AÑO XIII, SERIE II, N.º 50

1925, sep.

REVISTA

DE

CIENCIAS ECONÓMICAS

PUBLICACIÓN DE LA FACULTAD DE CIENCIAS ECONÓMICAS
CENTRO DE ESTUDIANTES Y COLEGIO
DE GRADUADOS

DIRECCIÓN Y ADMINISTRACIÓN
CALLE CHARCAS, 1835
BUENOS AIRES

La Dirección no se responsabiliza de las afirmaciones, los juicios y las doctrinas que aparezcan en esta Revista, en trabajos subscritos por sus redactores o colaboradores.

Imprenta de la Universidad

Apuntes sobre el cálculo de las probabilidades

(Continuación)

V

El teorema de Santiago Bernoulli

30. Hemos visto que siendo p la probabilidad de que se produzca un acontecimiento A, y $q = 1 - p$ la de que ocurra el acontecimiento contrario B, si se hacen n experiencias, la eventualidad *más probable* es la que corresponde a la presentación np veces de A y nq veces de B. Y hemos visto, también, que la probabilidad correspondiente, no obstante ser la mayor de todas, es sumamente pequeña. Por lo tanto será *muy raro* que el acontecimiento A se presente np veces exactamente.

Lo más probable es que se presente un número de veces comprendido entre ciertos límites $np - h$ y $np + h$. Esta cantidad variable, h, que representa la diferencia entre el número de veces, $m = np \pm h$, que se presente A en n experiencias, y el que corresponde a la probabilidad máxima, es lo que se llama *desvío*.

Así, si se tira un dado una vez, la probabilidad de que salga el *as* es $\frac{1}{6}$ Si se repite la experiencia 6000 veces, la probabilidad máxima corresponde a la salida del as *mil veces*. Es casi seguro que no saldrá las *mil veces exactamente*. Si se presenta 1028 veces, ó 975 veces, diremos que el *desvío* es $+ 28$ ó $- 25$, respectivamente.

31. Vamos a demostrar que « *aumentando el número de experiencias, aunque el desvío aumente, disminuye, en valor absoluto, la relación* $\left|\frac{h}{n}\right|$ *entre el desvío y el número de pruebas, de tal modo que haciendo « n » suficientemente grande, dicha relación puede hacerse tan pequeña como se desee* ».

O, en otros términos : « *que la relación entre el número de veces,* « *m* », *que se presentará en* « *n* » *experiencias un acontecimiento A de probabilidad* « *p* » *y el número total de experiencias, diferirá de* « *p* » *en una cantidad tan pequeña como se quiera, con tal que se tome* « *n* » *suficientemente grande* ».

Haciendo $m = np \pm h$ se ve enseguida que los dos enunciados son exactamente iguales. En efecto, siendo ε una cantidad tan pequeña como se quiera, el segundo enunciado implica que se verifique la desigualdad siguiente :

$$\left|\frac{m}{n} - p\right| = \left|\frac{np \pm h}{n} - p\right| < \varepsilon \quad \therefore$$

$$\left|\frac{h}{n}\right| < \varepsilon.$$

Es decir, el primer enunciado.

32. Para demostrarlo empezaremos por calcular la probabilidad de que se produzca un desvío dado : digamos — h. Esto equivale *a hacer* $m = np - h$, y, por lo tanto, la probabilidad buscada nos será dada por el término del desarrollo de $(p+q)^n$ en que sean

$$m = np - h. \qquad n - m = nq + h.$$

es decir, por el término

$$\frac{n!}{(np-h)!\,(nq+h)!}\, p^{np-h} q^{nq+h}.$$

Transformándolo, mediante la fórmula de Stirling, tenemos

$$\frac{e^{-n} n^n \sqrt{2\pi n}\, p^{np-h} q^{nq+h}}{e^{-(np-h)} (np-h)^{np-h} \sqrt{2\pi(np-h)}\, e^{-(nq+h)} (nq+h)^{nq+h} \sqrt{2\pi(nq+h)}} =$$

