TRABAJOS DE ESTADISTICA Y
DE INVESTIGACION OPERATIVA
Vol. 34. Núm. 1, 1983, pp. 21 a 42

PROBLEMAS DE OPTIMO QUE RELACIONAN LA INFORMACION DE KULLBACK Y EL CONJUNTO DE RIESGOS DE NEYMAN - PEARSON (*)

Ramiro Melendreras Gimeno
Dpto. de Estadística Matemática
Universidad de Granada

RESUMEN

Consideramos la conexión que existe entre la información de Kullback y los tests admisibles óptimos en el conjunto de riesgos de Neyman - Pearson, usando para ello el estudio de problemas de programación matemática de tipo infinito. Se obtienen resultados que caracterizan un subconjunto de soluciones Bayes como consecuencia del conocimiento de la información, así como una medida de discriminación entre hipótesis para el conjunto de riesgos.

INTRODUCCION

Gran parte de los problemas de óptimo que se plantean en la Estadística Matemática sugieren la aplicación de las modernas técnicas de la Programación Matemática. Las restricciones de linealidad y convexidad, que son las más estudiadas, suelen verificarse de forma natural debido a las propiedades de los conjuntos de medidas, riesgos, etc. ...

Son muchos los autores que han tratado de relacionar el contraste de hipótesis con la información de Kullback, unos basándose en la estimación de los multiplicadores de Lagrange, otros estudiando los problemas de tipo asintótico. En éste artículo hemos evitado éstas corrientes, nuestro objetivo ha sido la búsqueda de conexiones entre los problemas de optimización que surgen del modelo de hipótesis simple

(*) Recibido Febrero, 1981

contra alternativa simple, en su versión clásica y Bayes, con los de mínima información.

Una consecuencia fundamental de este artículo es la dependencia que existe entre la elección de una distribución a priori para la obtención de la regla de Bayes óptima y el concepto de información de Kullback, pues se da una medida de discriminación entre las hipótesis en términos de una diferencia de medias asociadas a las distribuciones a priori.

Hemos supuesto que las pérdidas sobre las hipótesis son la unidad, el motivo no es otro que evitar la complejidad de notación y cálculo. Si se usan otras la generalización es trivial, sin más que considerar los cambios de escala en los errores de tipo uno y dos, así como la normalización de las funciones de distribución que se obtienen de ellos.

INFORMACION OPTIMA

Supongamos que $F(x)$ y $G(x)$ son dos elementos fijos de un conjunto convexo y homogéneo de medidas de probabilidad M, dominado por un medida μ, sobre un espacio medible (Ω, A). Sea c un número real. Nuestro problema inicial consistirá en hallar la medida en probabilidad $\Delta(x)$ más próxima a $G(x)$ en el sentido de la información de Kullback, sujeta a la restricción de que la diferencia entre la información de $\Delta(x)$ a $G(x)$ y la información de $\Delta(x)$ a $F(x)$ sea el valor c.

Sea P el conjunto de las derivadas de Radon - Nikodym respecto de μ de la familia M, el problema de optimización considerado es *hallar un $\delta \in P$, si existe, tal que*

$$\inf I(\delta : g)$$

sujeto a la restricción de que

$$\delta(x) \in \Gamma(c) = \{\delta(x) \in P : \Delta(x) \in M, I(\delta : g) - I(\delta : f) = c\}$$

donde f, g y δ son las derivadas de Radon - Nikodym respecto de μ de F, G y Δ respectivamente; c es inicialmente una constante, que admitirá una excelente interpretación como parámetro.

Este problema fue inicialmente planteado por Kullback [8] para relacionar la teoría de la información con el concepto de razón de verosimilitud.

Nuestro problema de programación matemática es de tipo convexo e infinito, ya que $I(\delta : g)$ es una función convexa en δ y las restricciones son lineales en δ, luego es posible aplicar la teoría general de los problemas lagrangianos en la versión infinita de Hurwicz [5] e Isii [6] al problema primal

$$\inf \int_\Omega \log(\delta(x)/g(x))\, \delta(x)\, d\mu(x)$$

condicionado a que

$$\delta(x) \in \Gamma(c) = \{\delta(x) \in P : \int_\Omega \log(f(x)/g(x))\, \delta(x)\, d\mu(x) =$$

$$= c, \int_\Omega \delta(x)\, d\mu(x) = 1\}$$

siempre que el punto c sea interior al conjunto

$$\{ \int_\Omega \log(f(x)/g(x)) . d\Delta(x) : \Delta \in M\}$$

El *teorema de la dualidad* nos asegura que:

$$\inf_{\delta \in \Gamma(c)} \int_\Omega \log(\delta(x)/g(x))\, \delta(x)\, d\mu(x) =$$

$$= \sup_{(u,v)} \{u c + v + \inf_{\delta \in P} \int_\Omega (\log(\delta(x)/g(x)) -$$

$$- u \log(f(x)/g(x)) - v)\, \delta(x)\, d\mu(x)\}$$

La solución de

$$\inf_{\delta \in P} \int_{\Omega} (\log (\delta (x)/g(x)) - u \log (f(x)/g(x)) - v)\, \delta (x)\, d\mu(x)$$

