Confiabilidad de la probabilidad posterior

Por: Luis Pérez G.*

", , , , TEOREMA: Un apostador será un seguro perdedor si sus apuestas no son consistentes con el teorema de Bayes. . . . " Heath and Sudderth. Annals of Mathematics.

RESUMEN: Luego de hacer una introducción de los conceptos y del lenguaje utilizado por la teoría bayesiana y por la teoría de decisión, se procede a hacer un análisis acerca de la confiabilidad de la probabilidad posterior, y se introducen generalizaciones de expresiones matemáticas con múltiples interpretaciones eventuales. Como se sabe, es la probabilidad posterior, la causa principal del desfase entre bayesianos y clásicos.

TEORIA BAYESIANA

La teoría bayesiana es modernamente toda una ideología científica, y aunque dependiente en sus raíces de la teoría de la probabilidad y de la estadística clásica, crea situaciones y métodos científicos que difieren grandemente de lo clásico. Con el auge relativamente reciente de la teoría bayesiana, podría pensarse que la estructura de ella fue creada sólo hace algunas décadas; no, fue el sacerdote inglés Thomas Bayes quien en 1763 inspiró sus raíces en el libro "An essay towards solving a problem in the doctrine of chances". El desarrollo y axiomatización de la teoría bayesiana ha sido un proceso muy lento, y aunque grandes matemáticos - Laplace (1810), De Morgan (1838), Boole (1854), Edgeworch (1908),. . . - trabajaron con el convencimiento de poder convertirla en "una técnica sobresaliente e independiente", sólo en los últimos años ha tomado un auge de notables dimensiones en las diferentes áreas aplicadas y específicamente en la teoría de decisión, la cual no tendría tantas aplicaciones prácticas si no se le hubiese inyectado las técnicas bayesianas.

Contrario a lo que podría pensarse, la teoría de Bayes ha enriquecido de una manera notoria la teoría de la probabilidad y de la estadística clásica, en lugar de haber obstaculizado su desarrollo. A partir de ella, surgen permanentemente nuevos conceptos dentro de la inferencia estadística; los modelos lineales presentan enfoques diferentes si se trabajan en base a Bayes; la teoría de muestreo sufre cambios fundamentales con la inclusión de los conceptos bayesianos; los métodos de docimasia de hipótesis adquieren dimensiones distintas bajo esta teoría ya que los coeficientes de significancia (α , β) se conciben de una manera distinta pues son funciones dependientes del costo, lo cual ignora irrazonablemente la teoría clásica.

La teoría bayesiana apasiona y genera discusiones acaloradas entre sus seguidores y los clásicos; a tal punto, que con frecuencia se presentan dificultades entre las relaciones personales de los estudiosos por el solo hecho de pertenecer a una u otra tendencia. Existe también en la época moderna una paulatina emigración de clásicos hacia la teoría Bayesiana, lo cual es acogido con beneplácito por los disidentes establecidos. Resulta de interés conocer la opinión de uno de los valores jóvenes más prominentes de la estadística americana, James Berger, al desertar del campo clásico: ". . . I turned into a rabid bayesian. There was no single cause for this conversion; just a gradual realization that things seemed to ultimately make sense only when looked at from bayesian point of view. . . "

Para Berger, la teoría de decisión estudiada desde el punto de vista clásico es sólo un desarrollo teórico de diferentes principios de optimización aplicables a procedimientos estadísticos; y vista desde el ángulo bayesiano, mediante la inclusión de la probabilidad anterior y posterior y demás conceptos propios de esta teoría, adquiere inagotable aplicabilidad y la posibilidad de mayor expansión teórica.

La teoría de Bayes no es infalible, a sus ventajas tiene tácitamente adheridas múltiples posibilidades de error. Los siguientes comentarios que se transcriben ayudan a la formación de un juicio acerca de la teoría.

1. Una razón filosófica para la utlización de Bayes es la aplicación del principio del comportamiento humano donde siempre se actúa asumiendo algún tipo de información previa.

2. Una gran debilidad de la teoría es, a veces, la falta de objetividad en la escogencia de la probabilidad anterior y la posibilidad de generar sistemáticamente graves errores cuando se parte de una probabilidad anterior falsa.

3. La creación de técnicas que permiten denunciar como sospechosa a una probabilidad anterior falsa a través del proceso, le crea fuerza como técnica científica.