$$= \frac{e^{-n} n^n \sqrt{2\pi n}\, p^{np-h} q^{nq+h}}{e^{-np+h-nq-h}\left[np\left(1-\frac{h}{np}\right)\right]^{np-h}\left[nq\left(1+\frac{h}{nq}\right)\right]^{nq+h}\sqrt{2\pi np\left(1-\frac{h}{np}\right) 2\pi nq\left(1+\frac{h}{nq}\right)}} =$$

$$= \frac{e^{-n} n^n \sqrt{2\pi n}\, p^{np-h} q^{nq+h}}{e^{-n} n^{np-h} p^{np-h}\left(1-\frac{h}{np}\right)^{np-h} n^{nq+h} q^{nq+h}\left(1+\frac{h}{nq}\right)^{nq+h}\sqrt{2\pi n}\sqrt{2\pi npq\left(1-\frac{h}{np}\right)\left(1+\frac{h}{nq}\right)}}.$$

Pero

$$n^{np-h} \,.\, n^{nq+h} = n^{n(p+q)} = n^n.$$

Luego, simplificando, queda

$$\frac{1}{\sqrt{2\pi npq}\left(1-\frac{h}{np}\right)^{np-h}\left(1+\frac{h}{nq}\right)^{nq+h}\sqrt{1-\frac{h}{np}}\cdot\sqrt{1+\frac{h}{nq}}}=$$

$$=\frac{1}{\sqrt{2\pi npq}}\cdot\frac{1}{\left(1-\frac{h}{np}\right)^{np-h+\frac{1}{2}}\left(1+\frac{h}{nq}\right)^{nq+h+\frac{1}{2}}}. \quad (1)$$

Representemos ahora por α el denominador de la segunda fracción. Tendremos

$$\alpha=\left(1-\frac{h}{np}\right)^{np-h+\frac{1}{2}}\left(1+\frac{h}{nq}\right)^{nq+h+\frac{1}{2}}. \quad (2)$$

Tomando logaritmos neperianos

$$l.\alpha=\left(np-h+\frac{1}{2}\right)l.\left(1-\frac{h}{np}\right)+\left(nq+h+\frac{1}{2}\right)l.\left(1+\frac{h}{nq}\right). \quad (3)$$

Pero $l.\ (1\pm x)$, cuando es $x<1$ se puede desarrollar en serie, y se tiene

$$l.(1+x)=x-\frac{x^2}{2}+\frac{x^3}{3}-\frac{x^4}{4}+\ldots$$

$$l.(1-x)=-x-\frac{x^2}{2}-\frac{x^3}{3}-\frac{x^4}{4}-\ldots$$

Por lo tanto

$$l.\left(1-\frac{h}{np}\right)=-\frac{h}{np}-\frac{h^2}{2n^2p^2}-\ldots \quad (4)$$

$$l.\left(1+\frac{h}{nq}\right)=\frac{h}{nq}-\frac{h^2}{2n^2q^2}+\ldots \quad (5)$$

Tanto en (4) como en (5) prescindiremos de los términos que siguen al segundo, debido a su pequeñez. Llevando, pues, esos valores a la (3) tendremos :

$$l.\alpha=\left(np-h+\frac{1}{2}\right)\left(-\frac{h}{np}-\frac{h^2}{2n^2p^2}\right)+\left(nq+h+\frac{1}{2}\right)\left(\frac{h}{nq}-\frac{h^2}{2n^2q^2}\right) \quad \therefore$$

$$l.\alpha=-h+\frac{h^2}{np}-\frac{h}{2np}-\frac{h^2}{2np}+\frac{h^3}{2n^2p^2}-\frac{h^2}{4n^2p^2}+$$

$$+h+\frac{h^2}{nq}+\frac{h}{2nq}-\frac{h^2}{2nq}-\frac{h^3}{2n^2q^2}-\frac{h^2}{4n^2q^2}.$$

Y simplificando, y prescindiendo de los términos que tienen n^2 en el denominador,

$$l\,.\,\alpha = \frac{h^2}{2np} + \frac{h^2}{2nq} + \frac{h}{2nq} - \frac{h}{2np} = \frac{h^2}{2n}\left(\frac{1}{p} + \frac{1}{q}\right) + \frac{h}{2n}\left(\frac{1}{q} - \frac{1}{p}\right).$$

