \frac{\frac{1/\sigma(\Omega)}{P_e(r)}}{\left[\frac{1 - P_e(r)}{1 - P_e(r)} + 1 \right] \rho} = \frac{1 - P_e(r)}{\rho \sigma(\Omega)} \quad (2.19)$$

Tomando logaritmos neperianos en (2.19) y sustituyendo en (2.8) vemos que

$$P_e(r) = \frac{1}{\ln \left[\frac{1 - P_e(r)}{P_e(r)} \cdot \frac{1 - \rho}{\rho} \right]} \left[H(S/r) + \ln(1 - P_e(r)) - \ln \rho - \ln \sigma(\Omega) + \varepsilon(r, \alpha) \right] \quad (2.20)$$

que simplificando y sacando factores comunes se reduce a

$$\begin{aligned} & - (1 - P_e(r)) \ln(1 - P_e(r)) - P_e(r) \ln P_e(r) + (1 - \rho) \ln \rho + \\ & + P_e(r) \ln(1 - \rho) = H(S/r) - \ln \sigma(\Omega) + \varepsilon(r, \alpha) \end{aligned} \quad (2.21)$$

que podemos escribir

$$P \ln(1 - a) + (1 - P) \ln a + H(P) = b \quad (2.22)$$

donde hemos expresado

$$P = P_e(r); \quad a = p \quad ; \quad b = H(S/r) - \ln \sigma(\Omega) + \varepsilon(r, \alpha) \quad (2.23)$$

y con $H(P)$ notamos la entropía de Shannon para el esquema probabilístico de dos resultados posibles; uno con probabilidad P y su complementario con probabilidad $1 - P$.

La función (2.22) se puede expresar como $b = F_0(P, a)$, para $0 \leq P \leq 1$ y $0 \leq a \leq 1$.

TEOREMA 1.1. Para cada a fijo, $F_0(P, a)$ es una función cóncava de P .

En efecto, de (2.22) se sigue que

$$\begin{aligned} F_0(P, a) &= P \ln(1 - a) + (1 - P) \ln a - P \ln P - \\ &\quad - (1 - P) \ln(1 - P) \end{aligned} \quad (2.24)$$

Sea ahora a cualquier valor fijo, $0 < a < 1$; tendremos que

$$\frac{\partial F_0}{\partial P} = \ln(1 - a) - \ln a - \ln P - 1 + \ln(1 - P) + 1 \quad (2.25)$$

y la segunda derivada

$$\frac{\partial^2 F_0}{\partial P^2} = - \frac{1}{P} - \frac{1}{1 - P}$$

que es menor que Cero ya que $0 < P < 1$, con lo que quedaría probado el teorema.

Señalemos que no hemos considerado los casos $a = 0$ ni $a = 1$ porque para ellos la función $F_0(P, a)$ vale $-\infty$ salvo para los puntos $P = 1$ y $P = 0$, respectivamente, en que vale Cero.

TEOREMA 1.2. Para cada P fijo, la función $F_0(P, a)$ es una función cóncava de a.

En efecto, sea P un valor fijo, $0 \leq P \leq 1$; de (2.24) se deduce, al hallar la primera derivada parcial con respecto a a de la función $F_0(P, a)$ que

$$\frac{\partial F_0}{\partial a} = - \frac{P}{1-a} + \frac{1-P}{a} \quad (2.27)$$

y hallando su segunda derivada parcial respecto de a

$$\frac{\partial^2 F_0}{\partial a^2} = - \frac{P}{(1-a)^2} - \frac{1-P}{a^2} \quad (2.28)$$

que es negativa al ser cero menor que a.

COROLARIO 1.1. La función $F_0(P, a)$ no es positiva para ningún valor de a y P.

Para demostrarlo vamos a calcular, para cada a fijo, el valor de P que hace máximo $F_0(P, a)$. Para ello igualaremos a cero la primera derivada parcial de dicha función con respecto a P, obteniendo como valor que la anula $P = P_1$, que cumpliría

$$\frac{1 - P_1}{P_1} = \frac{a}{1-a} \quad (2.29)$$

luego, para cada a fijo, el valor que maximiza $F_0(P, a)$ es ⁽¹⁾

$$P_1 = 1 - a \quad (2.30)$$

y como la función es cóncava, sería, efectivamente, un máximo.

(1) - - - - -
A un resultado análogo al expresado en (2.30), habríamos llegado al considerar, para cada P fijo, el valor de a que maximice la función $F_0(P, a)$.

Sustituyendo ahora (2.30) en (2.24), obtenemos que

$$F_0(P_1, a) = 0 \quad (2.31)$$

de lo cual deducimos que la función crece desde $\ln a$ hasta alcanzar el valor cero y luego decrece hasta el valor $\ln(1 - a)$.

Si fuese $a = 0$, sería $F_0(P, 0) = -\infty$ salvo para $P = 1$, en que vale cero; por otro lado, si $a = 1$, sería $F_0(P, 1) = -\infty$ salvo para $P = 0$, en que vale cero.

También podemos ver que para P fijo, la función crece desde $-\infty$ hasta el valor cero, que lo alcanza para $a = 1 - P$, y después decrece nuevamente hasta $-\infty$.

Si $P = 0$, se verifica que $F_0(0, a) = \ln a$, que es siempre negativo salvo para $a = 1$ en que vale cero; si fuese $P = 1$, entonces $F_0(1, a) = \ln(1 - a)$ que siempre es negativo excepto para $a = 0$, en que se anula.

Hemos probado, pues, que $b = F_0(P, a) \leq 0$ para todos los valores de a y de P , y que para aquellos valores de a y P que satisfagan $a + P = 1$, alcanzan su valor máximo, que es $b = 0$.

Por consiguiente, si consideramos (2.22) como una ecuación para P , se verifica que para cada a y b fijos obtenemos, en general, dos soluciones. Notemos por $F_1(b, a)$ la menor de ellas. De forma análoga, podemos considerar (2.22) como una ecuación para a ; para cada P y b fijos obtenemos, en general, dos soluciones. Notemos por $F_2(b, P)$ la menor de ellas.

Supongamos ahora el caso en que

$$P + a < 1 \quad (2.32)$$

con lo que tanto $F_1(b, a)$ como $F_2(b, P)$ tienen una única solución, baste recordar que para $a + P = 1$ se alcanzaba el máximo en la función $b = F_0(P, a)$. Por lo tanto, para aquellos valores en que $a + P < 1$, la función $b = F_0(P, a)$ va creciendo y para cada b y P fijos nos da una solución única para a , al igual que para cada b y a fijos nos la proporciona para P .

TEOREMA 1.3. $F_1(b, a)$, para cada a fijo y $F_2(b, P)$, para cada P fijo, son funciones estrictamente crecientes de b .

a) Veamos que, para a fijo, $P = F_1(b, a)$ es una función estrictamente creciente de b .

Consideremos para ello la función $F(a, b, P) = 0$, que obtenemos de (2.22), es decir

$$F(a, b, P) = P \ln(1 - a) + (1 - P) \ln a - P \ln P - (1 - P) \ln(1 - P) - b = 0 \quad (2.33)$$

Por el análisis elemental sabemos que

$$\frac{\partial P}{\partial b} = - \frac{\partial F / \partial P}{\partial F / \partial b} \quad (2.34)$$

y como

$$\begin{aligned} \frac{\partial F}{\partial P} &= \ln(1 - a) - \ln a - \ln P + \ln(1 - P) = \\ &= \ln \left[\frac{1 - P}{P} \cdot \frac{1 - P}{a} \right] \end{aligned} \quad (2.35)$$

y, por otro lado, es

$$\frac{\partial F}{\partial b} = -1 \quad (2.36)$$

resulta que, de (2.34)

$$\frac{\partial P}{\partial b} = \ln \left[\frac{1-a}{P} \frac{1-P}{a} \right] \quad (2.37)$$

y, si tenemos en cuenta (2.32), tanto $(1-a)/P$ como $(1-P)/a$ son estrictamente mayores que uno, por lo que $\partial P/\partial b$ es estrictamente mayor que cero quedando probada, por tanto, esta parte del teorema.

b) De forma análoga a la anterior veremos que $a = F_2(b, P)$ para cada P fijo, es una función estrictamente creciente de b .

Para ello nos apoyaremos en (2.33) y en

$$\frac{\partial a}{\partial b} = - \frac{\partial F/\partial a}{-\partial F/\partial b} \quad (2.38)$$

y como

$$\frac{\partial F}{\partial a} = - \frac{P}{1-a} + \frac{1-P}{a} \quad \text{y} \quad \frac{\partial F}{\partial b} = -1 \quad (2.39)$$

obtenemos, según (2.38) que

$$\frac{\partial a}{\partial b} = - \frac{P}{1-a} + \frac{1-P}{a} \quad (2.40)$$

y teniendo en cuenta (2.32), vemos que $P < 1-a$ y que $a < 1-P$ por lo que (1)

$$\frac{\partial a}{\partial b} > -1 + \frac{1-P}{a} = \frac{1-a-P}{a} > 0 \quad (2.41)$$

con lo que queda probado el teorema 1.3.

TEOREMA 1.4. $F_1(b, a)$ y $F_2(b, P)$, para cada a y P fijos - respectivamente, son funciones convexas de b .

a) Veamos en primer lugar que $P = F_1(b, a)$, para cada a fi

(1) Es de señalar que $a > 0$; basta considerar (2.32).

jo, es una función convexa de b . Para ello nos apoyaremos en [5]

$$\frac{\partial^2 F}{\partial b^2} + 2 \frac{\partial^2 F}{\partial b \partial P} \frac{\partial P}{\partial b} + \frac{\partial^2 F}{\partial P^2} \left(\frac{\partial P}{\partial b}\right)^2 + \frac{\partial F}{\partial b} \cdot \frac{\partial^2 P}{\partial b^2} = 0 \quad (2.42)$$

de donde obtenemos que

$$\frac{\partial^2 P}{\partial b^2} = \left[\frac{1}{P} + \frac{1}{1-P} \right] \ln \left[\frac{1-a}{P} \frac{1-P}{a} \right] > 0 \quad (2.43)$$

ya que $a + P < 1$, con lo que queda probada esta parte del teorema.

b) Apoyándonos en idéntico resultado, pero ahora para la función $a = F_2(b, P)$ y para P fijo, obtenemos que

$$\begin{aligned} \frac{\partial^2 a}{\partial b^2} &= \left[\frac{P}{(1-a)^2} + \frac{1-P}{a^2} \right] \cdot \left[-\frac{P}{1-a} + \frac{1-P}{a} \right] = \\ &= \left[\frac{P}{(1-a)^2} + \frac{1-P}{a^2} \right] \cdot \left[\frac{1-a-P}{a(1-a)} \right] > 0 \end{aligned} \quad (2.44)$$

sin más que considerar (2.32), con lo que queda probada también la segunda parte del teorema.

TEOREMA 1.5. Se verifica que

$$F_2(b, P) > \exp \left\{ [b - H(P)] / (1 - P) \right\}$$

y

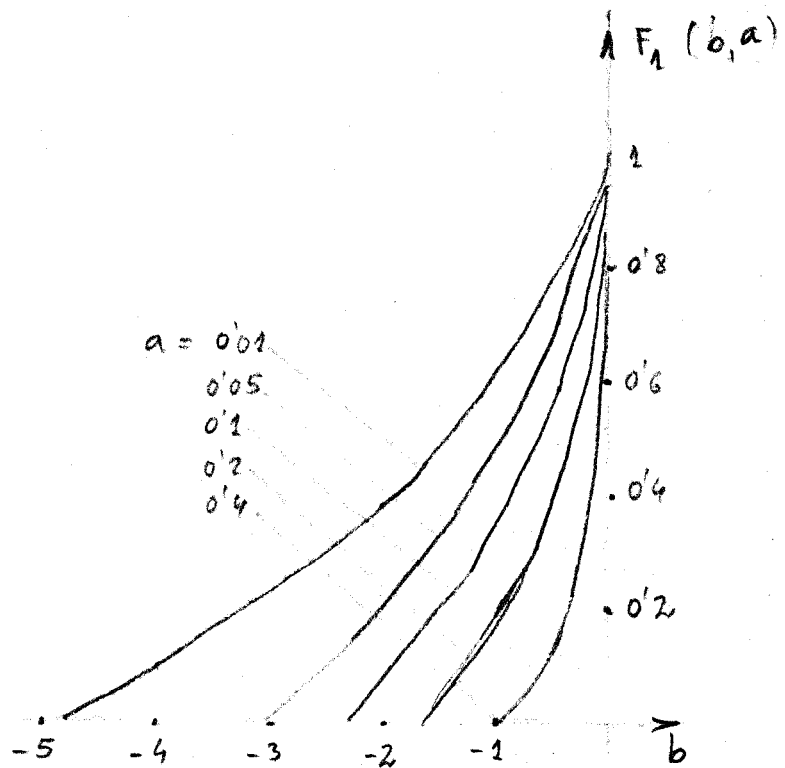
$$\lim_{P \rightarrow 0} F_2(b, P) e^{-b} = 1$$

