Las probabilidades y sus aplicaciones a la estadística

PROF. DGO. ALMENDRAS

Con este artículo empezamos una serie de otros que servirán como introducción a la estadística matemática. Los conocimientos básicos que se necesitan para la comprensión de esta materia serán dados simultáneamente. Se supone el conocimiento de los teoremas fundamentales del cálculo de probabilidades y del análisis combinatorio.

PROBLEMA DE LAS PRUEBAS REPETIDAS. — Se distinguen dos casos denominados de Bernouilli y de Poisson.

Caso de Bernouilli. —Supondremos una urna que contiene N_1 bolitas de clase A y N_2 de clase B, cada una de las cuales tiene la misma probabilidad de salir cuando se saca una al azar. Designando por p y q las probabilidades de sacar una de clase A y una de clase B, tendremos:

$$p = \frac{N_1}{N_1+N_2}$$

$$q = \frac{N_2}{N_1+N_2}$$

Si se saca n veces consecutivas una bolita reponiéndola después de cada salida, se pide la probabilidad para que k de ellas sean de clase A. Tal es el problema de las pruebas repetidas en el caso de Bernouilli. Un orden de salida podría ser B, B, B,. . . . . .A B, A BBA en que figuran k de clase A y n-k de clase B. La probabilidad para que ocurra este orden de salida será $p^k \, q^{n-k}$ (teorema de las probabilidades compuestas). Como el número de permutaciones distintas que resultan entre n elementos en que k son de una clase y n-k son de otra es:

$$\frac{n!}{k!\,(n-k)!} = C_n^k$$

por la aplicación del teorema de las probabilidades totales resulta que la probabilidad para que salgan k de clase A y n-k de clase B, en cualquier orden es:

$$1) \quad p_k = C_n^k \, p^k \, q^{n-k}$$

El problema de la extracción de una bolita con reposición en una urna determinada se puede reemplazar por el de la extracción de una sola bolita de de cada una de n urnas idénticas. Tal suposición es muy útil para la resolución de problemas más complicados.

Si en una urna hay bolitas de tres clases A, B y C y con probabilidades p, q y r de salir en cada extracción con reposicióu, en una extracción de n sucesivamente, la probabilidad de que salgan números k, h y l de cada clase será

$$\frac{n!}{k!\,h_{\text{i}}\,l!} \; p^k \, q^h \, l^r$$

Distribución binomial.—Puede observarse que la expresión $P_k = C_n^k \, p^k \, q^{n-k}$ es un término del desarrollo $(q+p)^n$, el cual se puede poner en la forma:

$$(q+p)^n = P_0 + P_1 + P_2 + \ldots . + P_k + \ldots . + P_n = 1$$

Si se realiza una experiencia en una urna haciendo n extracciones sucesivas y con reposición, el resultado obtenido podrá anotarse en una serie lineal, en la forma: BABBBAABA..........BABAAAB (en total n bolitas). Como esta experiencia se puede repetir N veces, siendo N tan grande como se quiera, podrá hacerse una estadística de todos los resultados obtenidos en un cuadro como el que sigue: en la primera columna se anotan los valores que puede tomar la variable aleatoria x, que representa el número de veces que puede salir A en cada experiencia; x varía de 0 a n; en la segunda columna se anotará las veces (frecuencia) que se repite la variable x en el número N de experiencias. Las frecuencias se designarán por f_0, f_1, ..., f_n. Conviene no olvidar que cada una de las N experiencias consiste de n extracciones de una bolita y con reposición.

N.° de A	f
0	f_0
1	f_1
2	f_2
:	:
n	f_n

Cuadro I

N.° de caras	frecuencia	N.° total de caras
0	56	0
1	196	196
2	154	308
3	54	162
	460	666

Cuadro II

En un caso práctico se lanzaron tres monedas (efigie de O'Higgins y Copihues) 460 veces siendo el caso favorable la salida de cara. En esta experiencia tenemos: x = 0, 1, 2, 3 (números de caras que puede salir en cada experiencia); $n = 3$; $N = 460$. El cuadro adjunto contiene los datos obtenidos. Las frecuencias relativas representan, aproximadamente, las probabilidades de salida de 0 cara; 1 cara; 2 caras y 3 caras. Estas frecuencias relativas tienden a valores más significativos si el número N de experiencias aumenta y diremos que tienden a las verdaderas probabilidades de salida de cara 0, 1, 2 ó 3 veces.

Si una persona no conoce estas probabilidades podría resolver el problema de calcular la probabilidad para que al lanzar 3 monedas resulte 0 cara, 1 cara, 2 caras o 3 caras y procederá matemáticamente admitiendo $p = q = 0{,}5$.

Las probabilidades para estos cuatro casos son los términos del desarrollo;
$(q + p)^3 = \frac{1}{8} + \frac{3}{8} + \frac{3}{8} + \frac{1}{8}$ y las frecuencias que habrían resultado al lanzar 460 veces tres monedas serían 58;172;172;58 en lugar del resultado experimental 56;196;154;54.

Se puede constatar que, de acuerdo con la experiencia, las probabilidades de salida de cara y sello no son iguales. Aunque el número de experiencias no es muy grande, podemos obtener una estimación buena de las probabilidades verdaderas. En efecto, lanzar 460 veces tres monedas es equivalente a lanzar 1.380 veces una sola y luego bastará saber cuantas caras se obtuvieron en estos lanzamientos para calcular la probabilidad a *posteriori*. En la tercera columna del cuadro II aparece el número de caras, que en total es 666, entonces la probabilidad a *posteriori* de salida de cara se puede estimar igual
$p' = \frac{666}{1380} = 0{,}483$. La probabilidad de salida de sello será $q' = 0{,}517$.

