SOBRE UNA DISTRIBUCION DISCRETA

Por

A. GIL CRIADO

Es bien conocido que todo experimento aleatorio que da lugar a dos únicos sucesos exclusivos, «éxito» o «fracaso», puede describirse mediante una variable aleatoria ξ

$$\xi = \begin{cases} 1 \text{ si ocurre «éxito»} \\ 0 \text{ si ocurre «fracaso»} \end{cases}$$

Si dicho experimento se repite en n pruebas independientes y llamamos ξ_r el valor de la variable correspondiente a la prueba r ésima, la variable aleatoria $\nu = \xi_1 + \xi_2 + \cdots + \xi n$ (número de éxitos en las n pruebas) tiene como función de probabilidad ($f.\ p.$):

$$P\ (\nu = r) = \binom{n}{r} \cdot p^r\ q^{n-r}, \quad (r = 0,1, \cdots n) \tag{1}$$

donde p representa la probabilidad de obtener «éxito» y $q = 1 - p$.

La distribución de la variable ν o «distribución binomial» de parámetros (n,p), es sin duda una de las más importantes en las aplicaciones estadísticas y de hecho será utilizada como base del presente estudio (*).

Si dos observadores realizan por separado n pruebas del experimento descrito anteriormente y llamados ν y ν', respectivamente, el número de éxitos obtenidos es fácil comprobar que la f. p. de la variable aleatoria $\nu - \nu'$ viene dada por la expresión·

$$P(\nu - \nu' = r) = \sum_{i=0}^{n-r} p^{2i+r}\ q^{2(n-i)-r} \binom{n}{i} \binom{n}{i+r} \tag{2}$$

(*) Para un estudio más a fondo de la distribución binomial y otras cuestiones relacionadas con ella, véase *Métodos Matemáticos de Estadística*, de H Cramer, «Aguilar, S A.» Madrid, 1953 para $r=0,1,2,\dots,n$, mientras que para los valores negativos de la variable:

$$P(\nu - \nu' = -r) = P(\nu - \nu' = r)$$

A continuación demostraremos la monotonía de la f. p. (2) en los puntos $r=0,1,2,\dots,n$ (la monotonía en los puntos de abscisa entera negativa se prueba de una manera totalmente análoga por la simetría de la función).

TEOREMA I: *Si* ν *y* ν' *son dos variables independientes que tienen como f. p. la binomial de parámetros* (n,p), p $\leqslant$q *entonces;*

$$P(\nu - \nu' = r) \quad < \quad P(\nu - \nu' = r-1) \quad , \quad (r = 1,2, \dots,n)$$

En efecto:

$$P(\nu - \nu' = r) - P(\nu - \nu' = r - 1) =$$

$$= \sum_{i=o}^{n-r} p^{2i+r} \cdot q^{2(n-i)-r} \binom{n}{i} \cdot \binom{n}{i+r} -$$

$$- \sum_{i=o}^{n-r+1} p^{2i+r-1} \cdot q^{2(n-i)-r+1} \binom{n}{i} \cdot \binom{n}{r-1+i} =$$

$$= \sum_{i=o}^{n-r} p^{2i+r-1} \cdot q^{2(n-i)-r} \binom{n}{i} \left\{ p \binom{n}{i+r} - q \binom{n}{i+r-1} \right\} -$$

$$- p^{2n-r+1} \cdot q^{r-1} \binom{n}{n-r+1} \binom{n}{n} \leqslant \sum_{i=o}^{n-r} \binom{n}{i} \left\{ \binom{n}{i+r} - \binom{n}{i+r-1} \right\} < 0$$

para $r = 1,2,\dots,n.$

Hay que hacer notar la simetría de la f. p. (2) en los parámetros p y q, de manera que la hipótesis $p \leqslant q$ no resta generalidad al teorema.

Puesto que la f. p. (2) alcanza su máximo en $r=o$, la distribución es simétrica y unimodal coincidiendo la moda con la *media;*

$$E[\nu - \nu'] = E[\nu] - E[\nu'] = np - np = o$$

Teniendo en cuenta la independencia de ν y ν' se determina la *varianza;*

$$E[\nu - \nu']^2 = E[(\nu - np) - (\nu' - np)]^2 =$$

$$= E[\nu - np]^2 + E[\nu' - np]^2 - E[2(\nu - np)(\nu' - np)] =$$

$$= E[\nu - np]^2 + E[\nu' - np]^2 - 2\, E[\nu - np] \cdot E[\nu' - np] = 2\, npq$$

Estamos interesados en estudiar la distribución de la variable aleatoria $|\nu - \nu'|$ porque, a pesar de guardar estrecha relación con la distribución definida en (2), presenta peculiaridades ciertamente interesantes.