En efecto, de (2.22) se deduce que

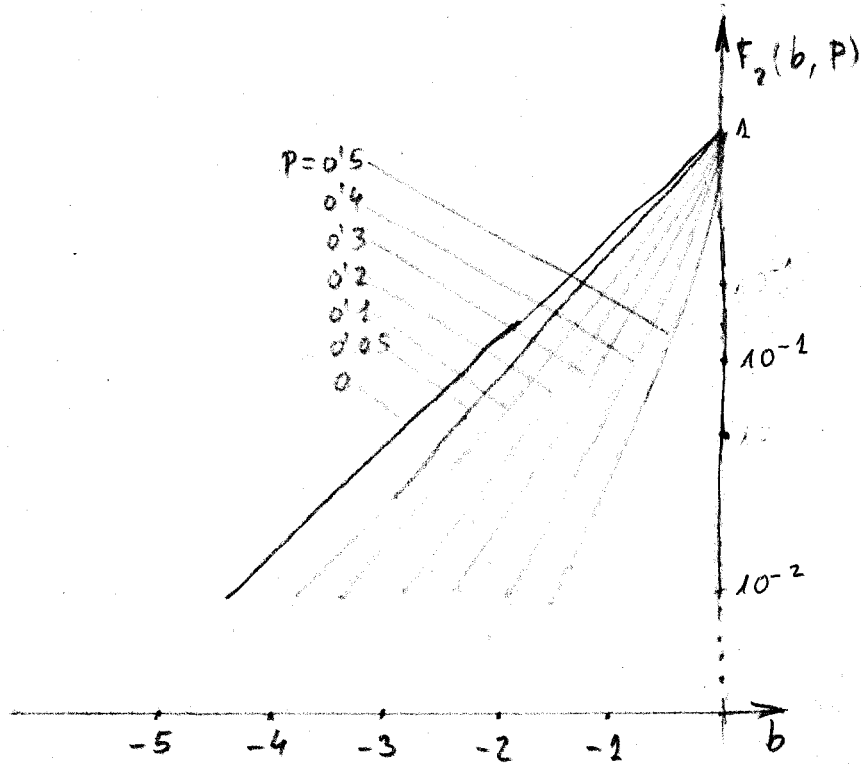
$$(1 - P) \ln a + P \ln(1 - a) = b - H(P) \quad (2.45)$$

y como $P > 0$ y $\ln(1 - a) < 0$, podemos escribir

$$(1 - P) \ln a > b - H(P) \quad (2.46)$$



Representación gráfica de la función $F_1(b, a)$



Representación gráfica de la función $F_2(b, p)$

y dividiendo ambos miembros por $1 - P$ que es mayor que cero, la igualdad subsiste y queda

$$\ln a > \frac{b - H(P)}{1 - P} \quad (2.47)$$

con lo que está probada la primera parte del teorema.

Para probar la segunda parte, sustituimos $a = F_2(b, P)$ en (2.45) y tomando límites para $P \rightarrow 0$, concluimos que⁽¹⁾

$$\lim_{P \rightarrow 0} \ln F_2(b, P) = b \quad (2.48)$$

o bien

$$\lim_{P \rightarrow 0} F_2(b, P) = e^b \quad (2.49)$$

con lo que probamos la segunda parte del teorema.

2.2. COTAS PARA EL RIESGO TOTAL

En la mayoría de las aplicaciones prácticas se considera el caso en que

$$P_e(r) + a < 1 \quad (2.50)$$

que es el mismo que hemos considerado hasta aquí a través de la referencia (2.32).

De (2.22) se sigue que

$$P_e(r) = F_1 \left[H(S/r) - \ln \sigma(\Omega) + \varepsilon(r, \alpha_1), \rho \right] \quad (2.51)$$

que se corresponde con la ecuación (1.8) del capítulo anterior, y

$$= F_2 \left[H(S/r) - \ln \sigma(\Omega) + \varepsilon(r, \alpha_1), P_e(r) \right] \quad (2.52)$$

que, en general, es una función de r .

Si notamos ahora con $\bar{\rho}$ el valor medio de ρ con respecto a r , y con \bar{P}_e el valor medio de $P_e(r)$ con respecto a r , y tenemos en cuenta el teorema 1.4. que nos indica que tanto (2.51) como (2.52) son funciones convexas de $H(S/r) - \ln \sigma(\Omega) + \varepsilon(r, \alpha_1)$ para ρ y $P_e(r)$ fijos respectivamente, podemos aplicarles la desigualdad de Jensen [] y así obtenemos, para (2.51)

$$\bar{P}_e \geq F_1 \left[- (I(S \parallel R) + I_0 - \bar{\varepsilon}), \bar{\rho} \right] \quad (2.53)$$

y para (2.52)

$$\bar{\rho} \geq F_2 \left[- (I(S \parallel R) + I_0 - \bar{\varepsilon}), \bar{P}_e \right] \quad (2.54)$$

donde con I_0 hemos notado

$$I_0 = \ln \sigma(\Omega) - H(S) \quad (2.55)$$

y el parámetro $\bar{\varepsilon}$ es el valor medio de $\varepsilon(r, \alpha_1)$ para r , que, nuevamente, se puede interpretar como la pérdida de información al realizar la detección, mientras que a I_0 se le podría interpretar como la información a priori.

Por otro lado, y teniendo en cuenta el teorema 1.3, F_1 y F_2 son funciones estrictamente crecientes, para P_e y ρ fijos respectivamente, y como $\bar{\varepsilon} \geq 0$, podemos prescindir de él, obteniendo

$$\bar{P}_e \geq F_1 \left[- (I(S \parallel R) + I_0), \bar{\rho} \right] \quad (2.56)$$

y

$$\bar{\rho} \geq F_2 \left[- (I(S \parallel R) + I_0), \bar{P}_e \right] \quad (2.57)$$

respectivamente.

2.3. RELACION CON LA COTA DE FANO

Consideremos ahora el caso en que Ω sea un conjunto discreto que contenga N elementos. Definamos $\sigma(S)$, para cualquier subconjunto S de Ω , como el número de elementos que posee y - consideremos nuevamente el proceso de decisión puntual; es decir, supongamos que $\hat{d}(r)$ contiene solamente un punto. Entonces $\rho = 1/N$ y es fácil ver que (2.56) es una cota equivalente a la desigualdad de Fano [1]-[6]

Para ello, sustituyendo los resultados que consideramos - en este caso en (2.21), obtenemos

$$H(S/r) + \varepsilon(r, \alpha_1) - \ln N = H[P_e(r)] - (1 - P_e) \ln N + P_e \ln \frac{N-1}{N} \quad (2.58)$$

y, operando en (2.58), llegamos a

$$H(S/r) = H[P_e(r)] + P_e(r) \ln(N-1) - \varepsilon(r, \alpha_1) \quad (2.59)$$

si tomamos valores medios en (2.59), con respecto a r ,

$$H(S/R) = H(\bar{P}_e) + \bar{P}_e \ln(N-1) + \bar{\varepsilon} \quad (2.60)$$

y como $\bar{\varepsilon} > 0$, al eliminarlo de (2.60) obtenemos la desigualdad dada por Fano.

Es evidente, pues, que dicha desigualdad se puede considerar como un caso particular de la desigualdad (2.57) obtenida anteriormente.

Finalmente, señalemos que $F_1[-(I(S||R) + I_0) \rho]$ y $F_2[-(I(S||R) + I_0) \bar{P}_e]$ están tanto más próximas a \bar{P}_e y a ρ respectivamente

cuanto más pequeño sea $\varepsilon(r, \alpha_1)$, cosa que es posible si la función de densidad $p(s/r)$ tiene la forma de la función $p(s, \hat{d}, \alpha)$.

BIBLIOGRAFIA

1. FANO R.M.: "Transmission of Information" John Wiley. New York 1961.
2. FOURGEBAUD-FUSH: "Statistique" Ed. Dunod.
3. KALTZ S. : "Some inequalities for convex Functions useful in - Information Theory SIAM. Rev.Vol. 7 nº 3 July 1965
4. MIDDLETON y VAN METER: "Detection and Extraction of Signals in Noise from the Point of View of Statistical Decision Theory" J. Soc. Ind. Appl. Math. Vol. 3 Nº 4 pp. 192-253, 1955; vol. 4 Nº 2, pp. 86-119, 1956.
5. SAN JUAN R. : Análisis Matemático.
6. THOMAS : "An Introduction to Statistical Communication Theory" Wiley, 1969.

ooo O ooo

CAPITULO III

DETECCION Y ESTIMACION SIMULTANEA DE SEÑALES PARA HIPOTESIS MULTIPLES

- 3.1. Planteamiento general del problema
- 3.2. Resolución del problema de hipótesis disjuntas
- 3.3. Casos binarios especiales

Caso 1: Detección pura

Caso 2: Detección y extracción conjunta

Caso 3: Detección y estimación conjunta para el
caso binario ($\vec{\theta}_1 \neq \vec{0}$ frente $\vec{\theta}_2 \neq \vec{0}$)

Bibliografía

DETECCION Y ESTIMACION SIMULTANEA DE SENALES PARA HIPOTESIS MULTIPLES

En este capítulo estudiamos el problema de detección y extracción en el caso de alternativas múltiples. Tratamos de determinar la estructura del receptor de forma que nos proporcione, a su salida, decisiones óptimas tanto para la detección como para la estimación de parámetros desconocidos, bien de las señales, bien del canal a través del cual se envían las mismas.

Consideramos, únicamente, el caso en que las hipótesis no se solapan y obtenemos, finalmente, una regla de decisión óptima.

3.1. PLANTEAMIENTO GENERAL DEL PROBLEMA.

Supondremos, a lo largo de este capítulo y del siguiente, que los datos solamente se pueden recoger durante un intervalo de tiempo de duración T . Asimismo, a cada elemento del conjunto finito de mensajes $\{m_j; 0 \leq j \leq M\}$, consistente en $N + 1$ elementos, siendo $N \geq 2$, le corresponderá una señal distinta seleccionada del conjunto $\{S_j(t); 0 \leq j \leq M\}$. Esta correspondencia entre cada mensaje m_j y su señal $S_j(t)$ supondremos que lie-

va asociada una etapa intermedia consistente en la selección de un parámetro vectorial $\bar{\theta}_j$, descriptivo de la señal, que va asociado a su vez a una señal ondulatoria $S(t, \bar{\theta}_j)$ que notaremos - abreviadamente también como $S_j(t)$.

Cada vez que la fuente emite un símbolo m_j , se envía a través del canal la señal ondulatoria correspondiente $S(t, \bar{\theta}_j)$, - llegando al receptor la onda $R(t, \bar{\theta}_j) = S(t, \bar{\theta}_j) + n(t)$, donde $n(t)$ es la realización de un proceso de ruido de tipo aditivo. La onda recibida, $R(t, \bar{\theta}_j)$, dependerá, en general, de un vector paramétrico desconocido, resultado del desconocimiento de la señal ondulatoria o bien del desconocimiento de los parámetros descriptivos del canal.

Necesitamos hacer algunas consideraciones acerca de la posibilidad de que la señal no esté presente durante el intervalo de observación $(0, T)$, y que, por consiguiente, los datos recibidos constan únicamente de ruido. Supondremos que esta situación se corresponde con el mensaje m_0 , que va a ir asociado a - la señal $S_0(t) = 0$, emitida por la fuente de señales. No obstante aparece un problema cuando tratamos de identificar sus - conjuntos de un espacio arbitrario de parámetros de señales con subconjuntos de un espacio de señales. Generalmente, un espacio de señales S se considera como una colección de señales

$\{S(t, \bar{\theta}) / \bar{\theta} \in \Omega\}$ donde el parámetro $\bar{\theta}$, que toma sus valores en un conjunto de parámetros Ω , las diferencia. Vamos a identificar el espacio paramétrico Ω con el espacio de señales S a fin de simplificar, en todo lo posible, la notación. Esta identificación nos va a permitir considerar el espacio de parámetros

como un caso especial del espacio de señales para el caso $\vec{\theta} = \vec{s}$ donde $\vec{\theta}$ es el vector de $2WT$ puntos muestrales⁽¹⁾[3]. Si no existen restricciones sobre el espacio paramétrico, no es cierto, en general, que $\vec{\theta} = \vec{0}$ implique que $\vec{s} = \vec{0}$, es decir, el parámetro nulo puede no corresponder a la señal nula⁽²⁾. Por tanto, si queremos llegar a identificar el espacio de señales con un espacio paramétrico, y tenemos en cuenta la posibilidad de que solo se presente ruido, deberemos restringir los espacios de parámetros. La restricción impuesta será que los parámetros considerados sean parámetros característicos de la energía de la señal⁽³⁾[2], para los cuales se verifica que $\vec{\theta} = \vec{0}$ implica que $\vec{s} = \vec{0}$. Notaremos Ω_0 al subconjunto del espacio de señales o espacio paramétrico de señales Ω correspondiente a la señal nula.

