Aplicaciones actuariales de las teorías de la "fiabilidad" y de la "credibilidad"

Por Angel VEGAS

Pretendemos con este trabajo precisar la "fiabilidad" o grado de confianza racional en la permanencia de una situación determinada, disponiendo de una información estadística con un cierto grado de "credibilidad" por una parte, y por otra, concretar cuál deba ser el valor que debe adjudicarse al parámetro que se haya elegido para definir la situación, en el caso en que ésta no permanezca vigente.

I

Fiabilidad.—De acuerdo con Bazovsky, definiremos la fiabilidad como la "probabilidad de que un sistema funcione de forma correcta durante un cierto tiempo y en adecuadas condiciones de utilización". Se trata, pues, de una probabilidad de supervivencia, en cierto modo.

Para la estimación de la "fiabilidad" procederemos de la siguiente forma:

Supongamos conocido el valor del parámetro Θ que hemos elegido para definir la estructura o situación cuya probabilidad de permanencia queremos estimar, y la $F_\Theta(x)$ función de distribución de probabilidad correspondiente a las variantes muestrales ξ.

Sabemos que si α es el nivel de significación correspondiente a la región crítica ω, para el contraste de la hipótesis $\Theta = \Theta_0$ tendremos

$$\int_{\omega} dF_{\Theta}(x) = \alpha \qquad [1]$$

Si r de las $X_1, X_2, \ldots, X_n$ muestrales no pertenecen a ω, tomaremos por valor de la fiabilidad

$$f = \frac{r}{n} \qquad [2]$$

Credibilidad.—La estimación de la fiabilidad depende del tamaño n de la muestra.

Siguiendo a Longley-Cook y a De Finetti, diremos que la "credibilidad" o nivel de información de la estadística de que se dispone para la obtención de la muestra, es una función creciente de n, asintóticamente igual a uno. Así, pues, la medida de la "credibilidad" podrá expresarse mediante la fórmula

$$c = \frac{n}{n + N} \qquad [3]$$

De Finetti propone para la determinación de N las siguientes consideraciones de carácter bayesiano:

Supondremos que la distribución de probabilidad *a priori* correspondiente a la situación cuya vigencia queremos estudiar es del tipo beta:

$$\frac{\Gamma(R + S + 2)}{\Gamma(R + 1)\ M(S + 1)} \int_0^x v^R(1 - v)^S\, dv \qquad [4]$$

en la que R y S representan el número de sucesos favorables y contrarios, respectivamente, de un total de $N = R + S$.

Sabemos que el valor medio y la varianza vienen expresados por las fórmulas

$$m = \frac{R + 1}{N + 2}$$

$$\sigma^2 = \frac{(R + 1)(S + 1)}{(N + 2)^2 (N + 3)} \qquad [5]$$

Determinamos el valor de N, R y S con la condición de que m coincide con $(1 - \alpha)$.

Luego de [5] se desprende

$$\frac{R + 1}{N + 2} = 1 - \alpha$$

$$\sigma^2 = \frac{\alpha(1 - \alpha)}{N + 3} \qquad [6]$$

$$N = \frac{\alpha\,(1 - \alpha)}{\sigma^2} - 3$$

De aquí que si $\sigma = t(1 - \alpha)$

$$N = \frac{\alpha}{t^2(1 - \alpha)} - 3 \qquad [7]$$

Este resultado puede interpretarse de la forma siguiente: N representa el tamaño de una muestra ficticia realizada en la situación cuya permanencia suponemos, con la probabilidad $(1 - \alpha)$.

La expresión de la "credibilidad" será entonces

$$c = \frac{n}{n + \dfrac{\alpha}{t^2(1 - \alpha)} - 3} \qquad [8]$$

De aquí que el tamaño de la muestra n que corresponde a una "credibilidad" de nivel β será

$$n = \frac{\beta}{1 - \beta}\left[\frac{\alpha}{t^2(1 - \alpha)} - 3\right] \qquad [10]$$

Como consecuencia de estos resultados estimaremos la "fiabilidad" según la fórmula

$$f = \frac{r + R}{n + N}$$

en la que

$$N = \frac{\alpha}{t^2(1 - \alpha)} - 3$$

y

$$R = N \cdot (1 - \alpha)$$

Estimación del parámetro.