Pero como $\frac{h}{2\,n}$ es muy pequeño, podemos desdeñar el término que lo contiene como factor, y queda, entonces,

$$l\,.\,\alpha = \frac{h^2}{2\,n}\left(\frac{1}{p} + \frac{1}{q}\right) = \frac{h^2}{2\,npq} \quad \therefore$$

$$\alpha = e^{\frac{h^2}{2\,npq}}$$

Pero en la (1) no tenemos α, sino $\frac{1}{\alpha}$ luego

$$\frac{1}{\alpha} = e^{-\frac{h^2}{2\,npq}}$$

Y la probabilidad que buscamos es, por consiguiente

$$\frac{1}{\sqrt{2\,\pi\,npq}}\,e^{-\frac{h^2}{2\,npq}}.$$

33. La probabilidad que acabamos de calcular es para un desvío negativo. Pero es evidente que si el desvío fuera positivo, llegaríamos a la misma expresión, desde que h está elevado a una potencia par.

34. Hagamos, ahora, algunas aplicaciones de la fórmula hallada. Sean

$$n = 1000 \quad \text{y} \quad p = q = \frac{1}{2}$$

Hallemos la probabilidad de que se produzca un desvío de $+\,40$. Tendremos

$$\frac{1}{\sqrt{2 \times 3\,.\,1416 \times 1000 \times \frac{1}{2} \times \frac{1}{2}}}\,e^{-\frac{1600}{2 \times 1000 \times {}^1/_2 \times {}^1/_2}} =$$

$$= \frac{1}{\sqrt{1570.8}}\,e^{-3.2} = 0.0010285$$

La probabilidad de que el desvío fuera $-\,40$ sería, como ya hicimos notar, la misma.

La probabilidad de que el desvío sea 60, será, supuestos para n, p y q, los mismos valores que antes,

$$\frac{1}{\sqrt{1570.8}} e^{-7.2} = 0.00001883 = 0.0^4 1883,$$

dándole, previamente, un signo a 60. Pero si sólo queremos que el desvío sea 60 *en valor absoluto, sin atender al signo,* su valor será, evidentemente doble, es decir, $0.00003776 = 0.0^4 3776$.

Hallemos, por fin, la probabilidad de que con los datos de los ejemplos anteriores, sea $+ 100$ el desvío.

Tendremos :

$$\frac{1}{\sqrt{1570.8}} e^{-20} = 0.0^9 52006.$$

35. Las cifras a que hemos llegado decrecen tan rápidamente a medida que crece el desvío, que surge de ellas la evidencia de que, *prácticamente,* es imposible que se produzcan desvíos superiores, en valor absoluto a un valor dado. O, en otros términos, que hay un valor h del desvío tal que será

$$\sum_{m = np - h}^{m = np + h} C_n^m p^m q^{n - m} = 1 - \varepsilon$$

siendo ε una cantidad tan pequeña como se quiera.

La probabilidad de que el desvío que se produzca no exceda de h en valor absoluto, tiende, pues, rápidamente hacia *uno* y hay una certidumbre *moral* de que ciertos desvíos no habrán de producirse nunca.

Tal es, en esencia, el teorema de Santiago Bernoulli. De precisar su alcance fijando el valor límite de h en función de las constantes n, p y q, trataremos en los siguientes párrafos.

36. Hemos llegado, para la probabilidad de que se produzca *exactamente* un desvío h, a la expresión

$$\frac{1}{\sqrt{2 \pi npq}} e^{-\frac{h^2}{2npq}}$$

que según vimos, nos da sólo un valor aproximado.

Admitamos ahora que el desvío h, en lugar de variar solamente por valores enteros, puede variar de un modo continuo, es decir, que se puede reemplazar la variable discontinua h por la continua z en la expresión anterior.