Al espacio de todos los datos observados R le notaremos Γ al igual que hemos notado al espacio de todas las posibles parámetros de señales Ω . Los espacios Ω y Γ son bastante arbitrarios y las probabilidades se pueden interpretar como apropiadas medidas definidas sobre ellos.

En la subsección 3.2. consideramos el caso en que los subconjuntos $\Omega_0, \Omega_1, \dots, \Omega_M$ del espacio Ω asignados a cada mensaje m_0, m_1, \dots, m_M son mutuamente disjuntos o no solapados.

(1) Esta condición nos la expresa el teorema muestral de Shannon que demuestra que cualquier onda de duración arbitraria T , pero de frecuencia limitada a un máximo de W ciclos por segundo, se puede describir completamente tomando una muestra de su amplitud ordenada a intervalos de $1/2W$ segundos.

(2) El ángulo de fase φ de una señal sinusoidal $\sin(\omega t + \varphi)$ es un buen ejemplo para mostrar este caso.

(3) Ejemplos típicos de estos parámetros son la amplitud y duración de una señal.

A estos subconjuntos les llamaremos conjuntos de señales o conjuntos de hipótesis. Conceptualmente es más fácil tratar el caso de no solapamiento, ya que siempre es posible, determinar, para el vector paramétrico θ seleccionado por el transmisor y la decisión γ tomada por el receptor, si ésta ha sido correcta o no. Asociando ahora una función de pérdida o costo $C(\theta, \gamma)$ con cada par (θ, γ) , que podemos interpretar como el costo de una decisión, bien correcta o no, trataremos de obtener la decisión que minimice el riesgo total.

Desgraciadamente, la suposición de que el espacio paramétrico está subdividido en conjuntos no solapados restringe el dominio de definición de las densidades de probabilidad empleadas en la teoría de la decisión bayesiana a subconjuntos propios de θ , en lugar de a todo el espacio, lo que excluye el empleo de ciertas funciones de densidad. Así, por ejemplo, si los parámetros son los $N = 2WT$ valores muestrales entonces el espacio es euclídeo N -dimensional y los diferentes espacios de señales, no solapados, $\Omega_1, \Omega_2, \dots, \Omega_M$, son subconjuntos de este espacio, lo que excluye el empleo de densidades normales multivariantes por ser su dominio de definición el espacio completo.

Las reglas de decisión óptimas las encontraremos mediante un proceso de minimización bietápico, que consiste en determinar, en primer lugar, una regla de detección unida a un estimador desconocido de la señal o del parámetro que la determina, que al ser sustituida en la expresión del riesgo total nos proporciona un mínimo relativo y, en segundo lugar, un estimador que minimice dicho riesgo total. La regla de decisión final para este pro-

ceso bietápico es la primera regla de detección junto con el estimador óptimo sustituido en el lugar del estimador desconocido antes del segundo paso. Esta aproximación analítica nos sugiere que la operación de estimación tal vez debiera preceder a la operación de detección ya que se emplea el estimador óptimo como parte del proceso de detección; no obstante, lo lógico de esta aproximación ha sido comentado por Kendall y Stuart [5].

Espósito en [1] y en [2], demuestra, siguiendo este proceso que la razón de verosimilitud para señales estocásticas normales distorsionadas por ruidos gaussianos se puede expresar como función de un estimador de mínima varianza de la señal bajo la hipótesis de que la señal está presente con probabilidad uno, y Kailath [4] ha generalizado recientemente estos resultados. No obstante, debemos señalar que en nuestro caso las dos operaciones están íntimamente relacionadas y ambas ocurren simultáneamente. A la salida del receptor obtendremos un par de decisiones: una consistente en una decisión para la detección y la otra una decisión para la extracción.

Nos interesa saber si existe o no una regla de decisión no aleatorizada que alcance el riesgo de Bayes. Middleton y Espósito (1970) han demostrado, por una aproximación análoga a la dada en [9] por Oggs que las reglas de decisión bayesianas para detección y extracción conjunta son no aleatorizadas. Siguiendo su trabajo es fácil demostrar que también se verifica para el caso de hipótesis múltiples.

3.2. RESOLUCION DEL PROBLEMA PARA HIPOTESIS DISJUNTAS

El minimizar el riesgo total asociado con las operaciones combinadas de detección y extracción nos proporciona un método, no solamente para analizar el caso binario (D. Middleton y R. Esposito, 1968), sino también para la misma situación en la recepción para el caso de alternativas múltiples, [6]. De acuerdo con ello, comenzamos con la expresión para el riesgo total en el caso de las operaciones combinadas de detección y extracción, que notaremos por R_{D+E} , y que viene dado por

$$R_{D+E} = \int_{\mathcal{A}} \int_{\mathcal{R}} \int_{\Omega} C(\bar{\theta}, \gamma) P(\bar{\theta}, \bar{r}, \gamma) d\bar{\theta} d\bar{r} d\gamma \quad (3.1)$$

En esta ecuación $\bar{\theta}$ denota el vector paramétrico de la señal, \bar{r} es el vector cuyas componentes son los datos muestreados de la onda recibida $r(t, \bar{\theta})$, y γ es la decisión tomada. El término $C(\bar{\theta}, \gamma)$ es el coste asociado a la decisión γ cuando $\bar{\theta}$ es el conjunto de parámetros de la señal presente en el intervalo de observación, mientras que $P(\bar{\theta}, \bar{r}, \gamma)$ es una ley de probabilidad conjunta.

Como cualquier decisión γ puede depender únicamente de los datos observados \bar{r} y no dependerá de los no observados, como es el caso del verdadero valor del parámetro $\bar{\theta}$, el riesgo total se puede expresar como

$$R_{D+E} = \int_{\mathcal{R}} d\bar{r} \int_{\mathcal{A}} d\gamma \int_{\Omega} d\bar{\theta} C(\bar{\theta}, \gamma) P(\gamma/\bar{r}) p(\bar{r}/\bar{\theta}) p(\bar{\theta}) \quad (3.2)$$

ya que siempre es posible escribir

$$P(\bar{\theta}, \bar{r}, \gamma) = P(\gamma/\bar{r}) p(\bar{r}/\bar{\theta}) p(\bar{\theta}) \quad (3.3)$$

y como la decisión γ es independiente de $\tilde{\theta}$ y depende únicamente de las señales recibidas \tilde{r} , la expresión (3.3) para $p(\tilde{\theta}, \tilde{r}, \gamma)$ se puede escribir nuevamente

$$p(\tilde{\theta}, \tilde{r}, \gamma) = p(\gamma / \tilde{r}) p(\tilde{r} / \tilde{\theta}) p(\tilde{\theta}) \quad (3.4)$$

y, de esta forma, hemos obtenido (3.2) a partir de (3.2).

El espacio de todos los patos vectoriales observados lo hemos notado por Γ , el espacio de todas las posibles decisiones con Δ y el de todas los posibles parámetros de la señal por Ω . En particular, cualquiera de los tres espacios pueden estar constituido por un conjunto discreto de puntos, en cuyo caso las integrales se interpretan como sumas para estos puntos [4].

Volviendo a nuestra expresión (3.2) para el riesgo total, observamos que la probabilidad a priori $p(\tilde{\theta})$ es posible descomponerla de la siguiente forma

$$p(\tilde{\theta}) = \sum_{i=1}^M p(\tilde{\theta} / H_i) p(H_i) \quad (3.5)$$

donde H_i representa la hipótesis de que en el intervalo de observación se encuentra presente una señal de la clase Ω_i , y $p(H_i)$ es la probabilidad de que se presenten señales de la clase Ω_i , siendo $p(\tilde{\theta} / H_i)$ la función de densidad que nos describe la probabilidad de que $\tilde{\theta}$ tome distintos valores en la clase Ω_i ; y como $p(\tilde{\theta} / H_i)$ solamente tiene significado para valores de $\tilde{\theta}$ contenidos en Ω_i , podemos suponer que $p(\tilde{\theta} / H_i)$ vale cero fuera de Ω_i y escribir

$$p(\tilde{\theta} / H_i) = p_i(\tilde{\theta}) I_{\Omega_i}(\tilde{\theta}) \quad (3.6)$$

donde I_{Ω_i} es la función indicador del conjunto Ω_i ; es decir,

$$I_{\Omega_i} = \begin{cases} 1 & \text{si } \theta \in \Omega_i \\ 0 & \text{en cualquier otro caso} \end{cases} \quad (3.7)$$

A partir de aquí, mientras no se exprese explícitamente, usaremos la notación abreviada p_i en lugar de $p(H_i)$ y, sustituyendo esta notación en (3.5) resulta que

$$P(\theta) = \sum_{i=1}^n p_i P_i(\theta) I_{\Omega_i}(\theta) \quad (3.8)$$

y si sustituimos (3.8) en la expresión (3.2) para el riesgo total

$$\begin{aligned} R_D + E &= \int_r \left[\int_a \int_{\Omega} d\theta C(\theta, \gamma) p(\gamma/\bar{r}) p(\bar{r}/\theta) \right. \\ &\quad \left. \left[\sum_{i=1}^n p_i P_i(\theta) I_{\Omega_i}(\theta) \right] \right] - \\ &= \int_r \left[\int_a d\gamma p(\gamma/\bar{r}) \left[\sum_{i=1}^n p_i \int_{\Omega} C(\theta, \gamma) p(\bar{r}/\theta) \right. \right. \\ &\quad \left. \left. P_i(\theta) I_{\Omega_i}(\theta) \right] d\theta \right] - \\ &= \int_r \left[\int_a p(\gamma/\bar{r}) \left[\sum_{i=1}^n p_i \int_{\Omega} d\theta C(\theta, \gamma) p(\bar{r}/\theta) P_i(\theta) \right] \right] \end{aligned} \quad (3.9)$$

donde la última igualdad se sigue del hecho de ser I_{Ω_i} cero fuera de Ω_i .

La siguiente etapa es decidir sobre la estructura del espacio de decisión Δ . Para ello tomamos como Δ el siguiente conjunto discreto

donde

γ_0 : decisión de considerar la hipótesis H_0 como cierta; es decir, solamente se presenta ruido, y entonces no necesitamos ningún estimador.

γ_1 : la decisión de que la hipótesis H_1 es cierta; es decir, una señal de la clase Ω está presente, y necesitamos un estimador $\hat{\theta}(\bar{r})$ del parámetro de la señal θ que se encuentra presente.

Como el espacio de decisión es un conjunto finito de elementos discretos, la integral sobre el espacio de decisión se convierte en una suma⁽¹⁾. Reemplazando ahora la integral por la suma en la expresión (3.9)

$$R_D + E = \int_{\bar{r}} dr \left\{ \sum_{\gamma_0} p(\gamma_0 / \bar{r}) \left[\sum_{\theta_0} P_1(\theta_0) C(\theta_0) p(\bar{r} / \theta_0) P_1(\theta_0) \right] + \sum_{\gamma_1} p(\gamma_1 / \bar{r}) \left[\sum_{\theta_1} P_1(\theta_1) C(\theta_1) p(\bar{r} / \theta_1) P_1(\theta_1) \right] \right\} \quad (3.11)$$

Al final de cada intervalo de observación hemos de tomar necesariamente una decisión y, además, la siguiente condición se verifica

$$\sum_{\gamma_0} p(\gamma_0 / \bar{r}) + \sum_{\gamma_1} p(\gamma_1 / \bar{r}) = 1 \quad (3.12)$$

Esta condición nos permite expresar $p(\gamma_1 / \bar{r})$ como

$$p(\gamma_1 / \bar{r}) = 1 - \sum_{\gamma_0} p(\gamma_0 / \bar{r}) \quad (3.13)$$

que, sustituido en $R_D + E$ queda

$$R_D + E = \int_{\bar{r}} p(\gamma_0 / \bar{r}) \left[\sum_{\theta_0} P_1(\theta_0) C(\theta_0) p(\bar{r} / \theta_0) P_1(\theta_0) \right] d\theta_0 dr + \int_{\bar{r}} p(\gamma_1 / \bar{r}) \left[\sum_{\theta_1} P_1(\theta_1) C(\theta_1) p(\bar{r} / \theta_1) P_1(\theta_1) \right] d\theta_1 dr \quad (3.14)$$

y sustituyendo en (3.14) la expresión (3.13)

$$R_D + E^- \int \left[\sum p_1 \right] C(\hat{\theta}, r) p(r/\hat{\theta}) p_1(\hat{\theta}) d\hat{\theta} dr + \\ + \int \left\{ \sum p(r/\hat{\theta}) \left[\sum p_1 \right] [C(\hat{\theta}, r) - C(\hat{\theta}, \bar{r})] d\hat{\theta} \right\} dr \quad (3.15)$$