nos viene dada por

$$\delta^*(x) = g(x) \exp \{u \log (f(x)/g(x)) + v - 1\}$$

donde u y v son los multiplicadores de Lagrange, sin restricción de signo, que dependen de c. Estos multiplicadores pueden determinarse obligando a la solución anterior a que verifique con el conjunto de restricciones. Como $\delta^*(x)$ debe ser función de densidad, tendrá que cumplir

$$\int_{\Omega} g(x) \exp \{u \log (f(x)/g(x)) + v - 1\}\, d\mu(x) = 1$$

de donde se deduce que

$$v = 1 - \log \rho (u)$$

siendo

$$\rho(u) = \int_{\Omega} (f(x))^u \, (g(x))^{1-u} \, d\mu(x)$$

por tanto tendremos que

$$\delta_u^*(x) = (f(x))^u \, (g(x))^{1-u}/\rho(u)$$

es la función de densidad solución a nuestro problema, dependiendo del multiplicador u.

El problema de dualidad queda reducido a

(1/2, 1/2) por las propiedades de los tests aleatorizados, así como que su interior es no vacio pues F y G no son idénticas, pues si lo fuesen la restricción $I(\delta : g) - I(\delta : f) = c$ no tendría sentido para c distinto de cero. Si como hemos visto $S(\Phi)$ no puede coincidir con D_0, igual sucede con D_1, debido a que F y G pertenecen a un conjunto homogéneo de medidas y por tanto los únicos puntos comunes de $S(\Phi)$ y D_1 son (1, 0) y (0, 1).

Representemos por $\partial S(\Phi)$ *la frontera inferior* de $S(\Phi)$ por ser éste cerrado, convexo y con interior no vacío, existe una recta soporte a él en todo punto de $\partial S(\Phi)$ tal que

$$l_1\,\alpha + l_2\,\beta \leqslant l_1\,y_1 + l_2\,y_2$$

para todo $(y_1, y_2) \in S(\Phi)$ y con $(\alpha, \beta) \in \partial S(\Phi)$. Las componentes del vector $(l_1, l_2) \in (R^2)^*$ son del mismo signo y además no pueden ser las dos nulas debido al teorema de separación de conjuntos convexos.

Sea $h = l_1/l_2$ el número real que representa el opuesto de la pendiente de la recta soporte a $S(\Phi)$ en (α, β) y que toma valores de cero a infinito.

Tomemos la función contínua y creciente

$$h = \exp c(u)$$

que admite como valores extremos a

$$h_0 = \exp c(0) = \exp\{-I(g:f)\} \quad \text{y} \quad h_1 = \exp c(1) = \exp\{I(f:g)\}$$

Definimos los conjuntos de la forma:

$$\partial S(u) = \{(\alpha, \beta) : h(u)\,\alpha + \beta = \inf\{h(u)\,y_1 + y_2 : (y_1, y_2) \in S(\Phi)\}\} \cap S(\Phi)$$

entonces llamamos *conjunto de soluciones admisibles Kullback* a

$$S(I) = \{(\alpha, \beta) \in \partial S(\Phi) : \exists\, u \in [0, 1] \Rightarrow (\alpha, \beta) \in \partial S(u)\}$$

Sean los $(l_1, l_2) \in (R^2)^*$ con $l_1/l_2 \geqslant 0$, tales que

$$\nu'(u) = -c(u)$$

$$\nu''(u) = -\mathrm{Var}\ \{(\log (f(x)/g(x)))^2/\delta_u^*\}$$

Consideremos ahora la variación del parámetro c. Si suponemos el problema de hallar $\delta \in P$, si existe, tal que

$$\inf I(\delta : g)$$

sujeto a la restricción de que

$$\delta(x) \in \Gamma(c) = \{\delta(x) \in P : \Delta(x) \in M, I(\delta : f) - I(\delta : g) = c\}$$

Chernoff [2] demuestra que su solución guarda una simetria con el problema original respecto a un punto u_0 del intervalo (0, 1) que puede ser determinado en nuestro caso haciendo $c = 0$, lo que implica

$$I(\delta_{u_0}^* : g) = I(\delta_{u_0}^* : f)$$

éste punto es fundamental en los trabajos del mencionado autor, pues en él basa su *función de eficiencia relativa*. Trataremos de caracterizar todos los valores de u, dando otro significado a u_0 que consideramos de igual relevancia.

Debido a la expresión $\rho'(u)/\rho(u) = c$ y la convexidad de $\rho(u)$, cuando c varía desde $-I(g : f)$ a $I(f : g)$ entonces u tiene una variación estrictamente monótona y contínua de cero a uno (Kullback [8]), esta monotonía estricta nos asegura que la variación de u en [0, 1] nos induce una variación de c en el intervalo $[-I(g : f), I(f : g)]$.