De lo anterior se deduce fácilmente la f. p. de la variable $|\nu - \nu'|$

$$\alpha(r\,;\,n,p) = P(|\nu - \nu'| = r) = 2\sum_{i=o}^{n-r} p^{2i+r}\cdot q^{2(n-i)-r}\cdot\binom{n}{i}\binom{n}{r+i}$$

si $r = 1,2,\ldots,n$

y $$\alpha(o\,;\,n,p) = P(|\nu - \nu'| = o) = \sum_{i=o}^{n} p^{2i}\cdot q^{2(n-i)}\binom{n}{i}\binom{n}{i}$$

si $r = 0$ (3)

La función (3), como la (2), es monótona creciente en los puntos $r=1,2,3,\ldots,n$, es decir $\alpha(r+1) < \alpha(r)$, $r \neq 0$

Sin embargo, no siempre alcanza el valor máximo en $r = o$, dependiendo del valor p (probabilidad de obtener «éxito» en el experimento aleatorio). Esta propiedad viene precisada en el siguiente:

TEOREMA 2: *Sean* ν *y* ν' *dos variables independientes distribuidas según la f. p.* (1). *Supuesto* $p \leqslant q$, $p \neq 0$ *se puede determinar un intervalo* $[1/2 - \varepsilon, 1/2 + \varepsilon]$ *de valores de* p *y* q *tal que la función definida en* (3) *tiene la moda en el punto* $r=1$, *cualquiera que sea el valor del parámetro* n.

En efecto:

Distinguiremos dos casos según que n (número de tiradas) sea par o impar.

Caso I; número de tiradas impar, $2n+1$

Haciendo $r=1$ en (3)

$$\left.\begin{aligned}\alpha(1\,;\,2n+1\,,\,p) = {} & 2\,p^{1}\,q^{4n+1}\binom{2n+1}{0}\binom{2n+1}{1} + \\ & + 2\,p^{3}\,q^{4n-1}\binom{2n+1}{1}\binom{2n+1}{2} + \\ & \cdots\cdots\cdots\cdots\cdots\cdots \\ & + 2\,p^{2n+1}q^{2n+1}\binom{2n+1}{n}\binom{2n+1}{n+1} + \\ & + 2\,p^{2n+3}q^{2n-1}\binom{2n+1}{n+1}\binom{2n+1}{n+2} + \\ & \cdots\cdots\cdots\cdots\cdots\cdots \\ & + 2\,p^{4n+1}\cdot q^{1}\binom{2n+1}{2n}\binom{2n+1}{2n+1}\end{aligned}\right\}\quad (4)$$

Ahora bien, teniendo en cuenta la hipótesis $p \leqslant q$

$$\left.\begin{aligned} \alpha(0\,;\,2n+1\,,\,p) \leqslant 2\,p^{0}\cdot q^{4n+2}\binom{2n+1}{0}^{2} &+ \\ +\,2\,p^{2}\cdot q^{4n}\binom{2n+1}{1}^{2} &+ \\ \cdots\cdots\cdots\cdots\cdots & \\ +\,2\,p^{2n}\cdot q^{2n+2}\binom{2n+1}{n}^{2} & \end{aligned}\right| \tag{5}$$

Para que la desigualdad $\alpha(1\,;\,2n+1\,,\,p) > \alpha(0\,;\,2n+1\,,\,p)$ sea cierta, basta con que:

a/

$$\frac{2\,p^{2k+1}\cdot q^{4n-2k-1}}{2\,p^{2r}\cdot q^{4n-2k+2}}\cdot\frac{\binom{2n+1}{k}\binom{2n+1}{k+1}}{\binom{2n+1}{k}^{2}} \geqslant 1 \quad \text{para} \quad k=0,1,\ldots,n-1$$

es decir

$$\frac{p\cdot(2n+1-k)!\,k!}{q\,(2n-k)!\,(k+1)!}=\frac{p}{q}\cdot\frac{2n+1-k}{k+1}\geqslant\frac{p}{q}\cdot\frac{n+2}{n}\geqslant 1 \qquad \text{Si}$$

