NUMEROS, 18 - 1988

PROPORCIONALIDAD Y PROBABILIDAD

Luis A. Santaló
Universidad de Buenos Aires.

En los problemas clásicos, la proporcionalidad aparece como una relación exacta, en el sentido de que compara magnitudes bien determinadas y con medidas que se supone conocidas exactamente. Es la manera como opera la llamada "regla de tres" de la escuela elemental. Así, en el movimiento uniforme, el espacio recorrido durante un tiempo fijo, es proporcional a la velocidad y, para una velocidad determinada, es proporcional al tiempo. También, el precio de una determinada mercadería es proporcional a la medida de la misma (longitud, si se trata de telas o alambres; peso, si se trata de azúcar o patatas; volumen, si de líquidos como vino o aceite). En las clases de nivel medio, conviene poner abundantes ejemplos de magnitudes proporcionales, como las que acabamos de mencionar, y otros de las que no lo son (espacio recorrido por un cuerpo que cae y el tiempo transcurrido, crecimiento de una población y el tiempo transcurrido, etc.). En general, es muy conveniente hacer la representación gráfica de una magnitud -de su medida- en función de la otra, para ver si es o no una recta.

La cosa es distinta cuando una de las magnitudes, o las dos, están determinadas por un procedimiento al azar. Un ejemplo típico es el siguiente. Supongamos un volumen V de un líquido que contiene ciertas partículas en suspensión (virus o bacterias, por ejemplo). Se toma, al azar, una pequeña porción de líquido (una muestra). Supongamos que sea de-volumen v, y se cuenta el número de partículas que contiene, sea n. Si N es el número total de partículas, se acepta la proporcionalidad

$$n / v = N / V \qquad (1)$$

es decir, la proporcionalidad entre los volúmenes y los números de partículas que contienen. Esto permite calcular el número total de partículas V, o el número de ellas por unidad de volumen. Así se procede, por ejemplo, para medir el número de globulos rojos de la sangre, por centímetro cúbico, a partir de su número en una gota o en una muestra cualquiera ; o bien, el número de bacterias por unidad de volumen del agua de una laguna o de un río. Sin embargo, aunque se supone que las partículas están "uniformemente" distribuidas en el volumen total V, lo que equivale a postular la proporcionalidad mencionada, el número n no es exacto, sino que depende de la muestra elegida. Es decir, interviene el azar. Por tanto, la igualdad (1) no tiene el sentido de una igualdad usual, pues el segundo miembro es una constante bien determinada, mientras que el primero puede variar con la muestra de volumen v que se elija.

La interpretación exacta de (1) es que, para cada volumen v, el número n es una variable aleatoria que depende de v, la cual tendrá una cierta esperanza matemática $\mathcal{E}(n)$. La ecuación (1) quiere decir que se cumple la igualdad

$$\mathcal{E}(n) = (N/V)v \qquad (2)$$

o sea, poniendo $(N/V) = k$ (constante), que se cumple

$$\mathcal{E}(n) = kv \qquad (3)$$

Si n puede tomar un solo valor para cada v, será

$$\mathcal{E}(n) = n$$

y la forma (3) es la de proporcionalidad en el sentido usual $n = kv$.

Veamos de ordenar todo esto en una definición general. Sea x una variable real (no aleatoria) que puede variar en un cierto dominio.- Para cada valor de x supongamos definida una variable aleatoria $y(x)$ que puede tomar, con cierta probabilidad, determinados valores. Sea $\mathcal{E}(y)$ la esperanza matemática de y para cada x. Si se cumple

$$\mathcal{E}(y) = kx \qquad (4)$$

siendo k una constante,diremos que la variable aleatoria y y la variable no aleatoria x son proporcionales. Si para un cierto valor $x=x_0$ la variable aleatoria y puede tomar un solo valor $y=y_0$, entonces $E(y(x_0))=y_0$ y (4) nos dice que el valor de la constante k es el cociente y_0/x_0.

Vamos a considerar algunos ejemplos :

Ejemplo 1.

Se desea calcular el número de peces N de una laguna. Para ello,se pescan m peces,los cuales se marcan de cierta manera y se arrojan nuevamente,con vida,al agua. Después de transcurrido un cierto tiempo,para que todos los peces,marcados y no marcados,se hayan mezclado,se pesca un lote de n peces,de los cuales resulta que hay m_1 marcados. Con estos datos, ¿qué puede decirse sobre el número buscado N?