4. Lo real de los resultados obtenidos por medio de la teoría de Bayes es lo que anima a buscar un mayor desarrollo de ella, paralelamente con técnicas correctivas en caso de errores iniciales en la escogencia de la información previa.

TEORIA DE DECISION

La teoría de decisión nace como una rama avanzada de la estadística matemática, y como se mencionó, es llevada a sus cauces prácticos mediante la inclusión de la teoría bayesiana. Para cumplir su objetivo maneja conceptos propios como: estado de naturaleza o ente desconocido, lo cual dificulta la decisión; acción a tomar, y su espacio de valores; función de riesgo, la cual varía con las distintas posibilidades del estado de naturaleza y de las acciones; admisibilidad; información previa y posterior; etc.

De esta manera, y en términos muy generales, podría

* Profesor Asociado
Universidad Nacional de Colombia
Seccional Medellín

decirse que la teoría de decisión en práctica opera a partir de información previa, y por medio de sus técnicas, consigue una información posterior la cual le permite generar decisiones aplicables a las situaciones en estudio.

Con la introducción de la probabilidad anterior y posterior en la teoría de decisión surge el mencionado desfase con lo clásico. Se notó que la probabilidad anterior puede contener errores al ser seleccionada; una rápida descripción de las diferentes probabilidades anteriores permitirá entender la posibilidad latente de imprecisión en su escogencia. Puede ser subjetiva, objetiva y empírica. La primera se da a partir de un grado de creencia personal; la segunda aparece como un concepto ajeno a todo punto de vista personal y subjetivo; y la probabilidad empírica nace a partir de experiencias pasadas y de alguna manera hace uso de la estadística clásica; es decir, la probabilidad anterior empírica es un último paso para tratar de conciliar lo bayesiano con lo clásico.

TERMINOLOGIA PRELIMINAR

Antes de emprender en detalle el trabajo que se quiere mostrar, es de vital importancia homogenizar las notaciones. Sea θ un parámetro el cual puede tomar valores $\theta_1, \theta_2 \ldots, \theta_K$; se está interesado en estimar con cierta confiabilidad la distribución de probabilidad de θ, esto es, la probabilidad de que θ tome el valor θ_i para i = 1, 2, . . ., k.

Sea $\pi_i = P(\theta = \theta_i)$ la probabilidad anterior asumida, bien sea objetiva, subjetiva o empírica.

Sea $\pi_i(X) = P(\theta = \theta_i / X)$ la probabilidad posterior de θ luego de haber tomado una muestra aleatoria X proveniente de una población con función de probabilidad f (x).

Aplicando el Teorema de Bayes:

$$\pi_i(X) = \frac{\pi_i \, f(x/\theta_i)}{\sum_{j=1}^{k} \pi_j \, f(x/\theta_j)} \qquad (1)$$

donde $f(x/\theta_j)$ es la función de probabilidad de X dado $\theta = \theta_j$

Si la muestra estudiada es $X_1, X_2, \ldots, X_n$, se obtiene una probabilidad posterior conjunta para las n muestras:

$$\pi_i(X_1, X_2, \ldots, X_n) = \frac{\pi_i \left[\prod_{j=1}^{n} f(x_j/\theta_i)\right]}{\sum_{r=1}^{k} \pi_r \left[\prod_{j=1}^{n} f(x_j/\theta_r)\right]}$$

PROBLEMAS A DISCUTIR

Los comentarios pasados están dirigidos a la necesidad de discutir la confiabilidad de la probabilidad posterior. De los cuatro puntos a analizar, uno de ellos es ampliamente conocido en la literatura de la teoría de decisión y se trae a mención por ser una propiedad de trascendencia; los demás, constituyen generalizaciones personales hechas a partir de sugerencias de los profesores B. Hill y M. De Groot, quienes son conocidos ideólogos de estas áreas. Dichos puntos a discutir son los siguientes:

1. Propiedades en el límite de la probabilidad posterior.
2. Esperanza matemática de $\pi_i(x)$ para i = 1, 2, . . , k.
3. Esperanza matemática de la probabilidad posterior condicionada a que θ tome un valor θ_i específico, esto es,

$$E(\pi_i(x) / \theta = \theta_i)$$

4. La probabilidad de que la distribución posterior asigne un valor muy pequeño al valor correcto de θ. Es decir,

$$P\left[\pi_i(x) \leqslant \delta \;/\; \theta = \theta_i\right]$$

si los valores de θ son igualmente probables.