Supondremos que el factor $C(\hat{\theta}, r)$ es mayor o igual que cero, cosa bastante razonable, pues en el caso en que estamos considerando, es decir, de conjuntos de señales disjuntos, existe un coste de decisión correcta si $\hat{\theta} = \bar{0}$, lo de decisión incorrecta sin $\hat{\theta} \neq \bar{0}$. Además, cuando decidimos $\hat{\theta}$, no necesitamos ningún estimador, por lo que $C(\hat{\theta}, r)$ no depende de \bar{r} a través de un estimador $\hat{\theta}(\bar{r})$. La ecuación (3.15) se puede escribir ahora de la siguiente forma

$$R_D + E^- \int A_0(\bar{r}) d\bar{r} + \int \sum p(r/\bar{r}) B_j'(\bar{r}) d\bar{r} \quad (3.16)$$

donde

$$A_0(\bar{r}) = \sum p_1 \int C(\hat{\theta}, r) p(r/\hat{\theta}) p_1(\hat{\theta}) d\hat{\theta} \quad (3.17)$$

$$B_j'(\bar{r}) = \sum p_1 \int [C(\hat{\theta}, r) - C(\hat{\theta}, \bar{r})] p(r/\hat{\theta}) p_1(\hat{\theta}) d\hat{\theta} \quad (3.18)$$

De nuestros comentarios anteriores acerca de $C(\hat{\theta}, r)$, es evidente que $A_0(\bar{r}) > 0$ y, por tanto, el primer término de (3.15) es una constante positiva independiente de cualquier regla de decisión, lo que nos indica que se puede omitir en el proceso de minimización que determina la regla de decisión óptima. Para rea

lizar el análisis de la manera más análoga a la situación de detección pura para alternativas múltiples, definimos la siguiente razón media de verosimilitud modificada

$$L_{1j}(\vec{r}) = \frac{\int_{\Omega} [C(\vec{\theta}, \gamma_j) - C(\vec{\theta}, \gamma_0)] p(\vec{r}/\vec{\theta}) p_1(\vec{\theta}) d\vec{\theta}}{[C(\vec{0}, \gamma_j) - C(\vec{0}, \gamma_0)] p(\vec{r}/\vec{0})} \quad (3.19)$$

para $i, j = 1, 2, \dots, M$; donde $\mu_1 = \frac{P_1}{P_0}$.

La razón de verosimilitud $L_{11}(\vec{r})$ corresponde a la razón de verosimilitud modificada Λ_g de Middleton y Espósito. Notando ahora $p_0(\vec{\theta}) = \delta(\vec{\theta} - \vec{0})$, donde $\delta(\cdot)$ es la función δ de Dirac, y usando la definición de $L_{1j}(\vec{r})$, podemos escribir (3.18) en la forma

$$\begin{aligned} B_j'(\vec{r}) &= p_0 [C(\vec{0}, \gamma_j) - C(\vec{0}, \gamma_0)] p(\vec{r}/\vec{0}) - \\ &\quad - \sum_{i=1}^M p_i \int_{\Omega} [C(\vec{\theta}, \gamma_j) - C(\vec{\theta}, \gamma_0)] p(\vec{r}/\vec{\theta}) p_1(\vec{\theta}) d\vec{\theta} = \\ &= p_0 [C(\vec{0}, \gamma_j) - C(\vec{0}, \gamma_0)] p(\vec{r}/\vec{0}) \left[1 - \sum_{i=1}^M L_{1j}(\vec{r}) \right] = \\ &= p_0 p(\vec{r}/\vec{0}) C_j(\vec{r}) B_j(\vec{r}) \quad (3.20) \end{aligned}$$

donde

$$B_j(\vec{r}) = 1 - \sum_{i=1}^M L_{1j}(\vec{r}) \quad i, j = 1, 2, \dots, M \quad (3.21)$$

$$C_j(\vec{r}) = C(\vec{0}, \gamma_j) - C(\vec{0}, \gamma_0) \quad j = 1, 2, \dots, M \quad (3.22)$$

con lo que el riesgo total sería

$$R_D + B = \int_{\mathcal{R}} A_0(\vec{r}) d\vec{r} + \int_{\mathcal{R}} p_0 p(\vec{r}/\vec{0}) \sum_{j=1}^M p(\gamma_j/\vec{r}) C_j(\vec{r}) B_j(\vec{r}) d\vec{r} \quad (3.23)$$

Suponemos que el coste de una decisión incorrecta es mayor que el coste de una decisión correcta, que, a menudo, se le asigna el valor cero [11]. Por tanto, $C_j(\vec{r}) \geq 0$ para todo $\vec{r} \in \Gamma$ y $j = 1, 2, \dots, M$. Ya que el primer término de (3.23) es una constante y tanto p_0 como $p(\vec{r}/\hat{\theta})$ son probabilidades y, por consiguiente, positivas, vemos que la siguiente regla de decisión no aleatoria minimiza el riesgo total.

Regla de decisión: Decidir γ_i ó $p(\gamma_i / \vec{r}) = 1$ si

$$a) B_j(\vec{r}) < 0 \quad (D.1.)$$

$$b) C_j(\vec{r}) B_j(\vec{r}) \leq C_k(\vec{r}) B_k(\vec{r}) \quad \forall i$$

Es evidente, de la definición (3.21) de $B_j(\vec{r})$ que la condición a) se puede reemplazar por $\sum_{i=1}^M L_{ij}(\vec{r}) > 1$, que es similar al resultado obtenido para detección en el caso de hipótesis múltiples por Middleton y Espósito [8]. De hecho, aún suponiendo que los costes $C(\vec{\theta}, \gamma_i)$ sean independientes de $\vec{\theta}$, llegaríamos a un resultado idéntico.

Como $C(\vec{\theta}, \gamma_i)$ refleja no solamente el costo de la operación de detección o clasificación, sino también el de la operación de estimación, debemos tener en cuenta la posibilidad de que $C(\vec{\theta}, \gamma_i)$ sea dependiente, tanto de $\vec{\theta}$ como de \vec{r} , así, por ejemplo, una posible función de costo, que usaremos más tarde, será la que, para cada $\vec{\theta} \in \Omega$, y cada $\gamma_i \in \Delta$ toma el valor

$$C(\vec{\theta}, \gamma_i) = c_{ij}^{(1)} + c_{ij}^{(2)} (\vec{\theta} - \hat{\theta}(\vec{r}))^T D_i \vec{\theta} - \hat{\theta}(\vec{r}) \quad (3.24)$$

donde $c_{ij}^{(1)}$ es el coste constante que se asigna para la detección

y $C_{i1}^{(2)}(\hat{\theta} - \hat{\theta}(\vec{r}))^T D(\hat{\theta} - \hat{\theta}(\vec{r}))$ es una función de coste cuadrático que aparece en la estimación de mínimos cuadrados.

La regla de decisión obtenida anteriormente depende de la función de costo $C(\hat{\theta}, \gamma_j)$, que, como indicamos en el párrafo anterior, puede contener una función desconocida de \vec{r} , como el estimador $\hat{\theta}(\vec{r})$ del parámetro de la señal presente. Por tanto, solamente la primera etapa de las dos de que consta el proceso de minimización se ha completado.

La estrategia obtenida en este proceso bietápico de minimización es determinar, en primer lugar, una regla de decisión que depende de un estimador $\hat{\theta}(\vec{r})$ y que será óptima cuando dicho estimador lo sea. Por consiguiente, la segunda etapa del proceso de minimización consistirá en calcular la función $\hat{\theta}_j^*(\vec{r})$ que minimice, además, el riesgo total. Teniendo en cuenta la regla de decisión anterior, el riesgo total para la detección y extracción conjunta se puede escribir de la siguiente manera

$$R_D + E \int_{\Gamma} A_0(\vec{r}) dr + \sum_{j=1}^N \int_{\Gamma_j} p_0 p(\vec{r}/\vec{0}) C_j(\vec{r}) B_j(\vec{r}) dr \quad (3.25)$$

donde Γ_j indica la región del espacio de observación para la que decidimos γ_j , es decir

$$\Gamma_j = \left\{ \vec{r} / B_j(\vec{r}) < 0 \text{ y } C_j(\vec{r}) B_j(\vec{r}) \leq C_k(\vec{r}) B_k(\vec{r}) \quad \forall k \right\} \quad (3.26)$$

Si en la primera etapa decidimos γ_j , el estimador óptimo que resulta del segundo proceso de minimización, y que notaremos $\hat{\theta}_j(\vec{r})$, se determinará mediante la siguiente condición

$$\hat{\theta}_j(\vec{r}) = \min_{\theta} \int_{\Gamma_j} p_0 p(\vec{r}/\vec{0}) C_j(\vec{r}) B_j(\vec{r}) dr \quad (j \neq 0) \quad (3.27)$$

Donde hemos incluido un subíndice j en el estimador óptimo para indicar, de esta forma, que hemos tomado la decisión $\hat{\theta}_j$, que es equivalente a decidir que la señal presente pertenece al subconjunto Ω_j , $j \neq 0$, siendo, además, el estimador resultante de (3.27) el mejor estimador que podemos obtener de la señal del subconjunto Ω_j . Para obtener la regla de detección óptima final, lo que haremos será sustituir el estimador óptimo que hemos encontrado en la segunda etapa en la regla de decisión obtenida en la primera, es decir (D.1).

3.3. CASOS BINARIOS ESPECIALES

En esta subsección consideraremos algunos casos especiales de los resultados obtenidos hasta ahora y, además, haremos comparaciones con resultados obtenidos anteriormente. Los casos considerados a continuación son: (1) detección pura; aquí no es necesario ninguna operación de estimación. Supondremos el caso binario típico $\bar{\theta} = \bar{0}$, frente a $\bar{\theta} = \bar{0}$ (2) el problema de detección y extracción conjunta para el caso $\bar{\theta} \neq \bar{0}$ frente a $\bar{\theta} = \bar{0}$, y, (3) el problema de detección y estimación conjunta para el caso $\bar{\theta}_1 \neq \bar{0}$ frente a $\bar{\theta}_2 \neq \bar{0}$.

Caso 1.- Detección pura.

Aquí nos interesamos únicamente por la operación de detección y, por tanto, los costes funcionales se reducen a los costes constantes del proceso de detección, es decir

$$\begin{aligned}
 c(\bar{\theta} \in \Omega_0, \gamma_0) &= C_{00} & c(\bar{\theta} \in \Omega_1, \gamma_0) &= C_{10} \\
 c(\bar{\theta} \in \Omega_0, \gamma_1) &= C_{01} & c(\bar{\theta} \in \Omega_1, \gamma_1) &= C_{11}
 \end{aligned}$$

(3.28)

para el caso binario $i, j = 0, 1$ y, por tanto, solamente existe una razón de verosimilitud del tipo (3.19), que será $L_{11}(\vec{r})$. Para la asignación de costes (3.28), dicha razón de verosimilitud se expresa

$$L_{11}(\vec{r}) = \frac{\int_0^1 (C_{10} - C_{11}) p(\vec{r}/\bar{\theta}) p_1(\bar{\theta}) d\bar{\theta}}{(C_{01} - C_{00}) p(\vec{r}/0)} = \frac{(C_{10} - C_{11})}{(C_{01} - C_{00})} \Lambda(\vec{r}) \quad (3.29)$$

donde $\Lambda(\vec{r})$ es la razón de verosimilitud usada normalmente en la detección [12], es decir

$$\Lambda(\vec{r}) = \frac{p_1 \int_0^1 p(\vec{r}/\bar{\theta}) p_1(\bar{\theta}) d\bar{\theta}}{p_0 p(\vec{r}/0)} \quad (3.30)$$

De acuerdo con nuestra regla de decisión (D.1), elegiremos γ_1 , es decir, hay señal presente, si $B_1(\vec{r}) < 0$; pero como $B_1(\vec{r}) = 1 - L_{11}(\vec{r})$, elegiremos γ_1 si $L_{11}(\vec{r}) > 1$, o, equivalentemente, si

$$\Lambda(\vec{r}) > \frac{C_{01} - C_{00}}{C_{10} - C_{11}} = K \quad (3.31)$$

resultado, muy conocido en la detección de señales en presencia de ruidos. Comparamos la razón de verosimilitud $\Lambda(\vec{r})$ con un umbral K , y tomamos la decisión de que existe señal durante el intervalo de observación si $\Lambda(\vec{r}) > K$; si no ocurriera esto decidimos que únicamente se encuentra presente el ruido.

Caso 2.- Detección y extracción conjunta.