SOLUCIONES ADMISIBLES KULLBACK

Como consecuencia de la correspondencia biunívoca que existe entre c y u, expuesta en el último párrafo del apartado anterior, se establece en el espacio muestral Ω una partición en función de u al considerar la restricción

$$\int_\Omega (\log (f(x)/g(x)) - c(u))\, \delta_u^*(x)\, d\mu(x) = 0$$

c.s. en μ, implicando la existencia de conjuntos de la forma

$$A^{<}(u) = \{x \in \Omega : \log\,(f(x)/g(x)) < c(u)\}$$

$$A^{=}(u) = \{x \in \Omega : \log\,(f(x)/g(x)) = c(u)\}$$

$$A^{>}(u) = \{x \in \Omega : \log\,(f(x)/g(x)) > c(u)\}$$

donde el signo que aparece en el superíndice indica la desigualdad que existe entre $\log(f(x)/g(x))$ y $c(u)$. La construcción de éstos conjuntos ha de permitir una interpretación estadística y geométrica de $c(u)$ para cada valor de u, basándonos en la teoría de Neyman - Pearson (Ferguson [4]).

Sea (y_1, y_2) un punto arbitrario de R^2 y consideremos los conjuntos

$$D_0 = \{(y_1, y_2) : y_1 + y_2 = 1,\; y_1 \geqslant 0, y_2 \geqslant 0\}$$

$$D_1 = \{(y_1, y_2) : 0 \leqslant y_1 \leqslant 1,\; 0 \leqslant y_2 \leqslant 1\}$$

De otra parte las medidas F y G, así como el conjunto de tests aleatorizados Φ, inducidos por el problema de contraste de hipótesis simple contra alternativa simple

$$H_0 : G \quad \text{frente a} \quad H_1 : F$$

definen el conjunto

$$S(\Phi) = \left\{ \int_\Omega \phi(x)\, dG(x),\; \int_\Omega (1 - \phi(x))\, dF(x) \quad : \phi \in \Phi \right\}$$

donde las coordenadas de los puntos de $S(\Phi)$ expresan el error de tipo uno y dos respectivamente. Se verifica de forma trivial que

$$D_0 \subset S(\Phi) \subset D_1$$

además por el teorema de Liapounov [10], $S(\Phi)$ es cerrado, acotado y convexo; de otra parte sabemos que es simétrico con respecto al punto

$$\inf_{\delta \in \Gamma(c)} \int_{\Omega} \log(\delta(x)/g(x))\, \delta(x)\, d\mu(x) = \sup_{u} \{u\, c - \log \rho(u)\}$$

Para el valor de u tal que

$$\rho'(u)/\rho(u) = c$$

se obtiene el óptimo de nuestro programa, pues

$$\frac{d^2}{du^2} \{u\, c - \log \rho(u)\} = - \operatorname{Var} \{(\log(f(x)/g(x)))^2 / \delta_u^*\} < 0$$

Debido a que nuestra intención es parametrizar el problema de programación matemática en c, nos quedaremos con la función de densidad $\delta_u^*(x)$, que se obtiene seleccionando para cada valor de c el correspondiente u en la expresión $\rho'(u)/\rho(u) = c$.

La existencia de $\delta_u^*(x)$ para valores de u en el intervalo $[0, 1]$ está asegurada, pues para $u = 0$ se transforma en $g(x)$, para $u = 1$ en $f(x)$ y para valores en el abierto $(0, 1)$ sabemos que si $a > 0$, $b > 0$ con $a + b = 1$, entonces

$$(f(x))^a\, (g(x))^b \leqslant a\, f(x) + b\, g(x)$$

por tanto

$$0 \leqslant \rho(u) \leqslant 1$$

y es posible normalizar la función $\delta^*(x)$ a $\delta_u^*(x)$. Por este mismo procedimiento es sencillo demostrar que $\rho(u)$ es una función estrictamente convexa. Para otras propiedades de las funciones $\rho(u)$ y $\delta_u^*(x)$ pueden verse Chernoff [1] y Kullback [8].

En cuanto a la función $\nu(u)$ sabemos que es cóncava en $[0, 1]$ y que alcanza su valor máximo en el mismo punto donde $\rho(u)$ alcanza su mínimo, además verifica que

$$-I(g:f)=\log h_0 \leqslant \log(l_1/l_2) \leqslant \log h_1 = I(f:g)$$

que originan los semiespacios

$$-l_1 + h_0 l_2 \leqslant 0 \quad \text{y} \quad l_1 - h_1 l_2 \leqslant 0$$

y por tanto el cono

$$K = \{(l_1, l_2) \in (R^2)^* : -l_1 + h_0 l_2 \leqslant 0,\ l_1 - h_1 l_2 \leqslant 0\}$$

Al ser las componentes de los elementos de K no negativos, se puede definir un conjunto $B(K)$ que son los elementos de K normalizados a que su suma sea la unidad, al que llamaremos *conjunto de soluciones Bayes generado por la información de Kullback.*

Sea $(\xi(u), 1-\xi(u)) \in B(K)$, entonces $\xi(u)/(1-\xi(u)) = \exp c(u)$, donde se define un valor de c y por tanto un u en $[0, 1]$, que nos define una función $\delta_u^*(x)$ solución del problema de óptimo. De igual forma, dado $u \in [0, 1]$, se puede obtener $\delta_u^*(u)$ y por tanto $c(u)$ lo que origina un $(\xi(u), 1-\xi(u)) \in B(K)$.