El número de peces n de la muestra extraída corresponde a la variable x anterior,y el número m_1 de peces marcados corresponde a la variable aleatoria que hemos indicado por $y=y(x)$. Queremos calcular $E(m_1)$. Si a cada pez de la muestra n asignamos la variable aleatoria y_i $(i=1,2,\ldots n)$,que vale 1 si el pez está marcado y 0 si no lo está,será

$$m_1 = y_1 + y_2 + \ldots + y_n$$

Puesto que la esperanza matemática de una suma es igual a la suma de las esperanzas matemáticas y $E(y_1)$ es la probabilidad de que un pez esté marcado,y por tanto vale m/N,resulta

$$E(m_1) = nm/N$$

Como m/N es una constante,independiente de la muestra,se cumple (4) y,en consecuencia,la variable aleatoria m_1(número de peces marcados de la muestra) y la variable no aleatoria n(número de peces de la muestra) son proporcionales.

Se suele decir que m_1/m es un *estimador* de m/N y se escribe

$$(m/N)^* = m_1/n \qquad (5)$$

Se trata de un estimador insesgado,puesto que su esperanza matemática $E(m_1)/n$ coincide con m/N. Esto permite estimar el número total de peces por la fórmula

$$N^* = mn/m_1 \qquad (6)$$

((Para aclaraciones sobre el método y abundante bibliografía relativa a este ejemplo, se puede ver *Azorín Poch, F.-Curso de Muestreo y Aplicaciones-Instituto Nacional de Estadística-2ª edición-Madrid, 1962*, principalmente los capítulos 7 y 27. Otros detalles están en el libro de *W. Feller "An Introduction to Probability and its Applications", vol. 1-2.-Wiley-New York, 1950-pág. 37*, que vincula el problema con la llamada distribución hipergeométrica, cuya esperanza es la $E(m_1) = mn/N$ mencionada. *Feller* considera el ejemplo en que $m=1000$, $n=1000$ y $m_1=100$, con lo cual resulta $N^* = 10.000$)).

Para comprobar el método seguido en el ejemplo 1, para estimar números grandes de objetos, se puede hacer la siguiente experiencia. Se toma un montón relativamente grande de palillos, fósforos o granos cualesquiera. Se marca un número m de ellos. Se devuelven al montón primiti y se mezclan todos bien. Luego, se saca, al azar, una muestra y se cuenta exactamente cuántos objetos la forman; supongamos que sean n. Si entre ellos hay m_1 de los marcados, el valor estimado del número total N nos lo da la fórmula *(6)*. Por ejemplo, si $m=80$, $n=60$ y $m_1=5$, es $N^* = 960$.

Ejemplo 2

En una población de N habitantes se desea estimar el número de ellos que son fumadores. Se toma una muestra de n personas elegidas, para que se pueda suponer que están uniformemente repartidas en el total N, en diferentes barrios, de diferentes edades, sexos y ocupaciones. Si el número de fumadores de la muestra es m_1, las mismas consideraciones del ejemplo anterior conducen a la proporcionalidad

$$E(m_1) = mn/N$$

entre el número de fumadores m_1 y el número de individuos n de la muestra. Esto permite tomar como estimador insesgado del número buscado m a

$$m^* = m_1 N/n \qquad (7)$$

Este método es la base de las encuestas de opinión para predecir resultados electorales, realizar análisis de mercados o evaluar la audiencia de programas de televisión.

Ejemplo 3

Dentro de un cuadrado de lado a se tiene una figura H cuya - área F se quiere determinar. Para ello se toman m puntos al azar en el cuadrado y supongamos que m_1 de ellos hayan caído dentro de H. El caso de los puntos que caen en el contorno de H es de probabilidad nula y - puede no ser tenido en cuenta. La probabilidad de que un punto dado al azar en el cuadrado caiga dentro de H, es igual al cociente de las áreas (F/a^2), lo que equivale a admitir que los puntos del cuadrado son igualmente probables y que han sido elegidos con probabilidad uniforme. Por tanto, la probabilidad de que dentro de H hayan caído m_1 puntos del total de los m puntos elegidos, responde a la ley binomial con probabilidad F/a^2. Luego la esperanza matemática de m_1 es $E(m_1) = (F/a^2)m$. Es decir, m_1 y F son proporcionales, según la definición *(4)*. Y, en consecuencia, se tiene para F el estimador