PROPIEDADES EN EL LIMITE

Se sabe que las propiedades en el límite de cualquier entidad matemática siempre son importantes, y si las propiedades en ese punto no son adecuadas a la realidad, nada promisorio debería esperarse de tal entidad. Por fortuna, las propiedades en el límite de la probabilidad posterior son adecuadas, y puede encontrarse en la literatura de teoría de decisión la siguiente demostración: Si

$X_1, X_2, \ldots, X_n$ es una muestra aleatoria de una distribución para la cual el valor de un parámetro θ es desconocido; y si el valor real de θ es θ_0; entonces a medida que $n \to \infty$, bajo ciertas condiciones, la distribución posterior de θ tiende a estar más concentrado alrededor de $\theta = \theta_0$

ESPERANZA DE LA PROBABILIDAD POSTERIOR

La esperanza de la probabilidad posterior genera un resultado interesante.

Por definición:

$$E\left[\pi_i(x)\right] = \int \pi_i(x) \, f(x) \, d\mu(x).$$

Aplicando el teorema de probabilidad total:

$$E\left[\pi_i(x)\right] = \int \pi_i(x) \left[\sum_{j=1}^{k} f(x/\theta_j)\, \pi_j\right] d\mu(x).$$

Asumiendo la expresión (1), la esperanza se convierte en:

$$E\left[\pi_i(x)\right] = \int f(x/\theta_i) \, \pi_i \, d\mu(x)$$

ó

$$E\left[\pi_i(x)\right] = \pi_i \quad \text{ya que} \int f(x/\theta_i)\, d\mu(x) = 1.$$

Es éste un resultado afortunado para algunos casos y desafortunado cuando no se tiene confianza en la probabilidad anterior. Muestra la clara dependencia entre las probabilidades anterior y posterior, y muestra también la posibilidad de que la distribución anterior genere todos sus defectos a la posterior. Con este resultado no se debe ser muy trascendental pues el concepto de esperanza matemática ha sido revaluado por muchos autores en la práctica, y en muchas ocasiones ellos escogen estimadores sesgados con propiedades especiales que los caracterizan.

ESPERANZA CONDICIONAL DE LA POSTERIOR

El siguiente problema a resolver está relacionado con la búsqueda de una expresión para:

$$E(\pi_i(x) / \theta = \theta_i)$$

Es decir, cuál es el valor esperado de $\pi_i(x)$ asumiendo que efectivamente θ es θ_i.

Por definición se sabe:

$$E[\pi_i(x) / \theta = \theta_i] = \int \pi_i(x) f(x / \theta_i) d\mu(x).$$

Utilizando las expresiones (1) y (2) dadas anteriormente:

$$E[\pi_i(x) / \theta = \theta_i] = \int \left[1 + \sum_{j \neq i} \frac{\pi_j f(x/\theta_j)}{\pi_i f(x/\theta_i)}\right]^{-1} f(x/\theta_i) d\mu(x).$$

Haciendo $Z_j = \dfrac{f(x/\theta_j)}{f(x/\theta_i)}$, se obtiene:

$$E[\pi_i(x) / \theta = \theta_i] = \int \left[1 + \sum_{j \neq i} \frac{\pi_j}{\pi_i} Z_j\right]^{-1} f(x/\theta_i) d\mu(x).$$

La expresión a la derecha también puede ser expresada como otra esperanza condicional; esto es,

$$E[\pi_i(x) / \theta = \theta_i] = E\left[\left(1 + \sum_{j \neq i} \frac{\pi_j}{\pi_i} Z_j\right)^{-1} \Big/ \theta = \theta_i\right]$$

Por propiedades de las funciones de probabilidad y utilizando el cálculo, puede probarse que la función:

$$\left[1 + \sum_{j \neq i} \frac{\pi_j}{\pi_i} Z_j\right]^{-1}$$

es convexa.