$$\boxed{\frac{p}{q}\geqslant\frac{n}{n+2}} \tag{6}$$

b/ la diferencia entre el sumando $(n+1)$—ésimo de (4) y el sumando $(n+1)$—ésimo de (5)

$$2\,(p^{2n}\cdot q^{2n+2}-p^{2n+1}\cdot q^{2n+1})\cdot\binom{2n+1}{n}^{2}=2\,p^{2n}\cdot q^{2n+1}(q-p)\cdot\binom{2n+1}{n}^{2}$$

se pueda hacer menor o igual al sumando $(n+2)$—ésimo de (4), esto es

$$\frac{2\,p^{2n}\cdot q^{2n+1}(q-p)}{2\,p^{2n+3}\cdot q^{2n-1}}\cdot\frac{\binom{2n+1}{n}^{2}}{\binom{2n+1}{n+1}\cdot\binom{2n+1}{n+2}}=\left(\frac{q}{p}\right)^{2}\frac{q-p}{p}\cdot\frac{n+2}{n}\leqslant 1$$

si se cumple la condición

$$\boxed{\left(\frac{p}{q}\right)^2 \geqslant \frac{n+2}{n} \cdot \frac{q-p}{p}} \qquad (7)$$

Multiplicando las desigualdades (6) y (7)

$$\boxed{\left(\frac{p}{q}\right)^3 \geqslant \frac{q-p}{p}} \qquad (8)$$

condición suficiente que ha de cumplir el parámetro p para que la distribución (3) tenga la moda en $r=1$.

La desigualdad (8) es evidente en el caso particular $p=q=1/2$, sin embargo, deja de ser cierta cuando $p \leqslant 1/3$ porque entonces

$$\frac{q-p}{p} = 1/p - 2 \geqslant 1 \quad \text{mientras que} \quad (p/q)^3 < 1.$$

De lo cual qe deduce la posibilidad de determinar un segmento $I = [1/2 - \varepsilon\,,\, 1/2 + \varepsilon]$ tal que si $p,q \in I$ la distribución (3) tiene la moda en $r=1$. Para ello basta llamar $p=1/2-\varepsilon$ y $q=1/2+\varepsilon$ y sustituir en (8)

$$(1/2 - \varepsilon)^4 \cdot \geqslant 2\,\varepsilon(1/2 + \varepsilon)^3$$

Resolviendo la ecuación

$$(1/2 - \varepsilon)^4 = 2\,\varepsilon\,(1/2 + \varepsilon)^3$$

$$-1/16 + 3/4\,\varepsilon + 5\,\varepsilon^3 + \varepsilon^4 = 0$$

se encuentra como única solución aceptable $\varepsilon \simeq 0{,}079$ (*)

Por consiguiente, el intervalo de valores de p y q es

$$I = [0{,}421\ ,\ 0.579]$$

y el teorema está demostrado.

(*) Se han determinado, con suficiente precisión, todas las raíces de la ecuación utilizando una rutina del Centro de Cálculo Electrónico de Madrid (C. S. I. C.). El resto del polinomio, cuando se sustituye en él el valor aproximado de la raíz, es menor que 25.10^{-7}.

CASO II: Número de tiradas par, $2n$

Seguiremos un procedimiento totalmente análogo al empleado en el Caso I

$$\left.\begin{aligned}
\alpha(1\,;\,2n\,.\,p) = {} & 2\,p^{1}\,q^{4n-1}\binom{2n}{0}\cdot\binom{2n}{1} + \\
& + 2\,p^{3}\,q^{4n-3}\binom{2n}{1}\cdot\binom{2n}{2} + \\
& \ldots\ldots\ldots \\
& + 2\,p^{2n-1}\cdot q^{2n+1}\binom{2n}{n-1}\cdot\binom{2n}{n} + \\
& + 2\,p^{2n+1}\,q^{2n-1}\binom{2n}{n}\cdot\binom{2n}{n+1} + \\
& + 2\,p^{2n+3}\cdot q^{2n-3}\binom{2n}{n+1}\cdot\binom{2n}{n+2} + \\
& \ldots\ldots\ldots\ldots\ldots\ldots \\
& + 2\,p^{4n-1}\cdot q^{1}\binom{2n}{2n-1}\cdot\binom{2n}{2n}
\end{aligned}\right\} \quad (4')$$

