AÑO 15. Nº 3-4. MAYO-JUNIO 1928

Sobre las frecuencias estadísticas de grupos ampliados

I

Supongamos una urna compuesta de m bolas (que en lo sucesivo llamaremos elementos), tales que por su color u otra convención cualquiera, conste de varias subclases de $|\alpha|, |\beta|, \ldots$ elementos cada una; siendo:

$$|\alpha| + |\beta| + \ldots\ldots = |m|$$

Indicando con p_α la probabilidad de un elemento cualquiera de la subclase α respecto a la urna de $|m|$ elementos, se desea obtener la P_α relativa a un elemento cualquiera de la subclase α cuando la clase m ha sufrido una agregación de $|n|$ elementos cualesquiera, esto es, indeterminados.

Servirá de esquema en el presente problema, una urna de la composición de la dada a la que se han agregado $|n|$ elementos extraídos al azar de otra urna de composición indeterminada (1).

Es casi obvio hacer ver cómo el problema de las urnas o grupos a composición agregada, se enlaza estrechamente con problemas de la Estadística. En la práctica se conoce una frecuencia por ejemplo $\frac{\alpha}{m}$ (que es una verdadera probabilidad cuando se considera la probabilidad para que un elemento de la clase m pertenezca a la subclase α) é interesa saber cuánto valdrá esta frecuencia cuando la clase m se ensancha con datos nuevos. El caso más general a este respecto de la agregación de elementos, es aquel en que se omite todo postulado referente

(1) Esta condición expresa es la que conforma el presente estudio. Se puede también, considerar la agregación de n elementos como producto de una extracción de otra urna de composición definida, fuera de nuestro plan y por demás, fácilmente calculable.

a su composición sobre cuya resolución nos ocuparemos inmediatamente.

II

Se trata de hallar aquí una expresión de P_α. De acuerdo con el enunciado del problema, los casos posibles son $m+n$. Si queremos ahora estudiar los casos que favorecerán a la subclase de los α, debemos observar (dado que ignoramos la calidad de los elementos agregados) que estos pueden repartirse respecto de α en la siguiente única forma:

$$0,1,2,3,\ldots\ldots\ldots,n.$$

Esto es, ninguno de los elementos de n es de la clase α, uno de los elementos de n es de la clase α, etc...

Todas estas formas son posibles y atento a ello, no podemos desechar ninguna en beneficio de otras (principio de la indiferencia de Laplace). Tomaremos pues, la media aritmética de ellas que es el valor más probable.

De acuerdo con las $n+1$ formas posibles, tendremos los siguientes casos favorables:

α, $\alpha+1$, $\alpha+2$, $\alpha+3$, ... $\alpha+n$, cuya media aritmética dá:

$$\frac{1}{n+1}\left[(n+1)\,\alpha+\frac{1+n.}{2}\,n\right] \text{ de donde: } P_\alpha=\frac{1}{n+1}\cdot\frac{(n+1)\alpha+\frac{1+n}{2}\,n}{m+n}$$

que simplificando conduce a:

$$P_\alpha=\frac{2\alpha+n}{2(m+n)}\cdot \quad (1)$$

III

Calculemos ahora, como es natural, la diferencia entre este valor de la probabilidad para la urna agregada con n elementos y la probabilidad $p_\alpha=\frac{\alpha}{m}\cdot$ Llamando esta diferencia con $\triangle_\alpha^n$ que leeremos diferencia de probabilidad de α para un agregado n, tendremos, por la (1).

$$\triangle_\alpha^n = P_\alpha - p_\alpha = \frac{2\,\alpha + n}{2\,(m+n)} - \frac{\alpha}{m} = \frac{2\,\alpha\, m + m\, n - 2\,\alpha\, m - 2\,\alpha\, n}{2\, m\,(m+n)} \;\therefore$$

$$\triangle_\alpha^n = \frac{m\, n - 2\,\alpha\, n}{2\, m\,(m+n)} \quad \text{o bien:} \quad \triangle_\alpha^n = \frac{n\,(m - 2\,\alpha)}{2\, m\,(m+n)} \qquad (2)$$

IV

Cualquiera que sea h (entero, por exigencia del problema estudiado), es facil demostrar que

$$\triangle_\alpha^{n+h} > \triangle_\alpha^n$$

En efecto:

$$\frac{\triangle_\alpha^{n+1}}{\triangle_\alpha^n} = \frac{[m\,(n+1) - 2\alpha\,(n+1)]\cdot[2m\,(m+n)]}{[2\, m\,(m+n+1)]\cdot[m\, n - 2\,\alpha\, n]} = \frac{m\,(n+1) - 2\,\alpha\,(n+1)}{m\, n - 2\,\alpha\, n} \times \frac{2\, m\,(m+n)}{2\, m\,(m+n+1)} =$$

$$\frac{(n+1)\,[m - 2\alpha]}{n\,(m - 2\alpha)} \times \frac{m+n}{m+n+1} = \frac{m\, n + n^2 + m + n}{m\, n + n^2 + n} = 1 + \frac{m}{m\, n + n^2 + n};$$

luego, finalmente:

$$\frac{\triangle_\alpha^{n+1}}{\triangle_\alpha^n} > 1. \quad \text{y en consecuencia:} \quad \triangle_\alpha^{n+h} > \triangle_\alpha^n$$

c.q.d.