Además sea $(\xi(u), 1-\xi(u)) \in B(K)$ para $u \in [0, 1]$, entonces

$$\inf\{\xi(u) y_1 + (1-\xi(u)) y_2 : (y_1, y_2) \in S(\Phi)\} =$$

$$= \inf\left\{1 - \xi(u) + \int_\Omega (\xi(u) g(x) - (1-\xi(u)) f(x))\, \phi(x)\, d\mu(x) : \phi \in \Phi\right\}$$

la solución de éste problema de programación matemática infinita nos viene dado por

$$\phi_u(x) = \begin{cases} 1 & \text{si } g(x)/f(x) < (1-\xi(u))/\xi(u) \\ \gamma_u & \text{si } g(x)/f(x) = (1-\xi(u))/\xi(u) \\ 0 & \text{si } g(x)/f(x) > (1-\xi(u))/\xi(u) \end{cases}$$

o lo que es igual:

$$\phi_u(x) = \begin{cases} 1 & \text{si } x \in A^{>}(u) \\ \gamma_u & \text{si } x \in A^{=}(u) \\ 0 & \text{si } x \in A^{<}(u) \end{cases}$$

Podemos observar que el conocimiento de la función de densidad óptima $\delta_u^*(x)$ nos lleva a definir conjuntos de la forma $A^{<}(u)$, $A^{=}(u)$ y $A^{>}(u)$, mediante los cuales se puede construir un subconjunto de soluciones óptimas para la teoría de Neyman-Pearson, sin más que hacer recorrer a u de cero a uno. Si $A^{=}(u)$ es de medida nula, podemos obtener el tamaño de un test $\phi_u(x)$ haciendo

$$\alpha(u) = E\,\{\phi_u(x)/G\} = \int_{A^{>}(u)} dG(x)$$

y será de potencia máxima

$$\beta(u) = 1 - E\,\{\phi_u(x)/F\} = \int_{A^{<}(u)} dF(x)$$

Si $A^{=}(u)$ es de medida no nula, tomemos

$$\alpha(u/\tau) = \Pr\,\{A^{>}(u)/G\} + \tau \Pr\,\{A^{=}(u)/G\}$$

con $0 \leqslant \tau \leqslant 1$, entonces el lema de Neyman-Pearson nos asegura que un test con ese tamaño (τ fijo) es de máxima potencia $\beta(u/\tau)$ si nos viene definido por un $\phi_u(x)$. Para cada extremo $(\alpha(u/0), \beta(u/0))$ y $(\alpha(u/1), \beta(u/1))$, que son puntos de $\partial S(\Phi)$, admiten un conjunto de rectas soportes a $S(\Phi)$, generadas por un cono de vectores $(-h(u), -1)$, que nos definen un subconjunto de valores de u en $[0, 1]$, para el cual el tamaño y la potencia son constantes.

De otra parte si conocemos $S(\Phi)$ y las informaciones $I(g:f)$ e $I(f:g)$, se puede construir un subconjunto de la frontera inferior de $S(\Phi)$ y mediante las rectas soportes a él obtener una función $c(u)$ que induce una función de densidad $\delta_u^*(x)$ que es óptima para el problema de programación infinita de información óptima.

Observemos que es posible obtener tests Bayes cuya distribución a priori que los genera no pertenezca a $B(K)$. Ahora bien, si introducimos una función de pérdida distinta a la usada, suponiendo que $p_1 > 0$ es la pérdida asociada al error de tipo uno y $p_2 > 0$ la correspondiente a error de tipo dos, si $(\xi, 1 - \xi)$ no pertenece a $B(K)$, entonces será solución de información óptima respecto a la pérdida inicial si se verifica la siguiente desigualdad

$$((1 - \xi)/\xi) \exp \{-I(g:f)\} \leqslant p_1/p_2 \leqslant ((1 - \xi)/\xi) \exp \{I(f:g)\}$$

Sabemos que a partir de $F(x)$ y $G(x)$ se puede definir una transformada de la forma

$$\rho(u) = \int_\Omega (f(x))^u \, (g(x))^{1-u} \, d\mu(x)$$

con $u \in [0, 1]$. El valor de u para el cual $\rho(u)$ es mínima tiene una gran importancia en la teoría de Neyman-Pearson, pues sabemos que para dicho punto u_0 existe una función de densidad $\delta^*_{u_0}(x)$ tal que

$$I(\delta^*_{u_0} : g) = I(\delta^*_{u_0} : f)$$

lo que implica $c = 0$ en el problema de óptimo, que puede ser interpretado cómo que la preferencia a priori sobre la hipótesis nula y alternativa son iguales.