$$F^* = m_1 a^2 / m \qquad (8)$$

Para realizar prácticamente este ejemplo en la clase, lo que es muy instructivo, conviene que todos los alumnos dispongan de una tabla de números al azar de unos 500 dígitos. Se puede tomar de algún texto de Probabilidades o de Estadística; por ejemplo, del libro de *Sixto Ríos, "Métodos Estadísticos"-Mc. Graw Hill-New York, 1967-págs. 225 y 226*. Existen muchos métodos para obtener números al azar, como los mencionados en *Les Probabilités a l'école, de M. Glayman y T. Varga-CEDIC, 1973 (pág. 158 de la traducción castellana de Ed. Teide.)*. También pueden tomarse de alguna computadora que los suministre.

Es importante que las tablas de cada alumno sean distintas, para evitar así que se copien los resultados y, sobre todo, porque al tomar el promedio la aproximación obtenida será mayor.

Tomemos entonces un cuadrado de 10 cm de lado, de papel milimetrado o de cuadriculado fino. Para tomar puntos al azar en el cuadrado, se toman cuaternas de números al azar de la tabla, de manera sucesiva. - Por ejemplo, la tabla de Sixto Ríos comienza con las cuaternas

2034 , 5600 , 2400 , 7583 , 1104

y los sucesivos puntos que se van tomando son los de coordenadas, en centímetros, $(2,0 : 3,4)$, $(5,6 : 0,0)$, $(2,4 : 0,0)$, $(7,5 : 8,3)$, $(1,1 : 0,4)$, ...y así sucesivamente hasta los 100 puntos (400 números de la tabla). De estos puntos, se cuentan los m_1 que han caído dentro de la figura H y, según (8) (para $a=10$ cm , $m=100$) el número m_1 será el área estimada de H en cm^2.

Como comprobación del método se puede tomar una figura de área conocida y comprobar el resultado que, naturalmente, no será exacto pero, en general, dará una bastante buena aproximación. Tomando valores mayores de m, se obtendría una mayor aproximación. Por ejemplo, tomando en el cuadrado de *10 cm* de lado, un círculo de radio *3 cm*, en cualquier posición, con las primeras 100 cuaternas de números al azar de las tablas se obtienen 100 puntos, y el número de ellos que resultan interiores al círculo, casi seguro que no será muy diferente de *28*, que es el área del círculo de radio *3* (en número entero de cm^2). Haciendo la experiencia tomando como centro del círculo el punto $(6,6)$ y usando las tablas de Ríos, se obtiene $m_1=29$ -con cierta duda acerca de dos o tres puntos del contorno-, resultado bastante aproximado.

Ejemplo 4 .- Determinación de π por el azar

Teniendo en cuenta que el área del círculo de radio *3* es 9π , del resultado anterior se deduce que un valor estimado de π es $29/9 \approx 3,22$. Tomando un número más grande de puntos dentro del cuadrado, el valor resultante sería más aproximado.

El método se puede aplicar a cualquier figura H cuya área se exprese mediante π y, entonces, la fórmula (8) permite despejar esta constante. Tomando, por ejemplo, el cuadrante de círculo de centro un vértice del cuadrado y radio igual al lado, un valor estimado de π resulta ser $\pi^* = 4m_1 / m$. Ver, por ejemplo, el libro citado de *Glayman y Varga*, págs. *199-202*.

El método de hallar ciertas constantes, y aun ciertas funciones, utilizando una simulación por el azar, se denomina *método de Monte Carlo* y, sobre todo con la ayuda de una computadora que permita simular un número grande de experiencias, tiene muchas aplicaciones.

Otros ejemplos

Hay ejemplos, cuya fundamentación teórica no es tan simple, pero que vamos a mencionar para el lector interesado en estas cuestiones. Pueden verse, por ejemplo, en *Luis A. Santaló-Integral Geometry and Geometric Probability-Addison Wesley, Reading- U.S.A., 1976-Cap. 8.*

Ej. 5

Supongamos el reticulado formado por rectángulos congruentes de lados a, b que cubren todo el plano, es decir, consideremos el plano dividido en rectángulos por dos haces de rectas paralelas y perpendiculares entre sí, a disancias respectivas a y b.