La probabilidad que corresponderá, entonces, a un desvío comprendido entre z y dz, es :

$$\frac{1}{\sqrt{2\pi npq}} e^{-\frac{z^2}{2npq}} dz$$

Admitamos, además, que el número de pruebas se hace infinitamente grande y que el desvío puede entonces variar entre $-\infty$ y $+\infty$. La suma de todas estas probabilidades, que nos habrá de dar la certeza, se obtiene integrando la expresión anterior entre los límites $-\infty$ y $+\infty$. Tendremos :

$$\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi npq}} e^{-\frac{z^2}{2npq}} dz = \frac{1}{\sqrt{2\pi npq}} \int_{-\infty}^{+\infty} e^{-\frac{z^2}{2npq}} dz = 1$$

Comprobémoslo.

Siendo la función a integrar tal que para valores positivos de z hemos de obtener los mismos resultados que para valores negativos, desde que dicha variable está elevada al cuadrado, integrarla entre $-\infty$ y $+\infty$ equivale a integrarla entre 0 y $+\infty$ y tomar luego el doble. Es decir, que tendremos

$$\frac{1}{\sqrt{2\pi npq}} \int_{-\infty}^{+\infty} e^{-\frac{z^2}{2npq}} dz = \frac{2}{\sqrt{2\pi npq}} \int_{0}^{+\infty} e^{-\frac{z^2}{2npq}} dz.$$

Hagamos

$$\frac{z}{\sqrt{2npq}} = t \quad \therefore \quad \frac{z^2}{2npq} = t^2 \quad \therefore \quad z = t\sqrt{2npq} \quad \text{y} \quad dz = \sqrt{2npq}\, dt$$

Entonces : para

$$z = 0, \quad t = 0$$

y para

$$z = \infty, \quad t = \infty$$

Nos resulta, pues, cambiando de variable :

$$\frac{2}{\sqrt{2\pi npq}} \int_0^{+\infty} e^{-\frac{z^2}{2npq}} dz = \frac{2}{\sqrt{2\pi npq}} \int_0^{+\infty} e^{-t^2} \sqrt{2npq}\, dt =$$

$$= \frac{2\sqrt{2npq}}{\sqrt{2\pi npq}} \int_0^{+\infty} e^{-t^2} dt = \frac{2}{\sqrt{\pi}} \int_0^{+\infty} e^{-t^2} dt.$$

Ahora bien : en todos los tratados de cálculo se demuestra (y nosotros vamos a dar en seguida una demostración), que

$$\int_0^{+\infty} e^{-t^2}\,dt = \frac{\sqrt{\pi}}{2}.$$

Luego :

$$\frac{2}{\sqrt{\pi}}\int_0^{+\infty} e^{-t^2}\,dt = \frac{2}{\sqrt{\pi}}\cdot\frac{\sqrt{\pi}}{2} = 1$$

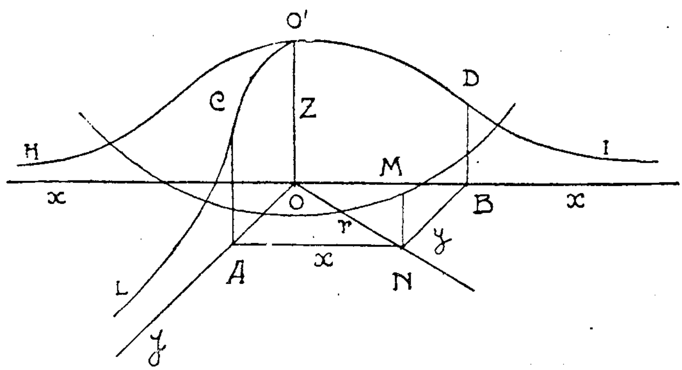
como queríamos demostrar.

Es decir, que admitiendo que el desvío pueda llegar a tener un valor infinito y por lo tanto que es también infinito el número de pruebas, nuestra fórmula nos ha dado un resultado exacto. Esto prueba la naturaleza asintótica de la misma. Luego cuanto mayor sea el número de experiencias, mayor será su grado de exactitud.

37. Demostremos ahora que realmente es :

$$\int_0^{+\infty} e^{-t^2}\,dt = \frac{\sqrt{\pi}}{2}.$$

Sean x, y y z tres ejes coordenados rectangulares que determinan



tres planos ortogonales xy, xz e yz, que pasan por el punto O.