En este nuevo apartado los estimadores forman parte de nues

tra regla de decisión en el caso de que decidamos que se encuentra presente una señal y, por consiguiente, tendremos que incluir en nuestra asignación de costes un factor que tenga presente el coste adicional de la estimación: ⁽¹⁾

$$\begin{aligned}
 c(\bar{\theta} \in \Omega_0, \gamma_0) &= c(\bar{0}, \gamma_0) = c_{00}^{(1)} + c_{00}^{(2)}(\bar{\theta}, \hat{\theta}) = c_{1-\alpha} + c_{00} \\
 c(\bar{\theta} \in \Omega_0, \gamma_1) &= c(\bar{0}, \gamma_1) = c_{01}^{(1)} + c_{01}^{(2)}(\bar{\theta}, \hat{\theta}) = c_{\alpha} + f_{01}(\hat{\theta}) \\
 c(\bar{\theta} \in \Omega_1, \gamma_0) &= c(\bar{\theta}, \gamma_0) = c_{10}^{(1)} + c_{10}^{(2)}(\bar{\theta}, \hat{\theta}) = c_{\beta} + f_{10}(\bar{\theta}) \\
 c(\bar{\theta} \in \Omega_1, \gamma_1) &= c(\bar{\theta}, \gamma_1) = c_{11}^{(1)} + c_{11}^{(2)}(\bar{\theta}, \hat{\theta}) = c_{1-\beta} + f_{11}(\bar{\theta}, \hat{\theta})
 \end{aligned}
 \tag{3.32}$$

Nuevamente, ya que estamos tratando el caso binario y, j = 0, 1 y solamente existe una razón de verosimilitud, $L_{11}(r)$, que viene dada por (3.19), tendremos, para los costes (3.32), reemplazando $\hat{\theta}(\bar{r})$ por $\bar{\theta}(\bar{r})$,

$$\begin{aligned}
 L_{11}(\bar{r}) &= \frac{(c_{10}^{(1)} - c_{11}^{(1)}) \Lambda(\bar{r}) + \int_{\hat{\theta}} (c_{10}^{(2)}(\bar{\theta}, \hat{\theta}) - c_{11}^{(2)}(\bar{\theta}, \hat{\theta}))}{c_{01}^{(1)} - c_{00}^{(1)} + c_{01}^{(2)}(\bar{\theta}, \hat{\theta}) - c_{00}^{(2)}(\bar{\theta}, \hat{\theta})} \\
 &\quad \cdot \frac{p(\bar{r}/\bar{\theta})}{p(\bar{r}/0)} p_1(\bar{\theta}) d\bar{\theta}
 \end{aligned}
 \tag{3.33}$$

que, teniendo en cuenta (3.32), podemos expresar

$$\begin{aligned}
 L_{11}(\bar{r}) &= \frac{(c_{\beta} - c_{1-\beta}) \Lambda(\bar{r})}{(c_{\alpha} - c_{1-\alpha}) + f_{10}(\bar{\theta}) - c_{00}} + \\
 &\quad + \frac{\int_{\hat{\theta}} [f_{01}(\hat{\theta}) - f_{11}(\bar{\theta}, \hat{\theta})] \Lambda(\bar{r}, \bar{\theta}) p_1(\bar{\theta}) d\bar{\theta}}{(c_{\alpha} - c_{1-\alpha}) + f_{10}(\bar{\theta}) - c_{00}}
 \end{aligned}
 \tag{3.34}$$

donde $\Lambda(\bar{r})$ es la razón de verosimilitud generalizada expresada

anteriormente y $\Lambda(\hat{r}, \hat{\theta})$ es la razón de verosimilitud ordinaria [10], es decir

$$\Lambda(\hat{r}, \hat{\theta}) = \frac{p_1 p(r/\hat{\theta})}{p_0 p(r/0)} \quad (3.35)$$

Supongamos ahora que $C_{1-x} = C_{1-\alpha} = C_{00} = f_{10}(\hat{\theta}) = 0$; entonces podemos escribir (3.34) en la forma

$$L_{11}(\hat{r}) = \frac{C_r}{C_x} \Lambda(\hat{r}) + \frac{1}{C_x} \int_{\hat{\theta}_1} f_{01}(\hat{\theta}) - f_{11}(\hat{\theta}, \hat{\theta}) \Lambda(r, \hat{\theta}) p_1(\hat{\theta}) d\hat{\theta} \quad (3.36)$$

que es idéntica a la expresión (4.13) del trabajo de Middleton y Espósito de 1968.

Según nuestra regla de decisión (D.1), si $L_{11}(r) > 1$, decidimos que, durante el intervalo de observación, se encuentra presente una señal y que el estimador óptimo de θ debe ser $\hat{\theta}^*(r)$ pero, por el contrario, si $L_{11}(\hat{r}) \leq 1$, debemos decidir que no hay ninguna señal presente, en cuyo caso no necesitamos ningún estimador.

Caso 3. - Detección y estimación conjunta para el caso binario ($\hat{\theta}_1 = \hat{\theta}$, frente a $\hat{\theta}_2 = \hat{\theta}$).

Mejor que tratar de modificar los resultados de la subsección 3.2. para acomodar la situación al caso en que la hipótesis nula sea excluida, es más fácil calcular la regla de decisión para la situación de recepción cuando una de las M señales distintas de cero pueda estar presente durante el intervalo de observación.

Para excluir la hipótesis nula, supondremos que $p_0 = 0$ y

$p(i/\bar{r}) = 0$. El riesgo total $R_D + E$, de (3.11), será, ahora

$$R_D + E = \int_{\bar{r}} \left[\sum_{i=1}^M p(i/\bar{r}) \left[\sum_{j=1}^M p_j \right] C(\bar{\theta}, y_j) p(\bar{r}/\bar{\theta}) p_1(\bar{\theta}) d\bar{\theta} \right] d\bar{r} \quad (3.37)$$

y como

$$p(y_k/\bar{r}) = 1 - \sum_{i=1}^M p(i/\bar{r}) \quad (3.38)$$

y si sustituimos esta expresión en (3.37), la expresión para $R_D + E$ sería

$$R_D + E = \int_{\bar{r}} \left[\sum_{j=1}^M p_j \right] C(\bar{\theta}, y_j) p(\bar{r}/\bar{\theta}) p_1(\bar{\theta}) d\bar{\theta} + \sum_{i=1}^M p(i/\bar{r}) B_{jk}(\bar{r}) \Big] d\bar{r} \quad (3.39)$$

donde con $B_{jk}(\bar{r})$ hemos expresado la cantidad

$$B_{jk}(\bar{r}) = \sum_{i=1}^M p_i \left[C(\bar{\theta}, y_i) - C(\bar{\theta}, y_k) \right] p(\bar{r}/\bar{\theta}) p_1(\bar{\theta}) d\bar{\theta} \quad (3.40)$$

Señalemos que el primer término de la parte derecha de (3.39), es una constante positiva, independiente de las probabilidades $p(i/\bar{r})$ que personifican las reglas de decisión. Por consiguiente, podemos elegir la siguiente regla de decisión no aleatorizada.

Decidir: y_1 (j+k) o bien $p(y_1/\bar{r}) = 1$, si

a) $B_{jk}(\bar{r}) < 0$

b) $B_{jk}(\bar{r}) \leq B_{ik}(\bar{r}) \quad \forall i \neq k$

En caso contrario, decidir y_k o $p(y_k/\bar{r}) = 1$

Sea, nuevamente, Γ_j la región del espacio de observación para la que decidimos γ_j . El riesgo total (3.39), para la regla de decisión anterior, queda como sigue

$$R_D + B = \int_{\Gamma} \left[\sum_{i=1}^M p_i \right]_{\Omega} C(\bar{\theta}, \gamma_k) p(\bar{r}/\bar{\theta}) p_1(\bar{\theta}) d\bar{\theta} \, d\bar{r} + \sum_{j=1}^M B_{jk}(\bar{r}) \, d\bar{r} \quad (3.41)$$

El estimador óptimo para el caso en que decidimos γ_j , lo notaremos por $\hat{\theta}_j(\bar{r})_{p, < 1}$, que es una función de sonido sobre Γ_j y que minimiza la expresión (3.41) para el riesgo total. Como el primer término de la parte derecha de (3.41) solamente concierne a la decisión γ_k y, además, es independiente de $\hat{\theta}_j(\bar{r})_{p, < 1}$ la condición que determina el estimador es

$$\min_{\hat{\theta}_j(\bar{r})_{p, < 1}} \left[\sum_{i=1}^M p_i \right]_{\Omega} (C(\bar{\theta}, \gamma_j) - C(\bar{\theta}, \gamma_k)) p(\bar{r}/\bar{\theta}) p_1(\bar{\theta}) d\bar{\theta} \, d\bar{r} \quad (3.42)$$

y como $C(\bar{\theta}, \gamma_k)$ es independiente de $\hat{\theta}_j(\bar{r})_{p, < 1}$, ésta última relación se convierte en

$$\min_{\hat{\theta}_j(\bar{r})_{p, < 1}} \left[\sum_{i=1}^M p_i \right]_{\Omega} C(\bar{\theta}, \gamma_j) p(\bar{r}/\bar{\theta}) p_1(\bar{\theta}) d\bar{\theta} \, d\bar{r} \quad (3.43)$$

Pero si fuese $p(\gamma_j / \bar{r}) = 0$ para todo j que sea distinto de k , es decir, decidimos γ_k , entonces el riesgo total viene dado por el primer término de (3.39) y el estimador óptimo $\hat{\theta}_k(\bar{r})_{p, < 1}$ para este caso se determina por la siguiente condición

$$\min_{\hat{\theta}_k(\bar{r})} \left[\sum_{i=1}^M p_i \right]_{\Omega} C(\bar{\theta}, \gamma_k) p(\bar{r}/\bar{\theta}) p_1(\bar{\theta}) d\bar{\theta} \, d\bar{r} \quad (3.44)$$

que es idéntica a (3.42) si hacemos en ésta última $j = k$.

La regla de decisión óptima sería la (D.2), con $\hat{\theta}_j(\bar{r})_{p, < 1}$

para $j = 1, 2, \dots, M$, en el lugar de los estimadores desconocidos que aparecen en dicha regla de decisión. Si especializamos ahora las expresiones anteriores para el caso binario con dos señales diferentes, podemos tomar $k = 2$ y $j = 1$ y entonces

$$B_{12}(\vec{r}) = \sum_{i=1}^2 p_i \int_{\Omega} (C(\vec{\theta}, r_1) - C(\vec{\theta}, r_2)) p(\vec{r}/\vec{\theta}) d\vec{\theta} \quad (3.45)$$

y los estimadores se determinan por las siguientes condiciones

$$\min_{\hat{\theta}_j(\vec{r})} \int_{\Omega} \left[\sum_{i=1}^2 p_i \int_{\Omega} C(\vec{\theta}, r_j) p(\vec{r}/\vec{\theta}) p_1(\vec{\theta}) d\vec{\theta} \right] d\vec{r} \quad (j = 1, 2) \quad (3.46)$$

Si notamos ahora por $B_{12}^*(\vec{r})$ a (3.42) con $\hat{\theta}_1(\vec{r})$ y $\hat{\theta}_2(\vec{r})$ sustituyendo a $\hat{\theta}_1(\vec{r})$ y $\hat{\theta}_2(\vec{r})$ respectivamente, la regla de decisión óptima viene dada, en este caso, por

$$p^*(r_1 / \vec{r}) = 1 \text{ si } B_{12}(\vec{r}) < 0$$

$$p^*(r_2 / \vec{r}) = 1 \text{ si } B_{12}(\vec{r}) \geq 0$$

BIBLIOGRAFIA

1. Esposito R. "A class of Estimation for Optimun Adaptive Detec-
tion" Information and Est. Vol. 12 pp 137-148 -
February 1967.
2. ESPOSITO R.: "On a Relation Between Detection and Estimation -
in Decission Theory" Inf, y Cont. Vol. 12 n^o 2 -
pp 116-120. February 1968.
3. GOLDMAN : "Information Theory" Ed. Dover Publications, INC.
New York. 1968.
4. KAILATH, T. y Duttweiler D.: An RKHS Approach to Detection and
Estimation problems. Part III Generalised Innova-
tions Representations and a Libelihood-Ratio For-
mula, IEEE Transactions on I.T. Vol.IT-18, N^o 6
Nov. 1972.
5. KENDALL M. G. y STUART A.: "The Advanced Theory of Statistics"
Second Edition, Vol, 2 Hafner Publishing Company
New York, n, 4, 1967.
6. MIDDLETON D. y VAN METER: "On Optimun Multiple Alternative De-
tection of Signals in Noise"
Trans, Prof, Groups, Information Theory (IRE)
Vol, II-1, Sept, 1955, pag. 1.
7. MIDDLETON y VAN METER: "Detection and Extraction of Signals in
Noise from the Point of View of Statistical Deci-
sion Theory" J. Soc. Ind, Appl. Math. Vol. 3 N^o 4
pp. 192-253. 1955; vol. 4 N^o 2, pp. 86-119, 1956.
8. MIDDLETON y ESPOSITO-: "Simultaneous Optimun Detection and Esti-
mation of Signals in Noise. IEEE Translation on
Information Theory. Vol. IT-14 N^o 3, Mayo-1968.
9. OGGS F. : "A Note on Bayes Detection of Signals" IEEE Transla-
tion on I.T. vol.IT-10 N^o 1. January 1964 pp. 57
a 60.
10. SAGE A. MELSA J. : "Estimation Theory with applications to -
Communications and Control, 1971.
11. SPATARU A.: "Theorie de la Transmission de L'Information, To-
mo 2. Editura Technica Boucarets, 1968.
12. VAN TREES, H. : Detection, Estimation and Modulation Theory
Ed. Wiley, 1968.