Además es conocido, sin más que aplicar la fórmula de Bayes, que:

$$I(f:g) = \int_\Omega \log (\Pr (H_1/x)/\Pr (H_0/x)) f(x) \, d\mu(x) - \log ((1 - \xi)/\xi)$$

donde la primera expresión del segundo miembro indica el valor medio de la variación a favor de H_1 frente a H_0 con respecto a F. Esta variación media es nula para

$$\xi = \exp I(f:g)/(1 + \exp I(f:g))$$

y aumenta según decrece esta probabilidad. Para valores de ξ superiores a éste extremo es negativa, lo que induce a pensar que nuestra confianza en H_0 es excesiva en términos de la información de Kullback. Igual argumento se puede usar con $I(g:f)$ y el extremo

$$\xi = \exp\{-I(g:f)\}/(1 + \exp\{-I(g:f)\})$$

Tratemos ahora de obtener una *desigualdad entre la información óptima y el riesgo Bayes asociado al contraste de hipótesis.* Supongamos que para todo $(\xi, 1-\xi) \in B(K)$ los conjuntos

$$A^{=}(\xi) = \{x \in \Omega : f(x)/g(x) = \xi/1-\xi\}$$

son de medida nula. Sea $\delta_\xi^*(x)$ la función de densidad óptima cuando $c = \log(\xi/1-\xi)$ y $R^*(\xi)$ el riesgo Bayes dado por

$$R^*(\xi) = \inf\{1 - \xi + \int_\Omega (\xi\, g(x) - (1-\xi) f(x))\, \phi(x)\, d\mu(x) : \phi \in \Phi\}$$

consideremos los conjuntos

$$A^{\rangle}(\xi) = \{x \in \Omega : f(x)/g(x) \geqslant \xi/1-\xi\} \quad \text{y}$$

$$A^{\langle}(\xi) = \{x \in \Omega : f(x)/g(x) < \xi/1-\xi\}$$

entonces

$$\rho(u) = \int_\Omega (f(x))^u\, (g(x))^{1-u}\, d\mu(x) = \int_{A^{\rangle}(\xi)} (f(x))^u\, (g(x))^{1-u}\, d\mu(x) +$$

$$+ \int_{A^{\langle}(\xi)} (f(x))^u\, (g(x))^{1-u}\, d\mu(x)$$

y por tanto

$$\xi(\xi/1-\xi)^{-u}\, \rho(u) \geqslant \xi \int_{A^{\rangle}(\xi)} g(x)\, d\mu(x) + (1-\xi) \int_{A^{\langle}(\xi)} f(x)\, d\mu(x)$$

de donde se deduce que

$$\log \xi - \{u \log (\xi/1 - \xi) - \log \rho(u)\} \geqslant \log R^*(\xi)$$

relación que se verifica para todo $u \in (0, 1)$, luego

$$\log \xi - \sup_{u \in (0, 1)} \{u \log (\xi/1 - \xi) - \log \rho(u)\} \geqslant \log R^*(\xi)$$

por tanto

$$\xi \exp \{- I(\delta_\xi^* : g)\} \geqslant R^*(\xi)$$

Ejemplos

A) Sea x una variable aleatoria $N(\mu, \sigma)$, con σ conocida, y $(x_1, x_2, ..., x_n)$ una muestra aleatoria simple de tamaño n. Representamos por $g(x)$ y $f(x)$ las funciones de verosimilitud bajo las hipótesis

$$H_0 : \mu = \mu_0 \qquad H_1 : \mu = \mu_1$$

Entonces la función de densidad del problema de óptimo es

$$\delta_u^*(x_1, x_2, ..., x_n) = \frac{1}{(2\,n)^{n/2}\,\sigma^n} \cdot$$

$$\cdot \exp \left\{ - \frac{1}{2\,\sigma^2} \sum_{i=1}^{n} (x_i - (u\,\mu_1 + (1 - u)\,\mu_0))^2 \right\}$$

siendo la función transformada

$$\rho(u) = \exp \{n\,((u\,\mu_1 + (1 - u)\,\mu_0)^2 - u\,\mu_1^2 - (1 - u)\,\mu_0^2)/2\,\sigma^2\}$$

Las informaciones de Kullback

$$I(\delta_u^* : g) = n\,\{(u\,\mu_1 + (1 - u)\,\mu_0) - \mu_0\}^2/2\,\sigma^2$$

$$I(\delta_u^* : f) = n\ \{(u\,\mu_1 + (1-u)\,\mu_0) - \mu_1\}^2/2\,\sigma^2$$

induce la función

$$c(u) = n\ (\mu_1 - \mu_0)\ \{2\ (u\,\mu_1 + (1-u)\,\mu_0) - (\mu_1 + \mu_0)\}/2\,\sigma^2$$

que nos lleva a la región crítica

$$A^{\rangle}(u) = \{(x_1, x_2, ..., x_n) \in R^n : \bar{x} \geqslant u\,\mu_1 + (1-u)\,\mu_0\}$$

B) Sea x una variable aleatoria que toma el valor uno con probabilidad p y cero con probabilidad $q = 1 - p$, su función de probabilidad nos viene dada por:

$$\Pr\{x/p\} = p^x\ q^{1-x}$$

con $x \in \{0, 1\}$ y $p \in (0, 1)$. Tomemos la muestra aleatoria simple $(x_1, x_2, ..., x_n)$, entonces las leyes de probabilidad de la variable $r = \sum_{i=1}^{n} x_i$ bajo las hipótesis

$$H_0 : p = p_0 \qquad H_1 : p = p_1$$

son

$$g(r/p_0) = \binom{n}{r} p_0^r\ q_0^{n-r} \qquad \text{y} \qquad f(r/p_1) = \binom{n}{r} p_1^r\ q_1^{n-r}$$