En el plano xz tracemos la curva

$$e^{-x^2}$$

y en el plano yz *la curva*

$$e^{-y^2}$$

Integrando ambas funciones entre los límites o y $+\infty$ y represendo por S cada una de las integrales, que tendrán igual valor toda vez que sólo se diferencian en el símbolo elegido para designar la variable; tenemos :

$$S = \int_0^{+\infty} e^{-x^2}\, dx \qquad S = \int_0^{+\infty} e^{-y^2}\, dy.$$

Multiplicando ambas ecuaciones, tenemos :

$$S^2 = \int_0^{+\infty}\int_0^{+\infty} e^{-x^2} e^{-y^2}\, dxdy = \int_0^{+\infty}\int_0^{+\infty} e^{-(x^2+y^2)}\, dxdy.$$

Pero $(x^2 + y^2) = r^2$, desde que x e y son los catetos del triángulo rectángulo OAN = OBN, cuya hipotenusa es ON, luego :

$$e^{-(x^2+y^2)} = e^{-r^2}.$$

Además, $dxdy$ es el área de un rectángulo (en el plano xy) de lados dx y dy, por lo tanto :

$$e^{-(x^2+y^2)}\, dxdy = e^{-r^2}\, dxdy$$

es el volumen de un paralelepípedo rectángulo que tiene por base el rectángulo infinitesimal $dxdy$ y por altura e^{-r^2}.

La integral que buscamos, S^2, es, pues, la suma de los volúmenes de todos los paralelepípedos análogos, que se forman en la región angular xOy del plano xy, o sea el volumen engendrado por la superficie plana $yOO'CL$ al girar sobre OO' hasta tomar la posición $xOO'DI$.

Pero dicho volumen es evidentemente la cuarta parte del que engendra $yOO'CL$ al girar sobre OO' hasta ocupar de nuevo su posición primitiva.

Calcular este nuevo volumen es, pues, calcular el anterior; representando por V el volumen total, tenemos :

$$V = 4\, S^2$$

y como, evidentemente, el volumen avaluado será el mismo, cualquiera que sea el procedimiento de descomposición que adoptemos, supondremos en el plano xy una serie de círculos concéntricos, cuyo radio indicaremos en función de la variable r.

Ahora bien, dos círculos cuyos radios son, rescpectivamente r y $r + dr$, determinan una corona circular que tiene por área :

$$\pi (r + dr)^2 - \pi r^2 = \pi (r^2 + 2rdr + [dr]^2 - r^2) = \pi (2rdr + [dr]^2)$$

que se reduce a $2\pi rdr$, prescindiendo del infinitesimal de segundo orden $(dr)^2$.

A cada una de esas coronas corresponde una capa cilíndrica cuya altura es e^{-r^2} y su volumen:

$$e^{-r^2} 2\pi r dr.$$

Hallar la suma de los volúmenes de todos los sólidos análogos es integrar la diferencial anterior entre los límites cero e infinito. Es decir, que tenemos :

$$4\,S^2 = V = \int_0^\infty e^{-r^2} 2\pi r dr = -\pi \int_0^\infty - e^{-r^2} 2rdr.$$

Pero la diferencial de e^{-r^2} es de la forma $e^u\, du$, en que $u = -r^2$, $du = -2rdr$. Luego $-e^{-r^2} 2rdr$ es la diferencial de e^{-r^2}. Por tanto :

$$4S^2 = -\pi \int_0^\infty - e^{-r^2} 2rdr = -\pi \left[e^{-r^2}\right]_0^\infty =$$

$$= \pi\left[e^{-\infty} - e^{-0}\right] = -\pi(-1) = \pi$$

Luego :

$$S^2 = \frac{\pi}{4} \quad \therefore \quad S = \frac{\sqrt{\pi}}{2}$$

38. Calculemos ahora el *valor probable del desvío* o sea la esperanza matemática del jugador a quien se le ofrece una suma igual al desvío que se produzca.