Aplicando la desigualdad de Jensen para funciones convexas

$$E\left[\pi_i(x) / \theta = \theta_i\right] \geqslant \left\{1 + E\left[\sum_{j=1}^{k} \frac{\pi_j}{\pi_i} Z_j \Big/ \theta = \theta_i\right]\right\}^{-1}$$

y como: $E(Z_j / \theta = \theta_i) = 1$,

entonces:

$$E\left[\pi_i(x) / \theta = \theta_i\right] \geqslant \left[1 + \sum_{j \neq i} \frac{\pi_j}{\pi_i}\right]^{-1}$$

Lo cual es equivalente a:

$$E\left[\pi_i(x) / \theta = \theta_i\right] \geqslant \pi_i$$

debido a la igualdad $\sum_j \pi_j = 1$

El resultado indica que la distribución posterior asignará un valor más adecuado a la probabilidad, que aquel que asigna la distribución anterior.

ULTIMO PROBLEMA

El último punto concierne con la consecución de un valor interpretable para la probabilidad de que la distribución posterior asigne un valor muy pequeño δ al correcto valor de θ. En el lenguaje de los símbolos se quiere analizar:

$$P\left[\pi_i(x) \leqslant \delta / \theta = \theta_i\right]$$

cuando los posibles valores del parámetro son equiprobables.

Usando expresiones (1) y (2) se nota que las siguientes expresiones son equivalentes:

$$P\left[\pi_i(x) \leqslant \delta / \theta = \theta_i\right]$$

$$= P\left\{\left[1 + \sum_{j \neq i} \frac{\pi_j f(x/\theta_j)}{\pi_i f(x/\theta_i)}\right]^{-1} \leqslant \delta \Big/ \theta = \theta_i\right\}$$

$$= P\left[\frac{1}{\delta} - 1 \leqslant \sum_{j \neq i} \frac{\pi_j f(x/\theta_j)}{\pi_i f(x/\theta_i)} \Big/ \theta = \theta_i\right]$$

La última probabilidad de la derecha permite decir que:

$$P[\pi_i(x) \leqslant \delta/\theta = \theta_i] = \int_\lambda f(x/\theta_i)\, d\mu\,(x),$$

cuando la integración se hace sobre el conjunto:

$$\lambda = \left[X: \frac{1}{\delta} - 1 \leqslant \sum_{j \neq i} \frac{\pi_j f(x/\theta_j)}{\pi_i f(x/\theta_i)} \right].$$

Ahora, sabiendo que:

$$\pi_i f(x/\theta_i) \leqslant \sum_{j \neq i} \pi_j f(x/\theta_j) \left[\frac{\delta}{1-\delta}\right],$$

la expresión de interés se convierte en:

$$\pi_i P[\pi_i(x) \leqslant \delta/\theta = \theta_i] \leqslant \int_\lambda \frac{\delta}{1-\delta} \sum_{j \neq i} \pi_j f(x/\theta_j)\, d\mu\,(x)$$

y ya que $\int_\lambda f(x/\theta_j) \leqslant 1$ para todo j,

$$P[\pi_i(x) \leqslant \delta/\theta = \theta_i] \leqslant \frac{\delta}{1-\delta} \left[\frac{1-\pi_i}{\pi_i}\right]$$

La desigualdad a la cual se llega posiblemente no tenga aplicaciones inmediatas; sin embargo, no es ésto lo que le da importancia. Eventualmente, esta desigualdad puede ser muy útil en problemas específicos. Veamos lo que ocurre en un problema de dos acciones equiprobables, el cual es bastante común en el trabajo de teoría de decisión. Haciendo los reemplazos en la fórmula principal, se consigue la expresión:

$$P(\pi_i(x) \leqslant \delta/\theta = \theta_i) \leqslant \frac{\delta}{1-\delta}.$$

Este resultado indica que existe una probabilidad pequeñísima de que se le asigne un valor pequeño a la probabilidad del valor correcto de θ.

Podrían considerarse otros tipos de aplicaciones para conocer los alcances de esta fórmula generalizada.

BIBLIOGRAFIA

BERGER, J. O. Lecture on Decision Theory,. University of Michigan, 1982.

BERGER, J. O. Statistical Decision Theory. Springer - Verlag. 1980.

DAWID, A. P. Posterior expectations for large observations. Biometrika 60, 664 - 666. 1973.

DE GROOT, M. Optimal Statistical Decisions. Mc Graw - Hill. 1970.