La función transformada es

$$\rho(u) = (p_0^u\ p_1^{1-u} + q_0^u\ q_1^{1-u})^n$$

siendo la ley de probabilidad óptima

$$\delta_u^*(r) = \binom{n}{r}\ (p_0^u\ p_1^{1-u})^r\ (q_0^u\ q_1^{1-u})^{n-r}/\rho(u)$$

La obtención de la función

$$c(u) = n\ \{p_0^u\ p_1^{1-u} \log(p_1/p_0) + q_0^u\ q_1^{1-u} \log(q_1/q_0)\}\ (\rho(u))^{-1/n}$$

nos lleva a la región crítica

$$A^{\rangle}(u) = \{r \in R : r \geqslant (c(u) - n \log(q_1/q_0))/\log(p_1 q_0/p_0 q_1)\}$$

cuyo signo depende del que tome $\log(p_1 q_0/p_0 q_1)$.

DISTRIBUCIONES ASOCIADAS AL CONJUNTO DE RIESGOS

Siguiendo la línea de H. Kudo [7] consideremos un conjunto S de R^2 verificando las siguientes condiciones:

1.– $D_0 \subset S \subset D_1$

2.– S es cerrado, convexo y con interior no vacio

3.– S es simétrico con respecto al punto $(1/2, 1/2)$ y $S \cap D_1 = \{(1, 0), (0, 1)\}$

Sea ξ un número real perteneciente al intervalo $[0, 1]$, que puede ser interpretado como una distribución a priori. Consideremos el siguiente problema de programación matemática, hallar $(y_1, y_2) \in R^2$, tal que

$$\inf \{\xi y_1 + (1 - \xi) y_2\}$$

sujeto a la restricción

$$(y_1, y_2) \in S$$

El conjunto de óptimos para éste problema, ξ fijo, lo representaremos por $\partial S(\xi)$ que puede ser un punto si la solución es única o bien un segmento contenido en la frontera inferior de S.

Sea $k = \xi/(1 - \xi)$ y definimos sobre la recta real las funciones (Kudo [7], Lehmann [9] pág. 65)

$$F^*(\log k) = \sup \{\beta : (\alpha, \beta) \in \partial S(\xi)\}$$

$$G^*(\log k) = \sup \{1 - \alpha : (\alpha, \beta) \in \partial S(\xi)\}$$

que verifican las propiedades

$$\lim_{t \to -\infty} F^*(t) = \lim_{t \to -\infty} G^*(t) = 0$$

$$\lim_{t \to +\infty} F^*(t) = \lim_{t \to +\infty} G^*(t) = 1$$

Sean las medidas $H(./F^*)$ y $H(./G^*)$, absolutamente contínuas una respecto de la otra y dominadas por una medida v, definidas sobre la recta real por

$$H((-\infty, t)/F^*) = \int_{-\infty}^{t} f^*(t)\, dv(t) = F^*(t)$$

$$H((-\infty, t)/G^*) = \int_{-\infty}^{t} g^*(t)\, dv(t) = G^*(t)$$

donde f^* y g^* son las derivadas de Radon-Nikodym de $H(./F^*)$ y $H(./G^*)$ con respecto a la medida v.

Observemos que la construcción de estas funciones ha sido posible a partir de ξ, por lo tanto son funciones de la distribución a priori que nos describen la frontera inferior de los conjuntos de riesgos S sin especificar las medidas que los originan. Hacemos notar de igual forma que las pérdidas que estamos usando son la unidad sobre los riesgos de tipo uno y dos, pero en el caso de usar otras tendríamos que normalizar estas funciones sobre el conjunto S, que lógicamente también se transforma.

Tratemos ahora de obtener una distribución II que sea la más próxima a la distribución G^* en el sentido de la información de Kullback, con la condición de que su momento de primer orden está prefijadoy sea un número real d, a este valor se le podrá dar la interpretación de un parámetro cuando nos interese. El problema de programación matemática infinita es el de hallar una función $\pi(t)$, tal que

$$\inf \int_{-\infty}^{+\infty} \log(\pi(t)/g^*(t))\, \pi(t)\, dv(t)$$

sujeto a la restricción de que

$$\pi(t) \in \Lambda(d) = \{\pi(t) : \int_{-\infty}^{+\infty} t\, \pi(t)\, dv(t) = d, \int_{-\infty}^{+\infty} \pi(t)\, dv(t) = 1\}$$

que corresponde al problema clásico de mínima información. Ahora bien, es fácil comprobar que:

$$\int_{-\infty}^{t} e^{s}\, dG^*(s) = F^*(t)$$

de donde se deduce que

$$d = \int_{-\infty}^{+\infty} t\, \pi(t)\, dv(t) = \int_{-\infty}^{+\infty} \log\,(f^*(t)/g^*(t))\, \pi(t)\, dv(t) =$$