Eso equivale a multiplicar cada término del desarrollo de

$$(p+q)^n$$

por h, *o*, si hacemos uso de la variable continua, a calcular la integral

$$\frac{2}{\sqrt{2\pi npq}} \int_0^\infty e^{-\frac{z^2}{2npq}} z dz.$$

Haciendo como antes :

$$\frac{z}{\sqrt{2npq}} = t \therefore z = t\sqrt{2npq} \therefore dz = \sqrt{2npq}\, dt$$

y substituyendo, resulta, teniendo en cuenta que, según hemos visto, los límites no cambian.

$$\frac{2}{\sqrt{2\pi npq}}\int_0^\infty e^{-\frac{z^2}{2npq}} zdz = \frac{2}{\sqrt{2\pi npq}}\int_0^\infty e^{-t^2}\, t\sqrt{2npq}\sqrt{2npq}\, dt =$$

$$= -\frac{\sqrt{2npq}\sqrt{2npq}}{\sqrt{2\pi npq}}\int_0^\infty -e^{-t^2} 2tdt =$$

$$= -\frac{\sqrt{2npq}}{\sqrt{\pi}}\left[e^{\cdot 2}\right]_0^\infty = -\frac{\sqrt{2npq}}{\sqrt{\pi}}(0-1) = \sqrt{\frac{2}{\pi}}\sqrt{npq} = 0{,}79789\sqrt{npq}.$$

Aproximadamente

$$0{,}8\sqrt{npq}$$

39. Calculemos ahora el valor probable del *cuadrado del desvío,* o sea la esperanza matemática del jugador a quien se le ofrece una suma igual al cuadrado del desvío que se produzca.

Si hacemos uso de la variable continua, tendremos que calcular la integral

$$\frac{2}{\sqrt{2\pi npq}}\int_0^\infty e^{-\frac{z^2}{2npq}} z^2\, dz.$$

Haciendo, como antes :

$$\frac{z}{\sqrt{2npq}} = t \quad \therefore \quad z = t\sqrt{2npq} \quad \therefore \quad dz = \sqrt{2npq}\, dt \quad \therefore \quad z^2 = 2npqt^2$$

y teniendo presente que los límites no varían, nos resultará, reemplazando valores :

$$\frac{2}{\sqrt{2\pi npq}}\int_0^\infty e^{-\frac{z^2}{2npq}} z^2\, dz = \frac{2}{\sqrt{2\pi npq}}\int_0^\infty e^{-t^2} 2\, npqt^2\sqrt{2npq}\, dt =$$

$$= -\frac{2\, npq}{\sqrt{\pi}}\int_0^\infty -e^{-t^2} 2\, t^2\, dt.$$

Consideremos la integral con prescindencia de la constante que la afecta

$$\int_0^\infty -e^{-t^2} 2\, t^2\, dt = \int_0^\infty -e^{-t^2} 2\, ttdt.$$

Hagamos :

$$t = u\,; \qquad -e^{-t^2} 2\, tdt = dv$$

$$\therefore \quad dt = du\,; \qquad e^{-t^2} = v$$

Nuestra integral se transforma en

$$\int_{o}^{\infty} udv,$$

Integrando por partes

$$\int_{o}^{\infty} udv = \Big[uv\Big]_{o}^{\infty} - \int_{o}^{\infty} vdu$$

y substituyendo valores :

$$\int_{o}^{\infty} - e^{-t^2} 2\, ttdt = \Big[te^{-t^2}\Big]_{o}^{\infty} - \int_{o}^{\infty} e^{-t^2} dt.$$

Pero el primer término del segundo miembro es nulo para ambos límites, pues en un caso se anula t y en otro e^{-t^2}

En cuanto al segundo término, sabemos que vale

$$\frac{\sqrt{\pi}}{2}.$$

Tenemos, pues :

$$\int_{o}^{\infty} - e^{-t^2} 2\, ttdt = -\frac{\sqrt{\pi}}{2}.$$

Luego :

$$-\frac{2\, npq}{\sqrt{\pi}} \int_{o}^{\infty} - e^{-t^2} 2\, t^2 dt = \left(-\frac{2\, npq}{\sqrt{\pi}}\right)\left(-\frac{\sqrt{\pi}}{2}\right) = npq.$$

Tal es el valor probable del cuadrado del desvío. El valor que corresponde al desvío, deducido de éste, y que es costumbre representar por μ es:

$$\mu = \sqrt{npq}.$$

Calculando directamente habíamos hallado

$$\sqrt{\frac{2}{\pi}}\sqrt{npq}$$

o, aproximadamente, $0{,}8\sqrt{npq}$, es decir, los cuatro quintos del valor que ahora resulta.