Hemos supuesto que la situación de que partíamos venía definida porque el parámetro Θ tenía el valor Θ_0. Supongamos ahora que el valor que se obtiene mediante una estimación a partir de la muestra es Θ_1, lo que equivaldría a decir que si la muestra era totalmente "creíble", éste debía ser el valor definitivo de Θ. Pero como la total "credibilidad" no se da, propondremos para valor de Θ el valor medio de la variante que puede tomar los valores Θ_0, Θ_1 con probabilidades f, 1 — f, es decir,

$$\Theta = \frac{r + R}{n + N}\,\Theta_0 + \frac{s + S}{n + N}\,\Theta_1 \qquad [11]$$

en la que $S = N - R$ y $s = n - r$.

Esta podría ser la solución al problema relativo al nivel de prima en el caso del seguro de accidentes, cuando se pretende modificarle teniendo en cuenta la información obtenida de las estadísticas más recientes, sin prescindir de los antecedentes recogidos en la determinación de la prima vigente hasta el momento.

La diferencia fundamental entre este resultado y los propuestos normalmente por Longley Cook y Allen Mayerson, entre otros, está en que estos autores, con razonamientos consistentes, en lugar de operar apoyándose en la fiabilidad para la estimación del nuevo nivel de primas, se sirven de la "credibilidad", es decir,

$$\Theta = (1 - c)\,\Theta_0 + c\Theta_1 \qquad [12]$$

El criterio seguido por tan ilustres autores se refiere exclusivamente a la calidad de la información estadística a través del tamaño de la muestra. El criterio sustentado por nosotros, no sólo tiene en cuenta la "credibilidad" de la muestra, sino también la probabilidad de permanencia de la situación definida con un cierto nivel de significación.

Estimación de $F_\Theta(x)$.—En las anteriores consideraciones hemos partido del conocimiento de la función de probabilidad de la variante ξ. Vamos ahora a tratar de su estimación en el caso de no poseer otra información que la del valor del parámetro Θ que vamos a suponer que se trata de la media, es decir, el momento estocástico de primer orden.

Para dar solución al problema nos vamos a apoyar en el principio de "Máxima entropía" de Jaynes. Dicho principio puede estudiarse en los siguientes términos: "La oportuna óptima estimación de la probabilidad, cuando se tiene en cuenta alguna información previa, no estadística o frecuencial, debe ser aquella que hace máxima la función de entropía".

La función de entropía, como sabemos, mide el grado de información que comporta todo fenómeno aleatorio cuando se realiza, y su expresión es

$$H = -\sum_{1}^{m} P_i \lg P_i$$

si se trata de un fenómeno discreto con m posibles sucesos. En el caso en que el fenómeno fuese continuo, tendríamos

$$H = -\int_a^b \lg f(x) \cdot f(x)\, dx$$

$$a \leqq x \leqq b$$

Se trata, pues, de hallar el máximo de H con las condiciones

$$\int_a^b f(x)\, dx = 1$$

$$\int_a^b x f(x)\, dx = \Theta_0$$

Mediante el cálculo de Variaciones llegamos a la solución

$$f(x) = \frac{1}{\Theta_0} e^{-x/\Theta_0}$$

Determinaremos la región crítica prepotente de tamaño α mediante la razón de máxima verosimilitud, que en este caso será

$$\lambda = \frac{\frac{1}{\Theta_0} e^{-x/\Theta_0}}{\frac{1}{x} e^{-1}} = \frac{x}{\Theta_0} e^{-1x/\Theta_0}$$

Se trata, pues, de una región crítica con dos ramas: $(0, x')$ y (x'', ∞)

$$\int_0^{x'} \frac{1}{\Theta_0} e^{-x/\Theta_0}\, dx = \int_{x''}^{\infty} \frac{1}{\Theta_0} e^{-x/\Theta_0}\, dx = \alpha$$

Aunque la región crítica de una sola rama no será centrada, podemos, cometiendo un despreciable error, considerar

$$\int_{x_0}^{\infty} \frac{1}{\Theta_0} e^{-x/\Theta_0}\, dx = \alpha$$

en cuyo caso

$$e^{-x_0/\Theta_0} = \alpha$$

luego

$$x_0 = \Theta_0 \mid \lg \alpha \mid$$

De aquí que todos los $x_r < x_0$ serán favorables a la hipótesis $\Theta = \Theta_0$ y los $x_s > x_0$ desfavorables.

La fiabilidad será estimada por $f = \dfrac{r + R}{n + N}$, siendo r el número de $x_i < x_0$.