CAPITULO IV

FUNCIONES DE PERDIDA ESPECIALES PARA DETECCION Y ESTIMACION SIMULTANEA

4.1. Funciones de pérdida cuadrática

4.1.1. Casos binarios especiales

Caso 1: Estimación pura

Caso 2: Estimación y detección conjunta

Caso 3: Detección y estimación binaria para

$$\vec{\theta}_1 \neq \vec{\theta} \quad \text{y} \quad \vec{\theta}_2 \neq \vec{\theta}$$

4.2. Función de pérdida simple

Bibliografía

FUNCIONES DE PERDIDA ESPECIALES PARA DETECCION Y ESTIMACION SIMULTANEA

En este capítulo proseguimos el estudio iniciado en el capítulo anterior, pero para obtener nuevos progresos en el análisis, es preciso introducir funciones de pérdida específicas y determinar $\hat{\theta}(\tilde{r})$ de acuerdo con la operación de minimización de - (3.27).

En primer lugar, consideraremos funciones de pérdida de tipo cuadrático pasando a continuación a considerar algunos ejemplos y casos particulares, viendo su equivalencia con los resultados obtenidos anteriormente por Middleton y Espósito, para el caso binario.

Finalmente consideramos la función de coste simple ⁽¹⁾ especializada para el caso que estamos estudiando y llegamos a obtener también la regla de decisión óptima.

4.1. FUNCIONES DE PERDIDA CUADRATICA

Seguimos considerando, como en el capítulo anterior, los subconjuntos del espacio paramétrico de señales disjuntos. Los costes usuales para las decisiones que tomamos en la detección, como indicamos anteriormente, se modifican teniendo en cuenta la influencia que tales decisiones tienen sobre si un estimador apa

rece al final del proceso de detección y estimación conjunta, o, por el contrario, no se presenta. La función de costo específica en esta subsección se corresponde con la expresada en (3.24) es decir,

$$C(\hat{\theta}, \gamma_i) = C_{ij}^{(1)} + C_{ij}^{(2)} (\hat{\theta} - \hat{\theta}(\hat{r}))^T D (\hat{\theta} - \hat{\theta}(\hat{r})) \quad (4.1)$$

para $\hat{\theta} \in \Omega_i$, $i, j = 0, 1, 2, \dots, M$, y donde D es una matriz definida positiva.

Los factores de costo $C_{ij}^{(1)}$ son los costos constantes asociados con la operación de detección pura; es decir, el costo de decidir que la hipótesis H_j es cierta cuando la que realmente lo es es la H_i . El factor de costo $C_{ii}^{(1)}$ está asociado con una decisión correcta y haremos la suposición de que los costos de detección correcta son nulos, es decir, $C_{ii}^{(1)} = 0$ para $i = 0, 1, 2, \dots, M$.

El segundo término de (4.1) es un coste adicional que refleja la interrelación entre las operaciones de detección y extracción; a continuación damos una discusión detallada del significado de los factores $C_{ij}^{(2)}$. Supongamos que $\hat{\theta}$ no es un parámetro ruidoso, es decir, que $\hat{\theta} = \theta$ implica que $\hat{r} = r$; entonces $\hat{\theta} \in \Omega_0$ implica que la hipótesis H_0 es cierta y si, además, elegimos γ_0 resulta que tomamos una decisión correcta y, en este caso, no precisamos ningún estimador a la salida del receptor. Por ello es razonable pensar que $C_{00}^{(2)} = 0$, lo que combinado con $C_{00}^{(1)} = 0$, nos indica que $C(\hat{\theta}, \gamma_0) = 0$ si $\hat{\theta} \in \Omega_0$. Pero si elegimos γ_i y es $\hat{\theta} \in \Omega_0$ ($i \neq 0$) resulta que hemos tomado una decisión incorrecta, no habiendo, sin embargo, estimación a la salida del receptor. Por último, queda el caso de decidir γ_j y que $\hat{\theta} \in \Omega_i$ ($i \neq j \neq 0$).

Para este caso se nos presenta el coste de detección incorrecta unido a $C_{ij}^{(2)}$. A continuación expresamos una relación entre estas últimas magnitudes que parece ser bastante razonable.

$$C_{0j}^{(2)} > C_{jj}^{(2)} > C_{j0}^{(2)} > C_{00}^{(2)} = 0 \quad (j \neq 0) \quad (4.2)$$

$$C_{ij}^{(2)} \geq C_{jj}^{(2)} \quad (i, j = 1, 2, \dots, M)$$

Teniendo en cuenta (4.2), la función de pérdida $C(\vec{\theta}, \gamma_j)$ para las diferentes combinaciones de los conjuntos paramétricos de señales Ω_j y decisiones γ_j son

$$\begin{aligned} C(\vec{0}, \gamma_0) &= 0 \\ C(\vec{0}, \gamma_j) &= C_{0j}^{(1)} + C_{0j}^{(2)} \hat{\theta}^T(\vec{r}) \hat{\theta}(\vec{r}) \quad j \neq 0 \\ C(\vec{\theta}, \gamma_j) &= C_{ij}^{(1)} + C_{ij}^{(2)} (\vec{\theta} - \hat{\theta}(\vec{r}))^T D (\vec{\theta} - \hat{\theta}(\vec{r})) \quad (i \neq 0) \end{aligned} \quad (4.3)$$

Nuestro propósito inmediato es determinar el estimador $\hat{\theta}(\vec{r})$ que minimiza la integral que aparece en (3.27), cuyo integrando contiene el factor característico de estos problemas $C_j(\vec{r}) B_j(\vec{r})$.

De las definiciones de $B_j(\vec{r})$ y $C_j(\vec{r})$ dadas en (3.21) y (3.22) respectivamente, vemos que para $j \neq 0$

$$\begin{aligned} C_j(\vec{r}) B_j(\vec{r}) &= C(\vec{0}, \gamma_j) - C(\vec{\theta}, \gamma_j) - \\ &= \frac{1}{P_0 P(r/\theta)} \sum_{i=1}^M P_i \int_{\Omega_i} (C(\vec{\theta}, \gamma_j) - C(\vec{s}, \gamma_j)) p(r/\vec{s}) P_i(\vec{s}) d\vec{s} \quad (4.4) \end{aligned}$$

Multiplicando ahora ambos miembros de (4.4) por $P_0 p(r/0)$ y sus

tituyendo los valores de $c(\hat{\theta}, \hat{r}_j)$ por los dados en (4.2), obtenemos

$$\begin{aligned}
 P_0 p(\hat{r}/\hat{0}) C_j(\hat{r}) B_j(\hat{r}) &= P_0 C_{0j}^{(1)} p(\hat{r}/\hat{0}) + \\
 + P_0 C_{0j}^{(2)} \hat{\theta}^T(\hat{r}) D \hat{\theta}(\hat{r}) p(\hat{r}/\hat{0}) &- \sum_{i=1}^M P_i (C_{i0}^{(1)} - C_{ij}^{(1)}) \int_{\Omega} p(\hat{r}/\hat{\theta}) P_i(\hat{\theta}) d\hat{\theta} \\
 - \sum_{i=1}^M P_i (C_{i0}^{(2)} - C_{ij}^{(2)}) \int_{\Omega} (\hat{\theta} - \hat{\theta}(\hat{r}))^T D (\hat{\theta} - \hat{\theta}(\hat{r})) &p(\hat{r}/\hat{\theta}) P_i(\hat{\theta}) d\hat{\theta}
 \end{aligned} \tag{4.5}$$

Integrando la última expresión sobre el espacio de observaciones \hat{r} y teniendo en cuenta que $p_0(\hat{\theta}) = \delta(\hat{\theta} - \hat{0})$ y $C_{00}^{(2)} = 0$, la integral de (3.27) queda

$$\begin{aligned}
 P_0 p(\hat{r}/\hat{0}) C_j(\hat{r}) B_j(\hat{r}) &= P_0 C_{0j}^{(1)} - \sum_{i=1}^M (C_{i0}^{(1)} - C_{ij}^{(1)}) P_i \\
 \int_{\Omega} p(\hat{r}/\hat{\theta}) P_i(\hat{\theta}) d\hat{\theta} &+ \int_{\Omega} \sum_{i=1}^M [(C_{ij}^{(2)} - C_{i0}^{(2)}) P_i \\
 \int_{\Omega} (\hat{\theta} - \hat{\theta}(\hat{r}))^T D (\hat{\theta} - \hat{\theta}(\hat{r})) &p(\hat{r}/\hat{\theta}) P_i(\hat{\theta}) d\hat{\theta} dr \tag{4.6}
 \end{aligned}$$

donde el factor -1 se ha absorbido en el último término como indica el orden inverso en los factores de costo. Los dos primeros términos de esta expresión no dependen del estimador $\hat{\theta}(\hat{r})$ y, por tanto, no intervienen en el proceso de minimización indicado en (3.27).

Como de la redacción (4.2) se desprende que $C_{ij}^{(2)} - C_{i0}^{(2)} > 0$ para $j \neq 0$, nuestro problema se reduce, entonces a calcular la función $\hat{\theta}(\hat{r})$ que minimiza el tercer término de (4.6). Sea

$$K(\hat{r}, \hat{\theta}(\hat{r})) = \sum_{i=1}^M (C_{ij}^{(2)} - C_{i0}^{(2)}) P_i \int_{\Omega} (\hat{\theta} - \hat{\theta}(\hat{r}))^T D (\hat{\theta} - \hat{\theta}(\hat{r})) p(\hat{r}/\hat{\theta}) P_i(\hat{\theta}) d\hat{\theta} \tag{4.7}$$

La función $K(\vec{r}, \hat{\theta}(\vec{r}))$ de (4.7) es positiva para todo valor de r y, por tanto, la función $\hat{\theta}(\vec{r})$ que minimice $K(\vec{r}, \hat{\theta}(\vec{r}))$ minimizará también a $\int K(\vec{r}, \hat{\theta}(\vec{r})) dr$. Por tanto una condición necesaria para que la función anterior tenga un mínimo es que exista un valor $\hat{\theta}(\vec{r})$ que satisfaga

$$\left. \nabla_{\hat{\theta}} K(\vec{r}, \hat{\theta}(\vec{r})) \right|_{\hat{\theta} = \hat{\theta}(\vec{r})} = 0 \quad (4.8)$$

donde $\nabla_{\hat{\theta}}$ es el operador gradiente, es decir, un vector columna de primeras derivadas parciales con respecto a la componente $\hat{\theta}(\vec{r})$. Si la matriz Hessiana, es decir, la matriz de derivadas parciales segundas con respecto a $\hat{\theta}(\vec{r})$ evaluadas en $\hat{\theta}(\vec{r})$ es definida positiva, entonces la función $\hat{\theta}(\vec{r})$ determinada por la condición (4.8) es la función que minimiza a $K(\vec{r}, \hat{\theta}(\vec{r}))$. Por tanto, si la matriz

$$\left. \nabla_{\hat{\theta}} (\nabla_{\hat{\theta}})^T K(\vec{r}, \hat{\theta}(\vec{r})) \right|_{\hat{\theta} = \hat{\theta}(\vec{r})} \quad (4.9)$$

es definida positiva, entonces la función determinada por (4.8) es una función que nos produce un mínimo. La sustitución de (4.7) en (4.9) nos proporciona

$$\left. \nabla_{\hat{\theta}} (\nabla_{\hat{\theta}})^T K(\vec{r}, \hat{\theta}(\vec{r})) \right|_{\hat{\theta} = \hat{\theta}(\vec{r})} = 2 \left[\sum_{j=1}^M (c_{1j}^{(2)} - c_{10}^{(2)}) p_1 \right] p(\vec{r}/\hat{\theta}) \underline{D}(\hat{\theta}) d\hat{\theta} \underline{D} \quad (4.10)$$

y como la matriz D es, por definición, definida positiva y el factor escalar que la multiplica también es positivo, resulta que la matriz Hessiana es definida positiva y, por tanto, la solución de (4.8) nos minimizará la expresión

$$\int K(\vec{r}, \hat{\theta}(\vec{r})) dr \quad (4.11)$$

Para obtener ahora el estimador óptimo $\hat{\theta}^*(\vec{r})$, sustituiremos (4.7) en (4.8). teniendo en cuenta que

$$\nabla_{\theta} (\hat{\theta} - \theta(\vec{r}))^T D (\hat{\theta} - \theta(\vec{r})) = -2D(\hat{\theta} - \theta(\vec{r})) \quad (4.12)$$

vemos que la solución buscada satisface la ecuación

$$\sum_{j=1}^m (c_{ij}^{(2)} - c_{i0}^{(2)}) p_i \int_{\Omega} D(\hat{\theta} - \theta_j(\vec{r})) p(\vec{r}/\theta) p_1(\theta) d\theta = 0 \quad (4.13)$$