Se lanza una curva de longitud L al azar sobre el plano y se cuenta el número de puntos de intersección-sea n-con las rectas del reticulado. La forma de la curva puede ser cualquiera e incluso puede ser un hilo deformable cuya forma sea variable en cada experiencia, mientras mantenga la longitud L. Se puede demostrar que el valor medio o esperanza matemática del número n vale

$$\mathcal{E}(n) = 2(a + b)\, L / \pi a b \qquad (9)$$

Según la definición (4) esto significa que la longitud L es proporcional a la variable aleatoria n. Y (9) nos dice que se puede tomar como estimador de n el valor

$$n^{*} = 2(a + b)\, L \;/\; \pi a b$$

o bien calcular la longitud estimada de L por la fórmula

$$L^{*} = \pi a b n \;/\; 2(a + b)$$

Experimentalmente se puede calcular n^{*} haciendo la experiencia un número grande de veces y tomando el valor medio del número de puntos de intersección de la curva con el reticulado de rectángulos.

Considerando el caso $a \to \infty$, el reticulado se reduce al plano dividido por un haz de paralelas a distancia b y entonces es

$$\mathcal{E}(n) = 2L \;/\; \pi b$$

En particular, si la curva es un segmento de longitud $L \leq b$, el número n sólo puede tomar los valores 1 ó 0 y, por tanto, $\mathcal{E}(n)$ es la probabilidad de que el segmento corte a alguna paralela. Es el resultado clásico del llamado *problema de la aguja de Buffon*. También en este caso, como se menciona en muchos textos clásicos de Probabilidades, se puede calcular π por el azar, tomando

$$\pi^* = 2L / b\mathcal{E}(n)$$

y calculando experimentalmente $\mathcal{E}(n)$ lanzando al azar el segmento o aguja repetidas veces sobre el haz de paralelas, y calculando el valor medio del número de veces que corta a alguna paralela.

En términos de la teoría de juegos de azar, podemos considerar el de lanzar al azar una curva de longitud L y forma cualquiera sobre el reticulado de rectángulos de lados a, b y convenir que el jugador cobrará como premio tantas unidades monetarias como puntos de intersección n de la curva con el reticulado. Para que el juego sea equitativo, el valor de la puesta debe ser igual al valor medio $\mathcal{E}(n)$ dado por *(9)*.- Si se cobra más, el juego es favorable a la banca; si menos, lo es al jugador.

Por ejemplo, si el reticulado está formado por cuadrados de lado unidad, la puesta debe valer

$$\mathcal{E}(n) = 4L\pi$$

cualquiera que sea la forma de la curva.

En el caso de *Buffon* del reticulado de rectas paralelas a distancia b, la puesta debe valer

$$\mathcal{E}(n) = 2L / \pi b$$

Ej. 6

Supongamos el plano cuadriculado por cuadrados de lado a. Se lanza al azar sobre el mismo una aguja de longitud $L \leqslant a$ (para que sólo pueda cortar al reticulado en 2 puntos a lo sumo). Se desea saber las probabilidades P_i de que la aguja corte al reticulado exactamente en i puntos ($i=0,1,2$). Se puede demostrar que estas posibilidades valen, respectivamente

$$P_0 = 1 - 4L / \pi a + (L^2 / \pi a^2)$$
$$P_1 = 4L / \pi a - 2L^2 / \pi a^2$$
$$P_2 = L^2 / \pi a^2$$

Es instructivo hacer la comprobación experimental de estos resultados. El valor medio del número de puntos de intersección será, de acuerdo con (9),

$$\mathcal{E}(n) = 0.P_0 + 1.P_1 + 2.P_2 = 4L / \pi a$$

Como juego de azar, los resultados anteriores pueden interpretarse de distintas maneras, a saber:

a) Si el jugador recibe un premio P cada vez que la aguja lanzada tenga con el reticulado i puntos de intersección, para cada valor de i fijado de antemano ($i=0,1,2$), la puesta debe valer PP_i.

b) Si el jugador recibe de premio tantas unidades como puntos de intersección logre, el valor de la puesta debe ser $\mathcal{E}(n) = 4L / \pi a$, que hemos visto es el mismo cualquiera que sea la longitud L y la forma de la curva lanzada.

Buenos Aires, Mayo de 1988