$$= I\,(\pi : g^*) - I\,(\pi : f^*)$$

por tanto el problema de mínima información pasa a ser uno de información óptima, y la solución nos viene dada por

$$\pi^*(t) = g^*(t) \exp\,\{t z + \omega - 1\}$$

donde z y ω son los multiplicadores de Lagrange asociados al problema. La normalización de la función $\pi^*(t)$ para obligarle a que sea de densidad nos lleva a

$$\pi_z^*(t) = g^*(t)\, e^{tz}/\eta(z)$$

donde

$$\eta(z) = \int_{-\infty}^{+\infty} e^{tz}\, g^*(t)\, dv(t) = \int_{-\infty}^{+\infty} (f^*(t))^z\, (g^*(t))^{1-z}\, dv(t)$$

que es una *función generatriz de momentos o transformada de Laplace* para la variable t que toma valores positivos y negativos con probabilidad mayor que cero.

Cuando z tiene una variación de cero a la unidad, sabemos que d tiene una variación estrictamente monótona entre

$$\int_{-\infty}^{+\infty} t\, g^*(t)\, dv(t) = -I\,(g^* : f^*) \quad \text{e} \quad \int_{-\infty}^{+\infty} t\, f^*(t)\, dv(t) = I\,(f^* : g^*)$$

de donde se deduce que las funciones F^* y G^* tienen medias positiva y negativa respectivamente, hecho que era de esperar por las propiedades de la función generatriz de momentos (Cox-Miller [3]) para estas funciones tan particulares.

La diferencia entre las medias de t por medio de F^* y G^* nos dan la divergencia entre las hipótesis H_0 y H_1, y como aseguran Chernoff [1] y Kullback [8] es una medida de la dificultad de discriminación entre ellas:

$$E\ \{t/F^*\} - E\ \{t/G^*\} = J\,(f^* : g^*)$$

Para otros valores de z será:

$$d = -\,\omega' = \frac{1}{\eta(z)} \int_{-\infty}^{+\infty} t\, e^{tz}\ g^*(t)\, dv(t)$$

Si S está generado por las funciones F y G, así como por un conjunto de tests Φ y si $\partial S(\xi)$ consta de un único punto para todo ξ en $[0, 1]$, entonces

$$F^*(\log k) = \beta \quad \text{y} \quad G^*(\log k) = 1 - \alpha$$

y tendremos:

$$F^*(t) = \int_{A^{\langle}(t)} f(x)\, d\mu(x) \quad \text{y} \quad G^*(t) = \int_{A^{\langle}(t)} g(x)\, d\mu(x)$$

con $A^{\langle}(t) = \{x \in \Omega : \log(f(x)/g(x)) \leqslant t\}$, si efectuamos el cambio de variable:

$$c(u) = \int_{\Omega} \log(f(x)/g(x))\, \delta_u^*(x)\, d\mu(x) = \frac{1}{\rho(u)} \int_{-\infty}^{+\infty} t\, e^{tu}\, dG^*(t)$$

por tanto

$$\int_{-\infty}^{+\infty} t\, g^*(t)\, dv(t) = -I(g:f) \quad \text{e} \quad \int_{-\infty}^{+\infty} t\, f^*(t)\, dv(t) = I(f:g)$$

de donde se deduce que las medias de las medidas F^* y G^* son las informaciones de Kullback inducidas por las medidas F y G que originan el conjunto S.

Tratemos ahora de usar la función generatriz $\eta(z)$ para dar cotas de las funciones F^* y G^*, cuando el conjunto S admite asimetría en el sentido de que la *solución mínimax* t_0 ($y_1 = y_2$ en la frontera inferior de S) no corresponde con $t = 0$ que es la *solución de indiferencia* entre H_0 y H_1.

Si $0 \leqslant a \leqslant t_0$ entonces

$$F^*(a) \leqslant 1 - G^*(a)$$

por tanto:

$$\int_{-\infty}^{a} f^*(t)\, dv(t) \leqslant 1 - \int_{-\infty}^{a} g^*(t)\, dv(t) \leqslant 1 - \int_{-\infty}^{a} e^{(t-a)z}\, g^*(t)\, dv(t) =$$

$$= 1 - e^{-az} \int_{-\infty}^{a} e^{zt}\, g^*(t)\, dv(t)$$

de donde se deduce que:

$$F^*(a) \leqslant 1 - e^{-az_0}\, \eta(z_0)\, \Pi_{z_0}^*(a) \leqslant 1 - e^{-t_0 z_0}\, \eta(z_0)\, \Pi_{z_0}^*(0)$$

pues la cadena de desigualdades es válida para todo z positivo y tomamos z_0 que hace mínima a $\eta(z)$.

Si $t_0 \leqslant b \leqslant 0$, por el mismo procedimiento se llega a

$$1 - G^*(b) \leqslant e^{-bz_1} \hat{\eta}(z_1) \hat{\Pi}_{z_1}(b) \leqslant e^{-t_0 z_1} \eta(z_1) \hat{\Pi}_{z_1}(0)$$

donde $\hat{\eta}(z) = \int_{-\infty}^{+\infty} (f^*(t))^{1-z} (g^*(t))^z \, dv(t)$ y z_1 es el valor de z que hace mínima a $\hat{\eta}(z)$.