y resolviéndola para $\theta_j(\vec{r})$ vemos que

$$\hat{\theta}_{j(\vec{r})}^{*} p_{j-1} = \frac{\sum_{i=1}^m (c_{ij}^{(2)} - c_{i0}^{(2)}) p_i \int_{\Omega} p(\vec{r}/\theta) p_1(\theta) d\theta}{\sum_{i=1}^m (c_{ij}^{(2)} - c_{i0}^{(2)}) p_i \int_{\Omega} p(\vec{r}/\theta) p_1(\theta) d\theta} \quad (4.14)$$

y recordando ahora que $p_0(\theta) = \delta(\theta - 0)$ y que $c_{00}^{(2)} = 0$, podemos escribir (4.14) en la forma

$$\hat{\theta}_{j(\vec{r})}^{*} p_{j-1} = \sum_{i=1}^m \frac{\frac{c_{ij}^{(2)} - c_{i0}^{(2)}}{c_{ij}^{(2)}} \Lambda_{i(\vec{r})}}{1 + \sum_{j=1}^m \left[\frac{c_{ij}^{(2)} - c_{i0}^{(2)}}{c_{0j}^{(2)}} \right] \Lambda_{i(\vec{r})}} \hat{\theta}_{i(\vec{r})}^{*} p_{i-1} \quad (4.15)$$

donde $\Lambda_{i(\vec{r})}$ es la razón media de verosimilitud y $\hat{\theta}_{i(\vec{r})}^{*} p_{i-1}$ es el estimador de mínimos cuadrados o estimador de mínima varianza para una señal de la clase Ω_i en presencia de certeza total; es decir

$$\Lambda_{i(\vec{r})} = \frac{p_i \int_{\Omega} p(\vec{r}/\theta) p_1(\theta) d\theta}{p_0 p(\vec{r}/\theta)} \quad (4.16)$$

y

$$\hat{\theta}_{i(\vec{r})}^{*} p_{i-1} = \frac{\int_{\Omega} p(\vec{r}/\theta) p_1(\theta) d\theta}{\int_{\Omega} p(\vec{r}/\theta) p_1(\theta) d\theta} \quad (4.17)$$

Si suponemos ahora que $c_{i0}^{(2)} = 0$ y $c_{ij}^{(2)} = c_{oj}^{(2)}$ 0 para $i \neq 0$, entonces (4.15) se reduce a la expresión

$$\hat{\theta}_j(\bar{r})_{p_j < 1} = \sum_{i=1}^m \left\{ \frac{\Lambda_i(\bar{r})}{1 + \sum_{i=1}^m \Lambda_i(\bar{r})} \hat{\theta}_i(\bar{r})_{p_i = 1} \right\} \quad (4.18)$$

que se puede expresar también de la siguiente forma

$$\hat{\theta}_j(\bar{r})_{p_j < 1} = \sum_{i=1}^m p(H_i / \bar{r}) \hat{\theta}_i(\bar{r})_{p_i = 1} \quad (4.19)$$

donde $p(H_i / \bar{r})$ es la probabilidad a posteriori de que sea cierta la hipótesis H_i . Se puede demostrar fácilmente el paso de la expresión (4.18) a (4.19) sin más que tener presente el teorema de Bayes y la definición, por el convenio que hicimos al principio, de ser $p_i = p(H_i)$. Tenemos

$$p(\bar{r}, \bar{\theta}, H_i) = p(\bar{r}, \bar{\theta} / H_i) p_i = p(\bar{r} / \bar{\theta}, H_i) p(\bar{\theta} / H_i) p_i \quad (4.20)$$

Dada la hipótesis H_i , entonces $\theta \in \Omega_i$ y podemos restringir el rango de integración al subconjunto Ω_i de Ω y de aquí

$$\begin{aligned} p_i p(\bar{r} / H_i) &= p_i \int_{\Omega} p(\bar{r}, \bar{\theta} / H_i) d\bar{\theta} = p_i \int_{\Omega_i} p(\bar{r} / \bar{\theta}, H_i) p_i(\bar{\theta}) d\bar{\theta} = \\ &= p_i \int_{\Omega_i} p(\bar{r} / \bar{\theta}) p_i(\bar{\theta}) d\bar{\theta} \end{aligned} \quad (4.21)$$

y teniendo en cuenta ahora que $p_{\Omega}(\bar{\theta}) = \int (\delta - \bar{\theta})$ podemos escribir

$$\begin{aligned} p(\bar{r}) &= \sum_{i=1}^m p(\bar{r} / H_i) p(H_i) = \sum_{i=1}^m p_i \int_{\Omega_i} p(\bar{r} / \bar{\theta}) p_i(\bar{\theta}) d\bar{\theta} = \\ &= p_0 p(\bar{r} / 0) + \sum_{i=1}^m p_i \int_{\Omega_i} p(\bar{r} / \bar{\theta}) p_i(\bar{\theta}) d\bar{\theta} \end{aligned} \quad (4.23)$$

y aplicando el teorema de Bayes

$$p(H_i / \bar{r}) = \frac{p(\bar{r} / H_i) p(H_i)}{p(\bar{r})} =$$

$$= \frac{p_1 \int_{\Omega} p(\vec{r}/\theta) p_1(\theta) d\theta}{p_0 p(\vec{r}/0) + \sum_{j \neq 0} p_1 \int_{\Omega} p(\vec{r}/\theta) p_1(\theta) d\theta} = \frac{\Lambda_1(\vec{r})}{1 + \sum_{j \neq 0} \Lambda_1(\vec{r})} \quad (4.24)$$

que es evidente por la propia definición (4.16) de $\Lambda_1(\vec{r})$.

Hemos determinado, pues, la forma de los estimadores para el caso que estamos estudiando y que son (4.15) y (4.18). El estimador dado por (4.18) no es tan complicado como el que nos proporciona (4.15) y, además, (4.16) nos proporciona una interpretación bastante razonable sobre él, al presentarlo como una suma ponderada de estimadores de mínimos cuadrados, para el caso en que no exista incertidumbre, con respecto a varias hipótesis, siendo los pesos las diferentes probabilidades a posteriori de las hipótesis respectivas. Por la simetría existente, el estimador que nos proporciona (4.18) es independiente de j , por lo que cualquier decisión X_j ($j \neq 0$) que tomemos nos proporcionará el mismo estimador a la salida de la recepción y, por tanto, podemos eliminar el subíndice j de $\hat{\theta}_j(\vec{r})$. No ocurre así con el estimador que nos proporciona (4.15) aunque tiene como ventaja que las suposiciones acerca del mismo son menos restrictivas. Se puede probar fácilmente por ejemplo que si $c_{0j}^{(2)} \neq c_{0k}^{(2)}$, entonces $\hat{\theta}_j(\vec{r})_{p_j < 1}$ diferirá de $\hat{\theta}_k(\vec{r})_{p_k < 1}$ si $j \neq k$ y $j, k \neq 0$. Por lo tanto es una ventaja importante poder elegir los costes de forma que nos conduzcan a un estimador tal como el que nos proporciona (4.18).

Después de determinar la forma del estimador óptimo para este caso, que es (4.15), o (4.18) la etapa final consistirá en la determinación de la regla de decisión óptima al sustituir los

estimadores obtenidos como hemos indicado anteriormente en la regla de decisión calculada en la primera etapa y que escribimos a continuación un poco modificada, después de indicar que $C_j(\vec{r})$ para todo $\vec{r} \in \Gamma$ y $j \neq 0$ e incluimos la dependencia de B_j y C_j con el estimador $\hat{\theta}_j(\vec{r})$.

Decidir γ_j ó $p(\gamma_j / \vec{r}) = 1$ si

$$a) C_j(\vec{r}, \hat{\theta}_j(\vec{r})) B_j(\vec{r}, \hat{\theta}_j(\vec{r})) < 0$$

$$b) C_j(\vec{r}, \hat{\theta}_j(\vec{r})) B_j(\vec{r}, \hat{\theta}_j(\vec{r})) < C_k(\vec{r}, \hat{\theta}_k(\vec{r})) B_k(\vec{r}, \hat{\theta}_k(\vec{r})) \quad \forall k$$

De la regla de detección anterior observamos que de entre todos los factores que intervienen en la determinación explícita del detector óptimo el más fundamental es $C_j(\vec{r}, \hat{\theta}_j(\vec{r})) B_j(\vec{r}, \hat{\theta}_j(\vec{r}))$. Calculamos a continuación este factor para el caso en que la asignación de costes sea tal que obtengamos el estimador dado por (4.18), suponiendo, además, que $C_{1j}^{(2)} \nabla 1$.

Sea $\hat{\theta}_j(\vec{r})$ el estimador que nos proporciona (4.18); bajo las hipótesis anteriores y teniendo en cuenta (3.21) y (3.22) obtenemos que

$$C_j(\vec{r}, \hat{\theta}_j(\vec{r})) B_j(\vec{r}, \hat{\theta}_j(\vec{r})) = (C_{0j}^{(1)} + (\hat{\theta}_j(\vec{r}))^T D \hat{\theta}_j(\vec{r}))$$

$$\left[1 - \sum_{i=1}^M \frac{p_1 \int_{\Omega} (C_{i0}^{(1)} - C_{ij}^{(1)} - (\hat{\theta}_j - \hat{\theta}_j)^T D (\hat{\theta}_j - \hat{\theta}_j)) p(\vec{r}/\theta) p_1(\theta) d\theta}{(C_{0j}^{(1)} + (\hat{\theta}_j)^T D \hat{\theta}_j) p_0 p(\vec{r}/\theta)} \right]$$

$$C_{0j}^{(1)} + (\hat{\theta}_j)^T D \hat{\theta}_j + \sum_{i=1}^M (C_{ij}^{(1)} - C_{i0}^{(1)}) \Lambda_i(\vec{r}) + (\hat{\theta}_j)^T D \sum_{i=1}^M \Lambda_i(\vec{r}) + \sum_{i=1}^M \Lambda_i(\vec{r}) \frac{\int_{\Omega} \hat{\theta}_j^T D \hat{\theta}_j p(\vec{r}/\theta) p_1(\theta) d\theta}{\int_{\Omega} p(\vec{r}/\theta) p_1(\theta) d\theta} \quad (4.25)$$

y si notamos

$$(\hat{\theta}^T D \hat{\theta})_1(\bar{r})_{P_1=1} = \frac{\int_{\theta_0}^{\theta_1} \theta^T D \hat{\theta} p(r/\theta) p_1(\theta) d\theta}{\int_{\theta_0}^{\theta_1} p(r/\theta) p_1(\theta) d\theta} \quad (4.26)$$

y tenemos en cuenta (4.19), podemos escribir (4.25) en la forma

$$\begin{aligned} C_j(\bar{r}, \hat{\theta}_j) B_j(\bar{r}, \hat{\theta}_j) &= C_{0j}^{(1)} + \sum_{i=1}^m (C_{ij}^{(1)} - C_{i0}^{(1)}) \Lambda_i(\bar{r}) + \\ &- (1 + \sum_{i=1}^m \Lambda_i(\bar{r})) \sum_{k=1}^m \sum_{l=1}^m p(H_k/\bar{r}) p(H_l/\bar{r}) \hat{X}_k(\bar{r}) \hat{X}_l(\bar{r})^T D(\hat{\theta}_i(\bar{r})) + \\ &+ \Lambda_i(\bar{r}) (\hat{\theta}^T D \hat{\theta})_1(\bar{r})_{P_1=1} \end{aligned} \quad (4.27)$$

4.1.1. Casos binarios especiales

Veremos a continuación tres casos particulares para los cuales alcanzamos los mismos resultados que se han obtenido en otros trabajos de detección y estimación simultánea para dos hipótesis.