Por la hipótesis del teorema:

$$\left.\begin{aligned}
\alpha(0\,;\,2n\,,\,p) \leqslant {} & 2\,p^{0}\,q^{4n}\binom{2n}{0}^{2} + \\
& + 2\,p^{2}\,q^{4n-2}\binom{2n}{1}^{2} + \\
& \ldots\ldots\ldots \\
& + 2\,p^{2n-2}\cdot q^{2n+2}\binom{2n}{n-1}^{2} + \\
& + \quad p^{2n}\cdot q^{2n}\binom{2n}{n}^{2}
\end{aligned}\right\} \quad (5')$$

Se consigue hacer $\alpha(1\,;\,2n\,,\,p) > \alpha(0\,;\,2n\,,\,p)$ mediante las dos condiciones siguientes:

a'/ Si cada uno de los n primeros sumandos de (4') es mayor o igual a su correspondiente en (5')

$$\frac{2\,p^{2k+1}\cdot q^{4n-2k-1}}{2\,p^{2k}\cdot q^{4n-2k}}\cdot\frac{\binom{2n}{k}\cdot\binom{2n}{k+1}}{\binom{2n}{k}^{2}} \geqslant 1 \quad \text{para} \quad k = 0,1,\ \ ,n-1$$

es decir

$$\frac{p}{q}\cdot\frac{(2n-k)\,!\,k\,!}{(2n-k-1)\,!\,(k+1)\,!}=\frac{p}{q}\cdot\frac{2n-k}{k+1}\geqslant\frac{p}{q}\cdot\frac{n+1}{n}\geqslant 1$$

Si $$\boxed{\frac{p}{q}\geqslant\frac{n}{n+1}} \qquad (6')$$

b'/ Si el sumando $(n+1)$—ésimo de (4') es mayor o igual al sumando $(n+1)$—ésimo de (5')

$$\frac{2\,p^{2n+1}\cdot q^{2n-1}\begin{pmatrix}2n\\ n+1\end{pmatrix}\cdot\begin{pmatrix}2n\\ n\end{pmatrix}}{p^{2n}\cdot q^{2n}\cdot\begin{pmatrix}2n\\ n\end{pmatrix}^2}=\frac{2p}{q}\cdot\frac{n!\,n!}{(n-1)!\,(n+1)!}=$$

$$=\frac{p}{q}\cdot\frac{2n}{n+1}\geqslant 1$$

Si $$\boxed{\frac{p}{q}\geqslant\frac{n+1}{2n}} \qquad (7')$$

Multiplicando miembro a miembro las desigualdades (6') y (7') se obtiene

$$\boxed{\left(\frac{p}{q}\right)^2\geqslant 1/2} \qquad (8')$$

El intervalo $I'=[1/2-\varepsilon'\,,\,1/2+\varepsilon']$ se determina llamando $p=1/2-\varepsilon'\,,\,q=1/2+\varepsilon'$ y sustituyendo en la desigualdad anterior

$$2\,(1/2-\varepsilon')^2\geqslant(1/2+\varepsilon')^2$$

cuya única solución aceptable es $\varepsilon'\simeq 0{,}085$ y por consiguiente el intervalo de valores de p,q en el Caso II es

$$I'=[0.415\,,\,0.585]$$

Obsérvese que $I'\supset I$, sin embargo, el teorema es cierto en ambos casos para el intervalo I.

Momentos de la distribución (3)

La función generatriz de momentos

$$\varphi(t) = \mathrm{E}\,[e^{tr}] = \sum_{i=o}^{n} p^{2i} \cdot q^{2(n-i)} \binom{n}{i}^2 +$$

$$+\ 2\,e^{t} \sum_{i=o}^{n-1} p^{2i+1} \cdot q^{2(n-i)-1} \cdot \binom{n}{i} \binom{n}{i+1} +$$

$$+\ 2\,e^{2t} \sum_{i=o}^{n-2} p^{2i+2} \cdot q^{2(n-i)-2} \binom{n}{i} \binom{n}{i+2} +$$

....

$$+\ 2\,e^{nt} \cdot p^{n} \cdot q^{n} \binom{n}{o} \binom{n}{n}$$

permite determinar todos los momentos de la distribución:

$$\alpha_i = \frac{d^{(i)}\,\varphi(t)}{d\,t^{i}} \Bigg|_{t = o} \qquad (i = 1, 2, \ldots)$$

No obstante, puesto que las expresiones que se derivan de aquí son excesivamente complicadas, nos limitaremos a construir intervalos que encierren a los dos valores típicos: media y desviación standard.