V

De acuerdo con lo demostrado en el párrafo anterior, la sucesión $\triangle_\alpha^n$ es monótona creciente con respecto a n. Además es facil ver que estará acotada superiormente.

Para n = 0 tenemos $\triangle_{\alpha}^{o} = 0$ y cuando $n \to \infty$ (1)·

$$\lim_{n \to \infty} \triangle_{\alpha}^{n} = \lim_{n \to \infty} \frac{n(m-2\alpha)}{2m(m+n)} = \lim_{n \to \infty} \frac{m-2\alpha}{2m\left(\frac{m}{n}+1\right)} = \frac{m-2\alpha}{2m}$$

resultado valedero para un valor de m cualquiera, prefijado.

Luego, $\lim\limits_{n \to \infty} \triangle_{\alpha}^{n} = \frac{m-2\alpha}{2m}$ (3) expresa que la sucesión $\triangle_{\alpha}^{n}$ que es monótona creciente, está además superiormente acotada.

Ejemplo: Supongamos $\alpha = 500$ $m = 2000$ y $n = 1.000.000$, será: $\triangle_{500}^{1.000.000} = \frac{1.000.000\,(2000-2\times 500)}{2\times 2000\,(2000+1.000.000)} = \frac{1000}{4008}$ valor que podemos comparar con el

$$\lim_{n \to \infty} \triangle_{\alpha}^{n} = \frac{m-2\alpha}{2m} = \frac{2000-2\times 500}{2\times 2000} = \frac{1000}{4000}$$

lo cual arroja la siguiente diferencia entre el

$\lim\limits_{n \to \infty} \triangle_{\alpha}^{n}$ y $\triangle_{500}^{1000000}$:

$$\frac{1000}{4000} - \frac{1000}{4008} = 0{,}000499.$$

Resulta un caso particular de este problema aquel en el cual $P_{\alpha} = p_{\alpha}$ y de consiguiente $\triangle_{\alpha}^{n} = 0$

Ello se verifica cuando $\alpha = 0{,}5 \times m$, y $n = m$, por cuanto entonces tenemos: $p_{\alpha} = \frac{\alpha}{m} = 0{,}5$ $P_{\alpha} = \frac{2\alpha + n}{2(m+n)} = \frac{2\times m}{4m} = 0{,}5$ de donde por diferencia directa o aplicando la (2), resulta:

$$\triangle_{0.5\,m}^{m} = \frac{m^2 - m^2}{4m^2} = 0 \text{ c.q.d.}$$

VI

El mismo valor $P_{\alpha} = 0{,}5$ se obtiene en el límite cuando $n = \infty$, y m y α, tienen valores asignados previamente.

---

(1)· El valor de n debe necesariamente, por exigencia del problema, ser entero. Ello no entraña impedimentos para que el límite tenga sentido, pues siendo cierto para los valores reales, lo será para los enteros que en ellos se incluyen.

En efecto: $\lim\limits_{n \to \infty} P_\alpha = \lim\limits_{n \to \infty} \dfrac{2\alpha + n}{2(m+n)} = 0{,}5$. Este resultado es intuitivamente evidente, ya que fijados α y m y para un agregado de $n = \infty$ elementos, nos hallamos sin poder discernir nada, y por tanto P_α que expresa la probabilidad para la salida de α se equilibra con la probabilidad contraria, sin tener ninguna de estas dos hipótesis mayor peso que la otra. Representa, pues, aquí el valor 0,5. la «duda práctica».

VI

Es sabido que en las investigaciones estadísticas, interesa calcular predicciones en base a ciertas frecuencias observadas, aceptándose el valor de éstas para grupos ampliados. La suposición aquí contenida, equivale a dar por establecida la igualdad siguiente:

$$p_\alpha = P_\alpha$$

Examinemos, pues, bajo qué condiciones ella se presenta. Supongamos que α y m están dados por la relación $\alpha = k.\, m$ siendo $0 < k < 1$.

En tal hipótesis resultan los dos siguientes resultados concordantes:

$$\lim_{m \to \infty} \triangle_\alpha^n = \lim_{m \to \infty} \frac{m.n - 2\alpha n}{2m(m+n)} = \lim_{m \to \infty} \frac{m.n - 2k\, m.n}{2m(m+n)} =$$

$$\lim_{m \to \infty} \frac{n - 2kn}{2(m+n)} = 0 \quad \text{y}$$

$$\lim_{m \to \infty} P_\alpha = \lim_{m \to \infty} \frac{2\alpha + n}{2(m+n)} = \lim_{m \to \infty} \frac{2k}{2} = k = p_\alpha$$

de lo cual se sigue que, para valores de m tanto más grandes $\triangle_\alpha^n$ es tanto más pequeña, haciéndose así posible mantener el valor p_α al caso de un conjunto agregado. He aquí precisamente la condición de los grandes números recabada de las observaciones cuando las frecuencias de las mismas se desea equipararlas a probabilidades de los fenómenos en consideración.

Carlos Dieulefait
Prof. Supl. Universidad Litoral