Caso 1.- Estimación pura

Aquí nos interesamos únicamente por la operación de estimación; los costes de detección $C_{ij}^{(1)}$ expresados anteriormente en (3.32), serán nulos. Tenemos nuevamente $i, j = 0, 1$ pues estamos interesados en el caso binario, es decir, suponemos que en el intervalo de observación se distingue ruido solamente o bien señal mezclada con ruido. Si suponemos que la función de pérdida tiene la forma

$$c(\hat{\theta}, \theta(\bar{r})) = C_{ij}^{(2)} (\hat{\theta} - \theta)^T D (\hat{\theta} - \theta) \quad (4.28)$$

y que $C_{00}^{(2)} = 0$, entonces (4.15) nos proporciona como forma del estimador la siguiente

$$\hat{\theta}_1(\bar{r})_{P_1 < 1} = \frac{\left[\frac{c_{11}^{(2)} - c_{10}^{(2)}}{c_{01}^{(2)}} \right] \Lambda_1(\bar{r})}{1 + \left[\frac{c_{11}^{(2)} - c_{10}^{(2)}}{c_{01}^{(2)}} \right] \Lambda_1(\bar{r})} \hat{\theta}_1(\bar{r})_{P_1 = 1} \quad (4.29)$$

por tanto, si suponemos que $c_{10}^{(2)} = 0$ y $c_{11}^{(2)} = c_{01}^{(2)}$, entonces la expresión (4.29) se reduce a

$$\hat{\theta}_1(\bar{r})_{P_1 < 1} = \frac{\Lambda(\bar{r})}{1 + \Lambda(\bar{r})} \hat{\theta}_1(\bar{r})_{P_1 = 1} = P(H_1 / \bar{r}) \hat{\theta}_1(\bar{r})_{P_1 = 1} \quad (4.30)$$

que es idéntico al resultado ya obtenido por Middleton y Espósito

Caso 2. - Estimación y detección conjunta

Ahora supondremos que existe interrelación entre las operaciones de detección y estimación. Suponemos que el coste asociado a la decisión γ , cuando $\hat{\theta} \in \Omega$, ($i, j = 0, 1$) tiene la siguiente forma

$$c(\hat{\theta}, \gamma) = c_{ij}^{(2)} + c_{ij}^{(2)} (\hat{\theta} - \hat{\theta}(\bar{r}))^T D (\hat{\theta} - \hat{\theta}(\bar{r})) \quad (4.31)$$

así como que $c(\hat{\theta} = \bar{\theta}, \gamma) = 0$; entonces, según las suposiciones que hagamos acerca de los factores de coste $c_{ij}^{(2)}$ obtendremos, - bien (4.15), bien (4.18) como formas del estimador. Si, además, suponemos que $c_{10}^{(2)} = 0$ y que $c_{11}^{(2)} = c_{01}^{(2)} = C$, entonces la forma del estimador viene dada por (4.18) y la estructura del detector vendrá dada por (4.27) que, para este caso, es

$$c_1(\bar{r}) \hat{\theta}_1(\bar{r})_{P_1 < 1} B_1(\bar{r}, \hat{\theta}_1(\bar{r})_{P_1 < 1}) = C - C \Lambda(\bar{r}) + (\hat{S}^T D \hat{S})_1(\bar{r})_{P_1 = 1} \\ - \frac{\Lambda^2(\bar{r})}{1 + \Lambda(\bar{r})} (\hat{S}_1(\bar{r})_{P_1 < 1})^T D (\hat{S}_1(\bar{r})_{P_1 < 1})$$

Caso 3. - Detección y estimación binaria para $\theta_1 \neq 0$ y $\theta_2 = 0$

Aquí excluimos el caso en que no se presente señal, y, por tanto, p_0 ha de ser cero en todos nuestros resultados anteriores como por ejemplo en las relaciones (4.15) y (4.18). Como la hipótesis nula se excluye, es $C_{10}^{(2)} = 0$ para $i = 1, 2$. Supondremos que $i, j = 1, 2$ y la función costo para este ejemplo será de la forma

$$C(\hat{\theta}, \theta) = C_{ij}^{(1)} + C_{ij}^{(2)} (\theta - \hat{\theta}_j(\bar{r}))^T D (\theta - \hat{\theta}_j(\bar{r})) \quad (4.33)$$

la estructura del estimador para este tercer caso se obtiene directamente al sustituir (4.33) en (3.46), o bien aplicando las restricciones expresadas anteriormente a (4.15). En este caso la forma del estimador viene dada por

$$\begin{aligned} \hat{\theta}_j(\bar{r})_{p_j=1} &= \sum_{i=1}^2 \frac{C_{ij}^{(2)}}{\sum_{k=1}^2 C_{kj}^{(2)} \Lambda_{ki}(\bar{r})} \hat{\theta}_i(\bar{r})_{p_i=1} \\ &= \sum_{i=1}^2 \frac{1}{\sum_{k=1}^2 \frac{C_{kj}^{(2)}}{C_{ij}^{(2)}} \Lambda_{ki}(\bar{r})} \hat{\theta}_i(\bar{r})_{p_i=1} \quad (4.34) \end{aligned}$$

donde $\hat{\theta}_i(\bar{r})_{p_i=1}$ está definido por (4.17) y $\Lambda_{ki}(\bar{r})$ viene dado por

$$\Lambda_{ki}(\bar{r}) = \frac{p_k \int_a^b p(\bar{r}/\theta) p_k(\theta) d\theta}{p_i \int_a^b p(\bar{r}/\theta) p_i(\theta) d\theta} \quad (4.35)$$

Por la definición (4.35) es muy fácil demostrar que $\Lambda_{ii}(\bar{r}) = 1$, así como $\Lambda_{ik}(\bar{r}) \Lambda_{ki}(\bar{r}) = 1$. Si suponemos ahora que $C_{kj}^{(2)} = C_{ij}^{(2)}$, para todos los valores de i y de k , entonces (4.34) se reduce a

$$\hat{\theta}_j(r)_{p_j=1} = \frac{1}{\sum_{k=1}^L k_i(r)} \hat{\theta}_i(r)_{p_i=1} \quad (4.36)$$

que se puede expresar también como

$$\hat{\theta}_j(r)_{p_j=1} = \frac{p(H_1/r)}{\sum_{k=1}^L k_i(r)} \hat{\theta}_i(r)_{p_i=1} \quad (4.37)$$

ya que $p(H_1/r) = (\sum_{k=1}^L k_i(r))$, como se puede demostrar fácilmente haciendo $p_0 = 0$ en la relación (4.24). La estructura del detector en este caso viene determinada por el factor $B_{12}(r)$ que podemos obtener de (3.40). Si sustituimos en esa expresión los estimadores obtenidos $\hat{\theta}_1(r)_{p_1=1}$ y $\hat{\theta}_2(r)_{p_2=1}$ y llamamos $B_{12}(r)$ al valor resultante, la regla de decisión para este caso será la siguiente

$$p(H_1/r) = 1 \quad \text{si } B_{12}(r) < 0$$

$$p(H_1/r) = 0 \quad \text{si } B_{12}(r) \geq 0$$

4.2. FUNCION DE PERDIDA SIMPLE

Hemos usado anteriormente una función de pérdida cuadrática para obtener nuevos resultados en estimación y detección conjunta bajo hipótesis múltiples. Otro caso importante, que se emplea a menudo en la estimación, es la función de pérdida simple y la razón de esta importancia es que conduce a estimadores incondicionales de máxima verosimilitud (1). Esta subsección la dedicamos a la detección y estimación de señales bajo hipótesis múl-

(1) Algunos autores (Middleton, 1955; Hancock y Wintz, 1966) llaman estimador condicional de máxima verosimilitud al que maximiza la función de verosimilitud; al estimador que maximiza la ley de probabilidad conjunta o la a posteriori, le llaman estimador incondicional de máxima verosimilitud (Middleton, 1955; Hancock y Wintz, 1966), o estimadores bayesianos de máxima verosimilitud (Lee 1964; Ho y Lee, 1964; Sage, 1968).

tiples cuando se supone como función de pérdida la que viene dada por la siguiente expresión

$$c(\hat{\theta}, \theta_j) = c_{ij}^{(1)} + c_i^{(2)} (A_i^{(2)} - J(\hat{\theta} - \theta_j)) \quad (\theta \in \Omega_j) \quad (4.38)$$

donde con $J(\hat{\theta} - \theta_j)$ hemos expresado

$$J(\hat{\theta}(\bar{r}) - \theta_j) = \sum_k J(\hat{\theta}_k(\bar{r}) - \theta_k) \quad (4.39)$$

siendo, en esta última relación, $J(\cdot)$ la función de Dirac, $\hat{\theta}_k(\bar{r})$ es la k -ésima componente de $\hat{\theta}(\bar{r})$, y $\hat{\theta}(\bar{r})$ y θ_j son vectores de n componentes. Los factores $c_{ij}^{(1)}$ son los costes constantes asociados al proceso de decisión, es decir, decidir que una señal de la clase Ω_j se encuentra durante el período de observación, cuando la señal que realmente está presente es del conjunto Ω_i . Los factores $c_i^{(2)}$ y $A_i^{(2)}$ están relacionados íntimamente con la estimación correcta o incorrecta⁽¹⁾.

Sustituyendo ahora (4.39) en la relación (3.15) para el riesgo total conjunto $R_D + B$, resulta que

$$\begin{aligned} R_D + B = & \int \sum_i (c_{i0}^{(1)} + c_i^{(2)} A_i^{(2)}) p_i \int p(r/\theta) p_i(\theta) d\theta dr + \\ & + \int \sum_j p(r/\bar{r}) \left[\sum_i (c_{ij}^{(1)} - c_{i0}^{(1)}) p_i \right] p(r/\theta) p_i(\theta) d\theta dr + \\ & + \int [c_0^{(2)} p_0 p(r/\theta) K(\theta) + \sum_i c_i^{(2)} p_i p(r/\theta) p_i(\theta)] dr \end{aligned} \quad (4.40)$$

El primer término de (4.40) es una constante positiva independiente de la regla de decisión y del estimador; el segundo término depende únicamente de la regla de decisión, pero no del

(1) Una discusión acerca de este tipo de costes puede encontrarse en el Cap. 23 de Middleton: "An introduction to Statistical Communication Theory" McGraw-Hill Book Company, New York. 1960.

estimador, mientras que el tercero depende del estimador y no de la regla de decisión. El integrando del tercer término siempre es positivo y, por tanto, la integral también será positiva. Al venir la integral referida por un signo negativo, tendremos que maximizarla si queremos minimizar el riesgo total. De esta forma elegiremos $\hat{\theta}(\hat{r})$ de la siguiente forma

$$\max_{\hat{\theta}(\hat{r})} \int_{\mathcal{R}} [C_0^{(2)} p_0 p(\hat{r}/0) J(\hat{\theta}) + \sum_{i=1}^M C_i^{(2)} p_i (p(\hat{r}/i) p_i(\hat{\theta}))] d\hat{r} \quad (4.41)$$

El estimador $\hat{\theta}(\hat{r})$ no es un estimador incondicional de máxima verosimilitud, pero se puede considerar como una versión generalizada de la estimación de máxima verosimilitud [3]. Si los factores $C_i^{(2)}$ son iguales a $C_0^{(2)}$, entonces (4.41) se convierte en

$$\max_{\hat{\theta}(\hat{r})} \int_{\mathcal{R}} [p_0 p(\hat{r}/0) J(\hat{\theta}) + \sum_{i=1}^M p_i p(\hat{r}/i) p_i(\hat{\theta})] d\hat{r} \quad (4.42)$$

que para el caso binario de señal mezclada con ruido o ruido solamente, $M = 1$ y (4.42) se reduce a

$$\max_{\hat{\theta}(\hat{r})} \int_{\mathcal{R}} [p_0 p(\hat{r}/0) J(\hat{\theta}) + p_1 p(\hat{r}/1) p_1(\hat{\theta})] d\hat{r} \quad (4.43)$$

que es equivalente a la relación (3.15) obtenida por Middleton y Espósito [1].

La regla de decisión óptima en este caso viene dada por

Decidir i_j ó $p(i_j/r) = 1$ si

a) $D_j(r) < 0$

b) $D_j(r) \leq D_k(r) \quad k = 1, 2, \dots, M.$

y en otro caso elegir i_0 , donde

$$D_j(r) = \sum_{i=0}^M (C_{ij}^{(1)} - C_{i0}^{(1)}) p_i \int_{\mathcal{R}} p(\hat{r}/i) p_i(\hat{\theta}) d\hat{\theta} \quad (4.44)$$

BIBLIOGRAFIA

1. MIDDLETON-VAN METER: "On optimum Multiple Alternative Detection of Signals in Noise" Trans. Prof. Group, Information Theory (IRE) Vol.II-I, sept. 1955, pg. 1.
2. MIDDLETON-VAN METER: "Detection and Extraction of Signals in noise from the Point of View of Statistical Decision Theory" J.Soc.Ind. Appl. Math. Vol. 3 N° 4 pp. 192-253, 1955; Vol. 4 N° 2 pp. 86-119, 1956.
3. MIDDLETON Y ESPOSITO: "Simultaneous Optimum Detection and Estimation of Signals in Noise, IEEE Translation on Information Theory. Vol. IT-14 n° 3, Mayo, 1968
4. THOMAS: "An Introduction to Statistical Communication Theory" Wiley, 1969.
5. VAN TREES: "Detection, Estimation and Modulation Theory" Ed. Wiley, 1968.

ooo 0 ooo

UNIVERSIDAD DE SEVILLA
FACULTAD DE CIENCIAS

Reunido el Tribunal integrado por los abajo firmantes en el día de la fecha, para juzgar la Tesis Doctoral

D. Luis Perse Guipón
titulada "Algunas cuestiones sobre selección y extinción de señales en tesis de la información"

Se acordó otorgarle la calificación de Sobresaliente
summa laude

Sevilla, 23 de Septiembre de 1976

El Vocal,

[Signature]

El Vocal,

[Signature]

El Vocal,

[Signature]

El Presidente,

[Signature]

El Secretario,

[Signature]

El Doctor

[Signature]



* 5 0 1 1 4 4 6 3 9 *

FMA C 043/269