El que acabamos de calcular,

$$\mu = \sqrt{npq},$$

se llama usualmente *desvío medio cuadrático.*

El que hallamos por medio del cálculo directo,

$$0.8\sqrt{npq},$$

se llama *desvío medio.*

Conviene recordar estas denominaciones y sobre todo no confundir a uno ni a otro con el llamado *desvío probable,* que es

$$0,476936\sqrt{2\,npq}$$

según más adelante veremos.

40. Si calculamos los valores del *desvío medio* y del *desvío medio cuadrático* mediante la fórmula del binomio, es decir, sin admitir que los desvíos puedan variar de un modo continuo, llegamos a los mismos resultados.

Calculemos previamente el *desvío medio cuadrático.*

Como antes, habremos de calcular la esperanza matemática del jugador a quien se le ofrece una suma igual al cuadrado del desvío que se produzca.

Siendo *h el desvío,* tendremos :

$$m = np \pm h \quad \therefore \quad h^2 = (np - m)^2 = (m - np)^2$$

valor del cuadrado del desvío en función de la variable m.

Por esa cantidad $(np - m)^2$ habremos de multiplicar cada término de $(p + q)^n$.

La esperanza matemática que buscamos es igual a :

$$\sum_{m=0}^{m=n} (np - m)^2\, C_n^m p^m q^{n-m} = \sum_{m=0}^{m=n} (n^2p^2 - 2npm + m^2)\, C_n^m p^m q^{n-m} =$$

$$= \sum_{m=0}^{m=n} n^2p^2\, C_n^m p^m q^{n-m} - \sum_{m=0}^{m=n} 2npm\, C_n^m p^m q^{n-m} +$$

$$+ \sum_{m=0}^{m=n} m^2\, C_n^m p^m q^{n-m}.$$

Calculemos por separado cada uno de los términos, que llamaremos, respectivamente, S_1, S_2, S_3.

Tendremos :

$$S_1 = \sum_{m=0}^{m=n} n^2p^2\, C_n^m p^m q^{n-m} = n^2p^2 \sum_{m=0}^{m=n} C_n^m p^m q^{n-m}$$

pero

$$\sum_{m=0}^{m=n} C_n^m p^m q^{n-m} = (p+q) \qquad (\alpha).$$

Luego $S_1 = n^2p^2$ desde que $(p+q)^n = 1$.
Ahora,

$$S_2 = \sum_{m=0}^{m=n} 2npm\, C_n^m p^m q^{n-m} = 2np \sum_{m=0}^{m=n} m\, C_n^m p^m q^{n-m}.$$

Pero teníamos (α):

$$\sum_{m=0}^{m=n} C_n^m p^m q^{n-m} = (p+q)^n.$$

Derivando esta expresión con respecto a p, tenemos :

$$\sum_{m=0}^{m=n} m\, C_n^m p^{m-1} q^{n-m} = n(p+q)^{n-1}$$

y multiplicando por p ambos miembros :

$$\sum_{m=0}^{m=n} m\, C_n^m p^m q^{n-n} = np\,(p+q)^{n-1} \qquad (\beta),$$

O sea, np puesto que $(p+q)^{n\text{-}1} = 1$. Luego :

$$S_2 = 2npnp = 2n^2p^2$$

Calculemos, ahora, S_3 :

$$S_3 = \sum_{m=0}^{m=n} m^2\, C_n^m p^m q^{n-m}.$$

Tomemos la derivada de (β) con respecto a p :

$$\sum_{m=0}^{m=n} m^2 C_n^m p^{m-1} q^{n-m} = np(n-1)(p+q)^{n-2} + n(p+q)^{n-1}$$

y multipliquemos ambos miembros por p

$$\sum_{m=0}^{m=n} m^2 C_n^m p^m q^{n-m} = n(n-1)p^2(p+q)^{n-2} + np(p+q)^{n-1} =$$

$$= n(n-1)p^2 + np$$

puesto que $(p+q)^{n-2}$ y $(p+q)^{n-1}$ valen 1, y como el último valor hallado es precisamente el de S_3, tenemos en definitiva, que el valor probable del cuadrado del desvío es :

$$S_1 - S_2 + S_3 = n^2p^2 - 2n^2p^2 + n(n-1)p^2 + np =$$
$$= np - np^2 = np(1-p) = npq.$$

O sea, el mismo resultado de antes.

