EL ESQUEMA DE LAS PRUEBAS REPETIDAS SIN REPOSICION

por F. L. GASPAR

Ya se sabe en qué consiste este esquema. Sea:

A el número de bolillas —coloradas y blancas— contenidas en una urna.

N el número de extracciones sucesivas que se hacen sin reponer las bolillas.

p la probabilidad de extraer una bolilla blanca antes de empezar las extracciones.

q la probabilidad de extraer una bolilla colorada en las mismas condiciones.

La composición de la urna, antes de comenzar las extracciones, ha quedado determinada:

$A\,p=B$ es el número de bolillas blancas $(N \leq B)$

$A\,q=C$ es el de las coloradas.

Si se pide la probabilidad de que, en las N extracciones, salgan r bolillas blancas y $N-r$ coloradas, en un orden cualquiera, simbolizando esa probabilidad por Y_r, su valor es:

$$\frac{B(B-1)\dots(B-r+1)\,C(C-1)\dots(C-N+r+1)}{A\,(A-1)\,\dots\,(A-N+1)}\binom{N}{r}$$

Llamando S a la suma de las probabilidades correspondientes a los $N+1$ casos que se pueden presentar en las N extracciones, desarrollando el número combinatorio y factoreando convenientemente queda

$$S=\frac{B\,(B-1)\dots(B-N+1)}{A\,(A-1)\dots(A-N+1)}\left[1+\frac{N\,C}{1!\,(B-N+1)}+\right.$$

$$+\frac{N(N-1)\,C\,(C-1)}{2!\,(B-N+1)\,(B-N+2)}+\dots$$

$$+\frac{N(N-1)\dots(N-r+1)\,C\,(C-1)\dots(C-N+r+1)}{r!\,(B-N+1)\,(B-N+2)\dots(B-r)}+\dots+$$

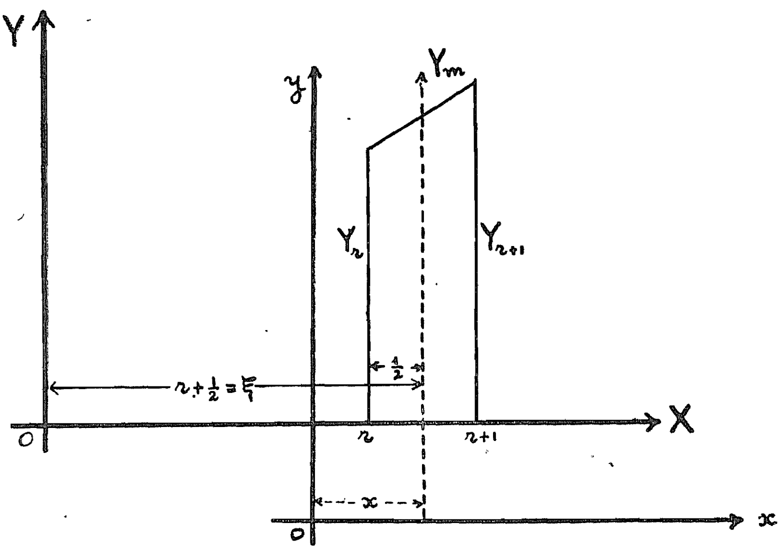
$$\left.+\frac{C\,(C-1)\dots(C-N+1)}{(B-N+1)\,(B-N+2)\dots B}\right]$$

Es decir, que la suma de dichas probabilidades, es un polinomio hipergeométrico.

Con el simbolismo usual, se escribe así:

$$S=\frac{B\,(B-1)\,\dots\,(B-N+1)}{A\,(A-1)\,\dots\,(A-N+1)}\;F\,(-N,-C,B-N+1,1)$$

Lo que queremos determinar es la función de repartición.



Siendo $\Delta Y_r = Y_{r+1} - Y_r$

si hacemos $r + \frac{1}{2} = \xi$

se llega a una expresión de esta forma

$$\frac{\Delta Y_r}{2\,Y_m} = \frac{A\left(N p + \frac{p-q}{2}\right) + N - \xi\,(A+2)}{A\,(N.p + \frac{1}{2}) + \frac{1}{2} - \xi\,(A\,(p-q) + 2N) + 2\,\xi^2}$$

Si suponemos que es $p = q = \frac{1}{2}$

y llevamos el origen al punto $\frac{N}{2}$, llamando x e y a las coordenadas del nuevo sistema resulta $x = \xi - \frac{N}{2}$

pero si tomamos x reducido, medido con la unidad de desvío, entonces es

$$\xi = \frac{N}{2} + x\,\sigma$$

$$\Delta x = \frac{1}{\sigma}$$

y por haber calculado x reducido corresponde

$$y = Y\,\sigma$$

Mediante sencillas transformaciones se llega, finalmente, a una expresión de esta forma

$$\frac{\Delta y}{y\,\Delta x} = \frac{-x}{\dfrac{N\,(A-N) + A + 1}{4\,(A+2)\,\sigma^2} + \dfrac{A+2}{1}\,x^2} \quad (1)$$

Como es, en este esquema,

$$\sigma = \sqrt{N \frac{A - N}{A - 1} \, p \, q}$$

en nuestro caso, al ser $p = q = ½$, resulta

$$\sigma^2 = ¼ \, \frac{N \, (A - N)}{A - 1}$$

Sustituyendo esta expresión de σ^2 en la (1) y suponiendo que N es muy grande, lo que importa que A también lo sea, se llega a tener, aproximadamente (v. G. Darmois - Statistique Mathématique - pág. 119)

$$\frac{y'}{y} = - x$$

es decir, una curva normal y simétrica, como en el esquema de las pruebas repetidas con reposición.

Se ve, pues, que si se quiere pasar de la ecuación en diferencias finitas (1) a la ecuación diferencial, por el clásico procedimiento del paso al límite, cuando $N \to \infty$, la fórmula se desnaturaliza, es decir, pierde todo el contenido del esquema que la individualiza.

Este resultado no debe sorprendernos; lo que ocurre es que, ese paso al límite, no tiene sentido en el esquema de las pruebas repetidas sin reposición, así como tampoco lo tiene la fórmula asintótica que, con él, se pretende obtener, pues se comprende, fácilmente, que a dicho esquema debe corresponder un tipo de curva a extremos nulos, del tipo de las de Jacobi.

Para que N pueda tender a infinito, A debe ser infinito y este sería un infinito actual. También lo serían B y C. Como se sabe de teoría de conjuntos, si a un infinito actual se quita un infinito potencial, que tal sería N, sigue siendo, aquel, un infinito actual; quiere ello decir que la composición de la urna no ha variado, después de cada extracción, a pesar de no haber reposición; entonces tampoco han variado las probabilidades iniciales $p = q = ½$ y se justifica el resultado a que se llegó.

En conclusión: el método clásico no es adecuado, en este caso, para el paso de la ecuación en diferencias a la ecuación diferencial, puesto que lo que interesa es hacer la variación contínua en un intervalo finito. El asunto parece invitar a que se lo estudie como un problema de teoría de conjuntos y, de cualquier manera, sería muy interesante lograr su solución.