Ejemplos

A) Sea el conjunto de R^2

$$C = \{(y_1, y_2) : (y_1 - 1)^2 + (y_2 - 1)^2 = 1,\ 0 \leqslant y_1 \leqslant 1,\ 0 \leqslant y_2 \leqslant 1\}$$

y construimos C' a partir de C tomando los puntos simétricos de éste conjunto con respecto al punto $(1/2, 1/2)$. Definimos el conjunto S como la intersección del epígrafo de C y el hipógrafo de C'.

Tomemos $y_2' = -(y_1 - 1)/(y_2 - 1) = h$, sea $y_1 - 1 = \text{Cos}\ \theta$, $y_2 - 1 = $ $= \text{Sen}\ \theta$, entonces

$$y_1 = 1 - h(1 + h^2)^{-1/2} \quad h \in R^- \Rightarrow G^*(h) = h(1 + h^2)^{-1/2} \quad h \in R^+$$

$$y_2 = 1 - (1 + h^2)^{-1/2} \quad h \in R^- \Rightarrow F^*(h) = 1 - (1 + h^2)^{-1/2} \quad h \in R^+$$

por tanto las funciones de densidad buscadas son

$$g^*(t) = e^t (1 + e^{2t})^{-3/2} \quad \text{y} \quad f^*(t) = e^{2t} (1 + e^{2t})^{-3/2}$$

con $t \in R$. La función de densidad de G^* se puede identificar con una t de Student con dos grados de libertad.

B) Sea el conjunto de R^2

$$C = \{(y_1, y_2) : y_1^{2/3} + y_2^{2/3} = 1,\ y_1 \geqslant 0,\ y_2 \geqslant 0\}$$

y construimos el conjunto S como en A). Entonces:

$$g^*(t) = 3\,e^{2t}\,(1 + e^{2t})^{-5/2} \quad \text{y} \quad f^*(t) = 3\,e^{3t}\,(1 + e^{2t})^{-5/2}$$

con $t \in R$.

C) Sea el conjunto de R^2

$$C = \{(y_1, y_2) : y_2 = (y_1 - 1)^2,\ 0 \leqslant y_1 \leqslant 1\}$$

y construimos el conjunto S como en A). Entonces

$$g^*(t) = \frac{1}{2}\,e^t \quad \text{y} \quad f^*(t) = \frac{1}{2}\,e^{2t}$$

con $t \in (-\infty, \log 2)$. Es sencillo calcular

$$\eta(z) = 2^z/(z + 1), \quad z_0 = (1/\log 2) - 1,$$

$$\Pi_z^*(t) = (z + 1)\,e^{t(z+1)}/2^{(z+1)}, \quad d = \log 2 - (1 + z)^{-1}$$

$$E\{t/G^*\} = -I\,(g^* : f^*) = -0{,}3069 \quad E\{t/F^*\} = I\,(f^* : g^*) = 0{,}1931$$

Además

$$G^*(\xi) = \xi/2\,(1 - \xi) \quad \text{y} \quad F^*(\xi) = \xi^2/4\,(1 - \xi)^2$$

para $\xi \in (0, 2/3]$, para los valores de $\xi \in (2/3, 1]$ las funciones son iguales a la unidad.

BIBLIOGRAFIA

[1] Chernoff, H. (1952). "A measure of asymptotic efficiency for tests of hypothesis based on the sum of observations". *Ann. Math. Statist.*, Vol. 23, pp. 493-507.

[2] Chernoff, H. (1956). "Large-sample theory: parametric case". *Ann. Math. Statist.* Vol. 27. pp. 1-22.

[3] Cox, D.R. - Miller, H.D. (1968). "The Theory of Stochastic Processes". *Chapman and Hall.* London.

[4] Ferguson, T.S. (1967). "Mathematical Statistics: A decision theoretic approach". *Academic Press.* New York - London.

[5] Hurwicz, L. (1958). "Programming in Linear Spaces" en Arrow, K., L. Hurwicz y H. Uzawa (eds): "Studies in Linear and Nonlinear Programming". *Stanford University Press.* Stanford. California. pp. 38-102.

[6] Isii, K. (1964). "Inequalities of the types of Chebychev and Cramér - Rao and mathematical programming". *Ann. Inst. Stat. Math.* Vol. 16, pp. 277-293.

[7] Kudo, H. (1952). "A Theorem of Kakutani on Infinite Product Measures". *Natural Science Report Ochanamizu University.* Vol. 3. pp. 10-22.

[8] Kullback, S. (1968). "Information Theory and Statistics". *Dover.* New York.

[9] Lehmann, E.L. (1959). "Testing Statistical Hypotheses". *John Wiley.* New York.

[10] Liapounov, A. (1940). "Sur les fonctions - vecteurs complétement additives". *Bull. Acad. Sci. URSS.* Vol. 4. pp. 465-478.