Pero antes demostraremos la siguiente propiedad:

TEOREMA 3: *El primer momento absoluto, respecto de la media, de la distribución binomial* (1) *viene dado por*

$$\mathrm{E}\,[|\nu - np|] = 2\,\mu \binom{n}{\mu} p^{\mu} \cdot q^{n-\mu+1}$$

donde μ *es el primer entero que sigue a* n.p , *es decir* $\mathrm{n} \cdot \mathrm{p} < \mu \leqslant \mathrm{n} \cdot \mathrm{p} + 1$

En efecto:

Por definición $\quad \mathrm{E}\,[|\nu - np|] = 2 \sum_{k > np} (\mathrm{k} - np) \binom{n}{k} \cdot p^{k}\, q^{n-k}$

expresión que se puede descomponer

$$E\,[|\nu - np|] = 2 \sum_{k > np} (k - kp - np + kp) \binom{n}{k} \cdot p^k \cdot q^{n-k} =$$

$$= 2 \sum_{k > np} (kq - (n-k)p) \binom{n}{k} \cdot p^k \cdot q^{n-k} =$$

$$= 2\,q \left\{ \sum_{k > np} k \binom{n}{k} p^k \cdot q^{n-k} - \sum_{k > np} (n-k) \binom{n}{k} p^{k+1}\, q^{n-k-1} \right\} =$$

$$= 2\,q \left\{ \mu \binom{n}{\mu} \cdot p^{\mu} \cdot q^{n-\mu} + \sum_{k > np} (k+1) \cdot \binom{n}{k+1} p^{k+1} \cdot q^{n-(k+1)} - \right.$$

$$\left. - \sum_{k > np} (n-k) \binom{n}{k} \cdot p^{k+1} \cdot q^{n-(k+1)} \right\} =$$

$$= 2\,q \left\{ \mu \binom{n}{\mu} p^{\mu} \cdot q^{n-\mu} + \sum_{k > np} p^{k+1} \cdot q^{n-(k+1)} \left[\binom{n}{k} \cdot (n-k) - \right.\right.$$

$$\left.\left. - \binom{n}{k} (n-k) \right] \right\} =$$

$$= 2\,q\,\mu \binom{n}{\mu} p^{\mu} \cdot q^{n-\mu}$$

donde μ es entero y tal que $np < \mu \leqslant np + 1$

Esta propiedad junto con las de la función «esperanza matemática» permiten obtener expresiones cómodas de los intervalos a que hacíamos relación anteriormente.

Para la media de la distribución (3):

$$o \leqslant E\,[|\nu - \nu'|] = E\,[|\nu - np - \nu' + np|] \leqslant$$

$$\leqslant E\,[|\nu - np|] + E[|\nu' - np|] =$$

$$= 4\,q\,\mu \binom{n}{\mu} p^{\mu} \cdot q^{n-\mu}$$

es decir $\quad o \leqslant \alpha_1 \leqslant 2\,q \cdot \mu \dbinom{n}{\mu} p^{\mu} \cdot q^{n-\mu}$

siendo μ entero y $np < \mu \leqslant np + 1$

para el cuadrado de la desviación standard:

$$o \leqslant \sigma^2 = \alpha_2 - \alpha^2_1 = E\,[|\nu - \nu'|^2] - \alpha^2_1 =$$
$$= E\,[\nu^2] + E\,[\nu'^2] - 2\,E\,[\nu] \cdot E\,[\nu'] - \alpha^2_1 =$$
$$= 2\,(npq + n^2\,p^2) - 2\,n^2\,p^2 - \alpha^2_1 =$$
$$= 2\,npq - \alpha^2_1$$

o bien $2\,npq - 16\,\mu^2 \binom{n}{\mu}^2 p^{2\mu} \cdot q^{2\,(n-\mu)} \leqslant \sigma^2 \leqslant 2\,npq$

Este tipo de distribuciones se presenta siempre que se trata de comparar dos variables binomiales: Pensemos, por ejemplo, en dos piezas de artillería que tienen una dotación diaria de n proyectiles cada una. Supuesto que ambas tienen la misma probabilidad p de hacer blanco en un disparo, se podría estudiar la discrepancia entre el número de impactos diarios de la primera pieza y el número de impactos diarios de la segunda.

Otro ejemplo: si dos jugadores disponen de n monedas cada uno y convienen en sacar al azar ν y ν', respectivamente, ganando aquel que acierte el número $|\nu - \nu'|$, está claro que interesa jugar al «1».