Las probabilidades de salida de 0 cara; 1 cara; 2 caras; 3 caras, con estos nuevos valores q' y p' son los términos del desarrollo;

$$(0{,}517+0{,}483)^3 = 0{,}138+0{,}387+0{,}362+0{,}113$$

En el cuadro III figuran las frecuencias obtenidas de la experiencia y las correspondientes a las distribuciones binomiales con las probabilidades a *posteriori* y en la hipótesis de $q = p = 0{,}5$. Vemos que se ajustan mejor a la experiencia los valores deducidos de las probabilidades estimativas q' y p'.

N.º de caras	f experimental	f $p'=0{,}483$ $q'=0{,}517$	f $p=0{,}5$ $q=0{,}5$
0	56	63	58
1	196	178	172
2	154	167	172
3	54	52	58
	460	460	460

Cuadro III

N.º de machos	frecuencia f	N.º total de machos	f ajustada
0	2	0	4
1	20	20	20
2	41	82	37
3	35	105	35
4	14	56	17
5	4	20	3
	116	283	116

Cuadro IV

EJEMPLO TOMADO DE APPLIED GENERAL STATITICS DE CROXTON AND COWDEN.

«De 116 pariciones de 5 chanchitos, siendo la variable el número de machos, se obtuvo el siguiente cuadro de frecuencia»:

Para representar esta distribución por una binomial debemos calcular un valor estimativo de la probabilidad de nacimiento macho. En la columna 3 tenemos el total de macho, que es 283; el total de nacimientos es 5.116 = 580.

Pondremos $p' = \frac{283}{580} = 0{,}4879$; $q' = 0{,}5121$.

Las probabilidades correspondientes a la distribución binomial son los términos del desarrollo $(0{,}5121 + 0{,}4879)^5$. Multiplicando los términos de este desarrollo por el número de pariciones, 116, obtendremos el ajuste a la distribución binomial. Estas frecuencias figuran en la columna 4 del cuadro anterior.

Caso de Poisson.—Este caso queda planteado en el siguiente problema: «De una urne que contiene N bolitas de las cuales *a* son blancas y *b*, negras, se sacan n de ellas al azar y se pide la probabilidad para que x de ellas sean blancas».

Para resolver este problema, pongamos:

$$a = Np$$
$$b = Nq$$

El número de casos posibles es C_N^n y el número de casos favorables es $C_a^x \cdot C_b^{n-x}$, luego la probabilidad pedida es:

$$P_x = \frac{C_a^x \, C_b^{n-x}}{C_N^n}$$

Para tener un caso real análogo al de Poisson, sería necesario hacer un número grande de experiencias sacando, simultáneamente, n bolitas y reponiéndolas todas, después de cada una de ellas. Finalmente, sería necesario hacer un cuadro estadístico en que en la primera columna se anotara la variable aleatoria, que puede tomar los valores x = 0, 1, 2, 3 . . . . . n y en la otra columna se anotará la frecuencia respectiva.

MOMENTOS DE PRIMER Y SEGUNDO ORDEN EN EL CASO DE BERNOUILLI Y POISSON

En general, los momentos de las distribuciones de Bernouilli y Poisson, son números que tienen una importancia fundamental en las aplicaciones del cálculo de probabilidades a la estadística, especialmente en la teoría del muestreo (sampling). En lo que sigue, nos limitaremos a la determinación de los momentos de primer y segundo orden.

Definición.—En distribuciones como las representadas en los cuadros III y IV, estando en la primera columna los valores que puede tomar la variable aleatoria y en la segunda columna las frecuencias respectivas, se llama momento de orden p de la variable aleatoria x al número:

$$(1) \quad m_p = \frac{\Sigma f_i x_i^p}{F} \qquad \text{siendo } F = f_1 + f_2 + f_3 + {}_4 + \ldots\ldots\ldots + f_n$$

El medio aritmético de la variable x, que designaremos por $\bar{x}$ o m_1 es el momento de primer orden.

Si los valores de la variable aleatoria se miden a partir del valor m_1, los momentos de orden p de la nueva variable se designarán por μ_p.

Evidentemente, tenemos:

$$\mu_p = \frac{\Sigma f_i (x_i - \bar{x})^p}{F} = m_p - C_p^1 m_{p-1} \bar{x} + C_p^2 m_{p-2} \bar{x}^2 \mp \ldots + (-1)^p \bar{x}^p$$

Para p = 1, tenemos:

$$\mu_1 = m_1 - \bar{x} = 0$$

Para p = 2, se tiene:

$$\mu_2 = m_2 - 2 m_1 \bar{x} + \bar{x}^2$$

$$\mu_2 = m_2 - \bar{x}^2$$

Si la distribución de la variable aleatoria x es equivalente a una de Bernouilli o Poisson, las frecuencias relativas $\frac{f_0}{F}; \frac{f_1}{F}; \frac{f_2}{F}; \ldots; \frac{f_n}{F}$ son iguales a las probabilidades P_0 ; P_1 ; P_2 ; ......; P_n y luego tendremos:

$$(2) \quad m_p = \frac{\Sigma F P_i x_i^p}{F} = \sum_{i=0}^{n} P_i x_i^p$$

Y como caso especial;

$$m_1 = x = \sum_{i=0}^{n} P_i x_i = \sum_{x=0}^{n} x P_x$$

(Continuará)