41. Calculemos ahora, por medio de la fórmula del binomio, el valor del *desvío medio,* que es, según hemos definido, la esperanza matemática del jugador a quien se ofrece una suma igual al *desvío* que se produzca.

Como aquí no tenemos el desvío elevado al cuadrado, hemos de considerar por separado el desvío positivo y el negativo.

Si es el desvío negativo, tenemos :

$$m = np - h$$

$$n - m = nq + h$$

multiplicando por q la primera ecuación y por p la segunda, nos queda :

$$mq = npq - hq$$

$$(n-m)\,p = npq + hp$$

y, restando la primera de la segunda :

$$(n-m)p - mq = hp + hq = h \therefore$$

$$h = \frac{n-m}{q}pq - \frac{m}{p}pq \quad \therefore$$

$$h = pq\left[\frac{n-m}{q} - \frac{m}{p}\right] \qquad (1)$$

Para tener ahora el valor del desvío positivo bastará, evidentemente, cambiar de signo la expresión anterior, y tendremos :

$$h = pq\left[\frac{m}{p} - \frac{n-m}{q}\right]. \qquad (2)$$

Multiplicar, pues, por h los términos del binomio $(p+q)^n$ es multiplicarlos respectivamente por (1) o por (2), según sea negativo o positivo el desvío.

Pero, ¿a qué equivale eso?

En el caso del desvío negativo, hemos de multiplicar cada término, primero por $\frac{n-m}{q}$, luego por $\frac{m}{p}$, hacer después la diferencia y multiplicar el resultado por pq.

Pero multiplicar cada término por $\frac{m}{p}$ es tomar su derivada con respecto a p; y multiplicarlo por $\frac{n-m}{q}$ tomar su derivada con respecto a q.

Sacamos en consecuencia que para los términos de desvío negativo se toman las derivadas de cada término con respecto a p y q, se resta la primera de la segunda y se multiplica la diferencia por pq.

Para los términos de desvío positivo, en lugar de restar la derivada con respecto a p de la con respecto a q, se resta ésta de aquélla.

¿Y el término máximo? Como su desvío es *nulo*, sus derivadas con respecto a p y a q han de ser iguales.

Podemos, pues, al derivar, incluir el término máximo en los dos grupos de términos que formaremos (de desvío positivo y de desvío negativo), sin ningún inconveniente.

La suma de los términos de desvío negativo (incluyendo el máximo) es :

$$\sum_{m=0}^{m=np} C_n^m p^m q^{n-m} = q^n + npq^{n-1} + \frac{n(n-1)}{2} p^2 q^{n-2} +$$

$$+ \ldots + \frac{n!}{(np-1)!\,(nq+1)!} p^{np-1} q^{nq+1} + \frac{n!}{(np)!\,(nq)!} p^{np} q^{nq}.$$

Derivando con respecto a q :

$$nq^{n-1} + n(n-1)pq^{n-2} + \frac{n(n-1)(n-2)}{2} p^2 q^{n-3} + \ldots +$$

$$+ \frac{n!}{(np-1)!\,(nq+1)!} (nq+1) p^{np-1} q^{nq} + \frac{n!}{(np)!\,(nq)!} nqp^{np} q^{nq-1}.$$

Simplificando y multiplicando y dividiendo por np el penúltimo término, queda :

$$nq^{n-1} + n(n-1)pq^{n-2} + \frac{n(n-1)(n-2)}{2}p^2q^{n-3} + \dots +$$

$$+ \frac{n!}{(np)!(nq)!}np^{np}q^{nq} + \frac{n!}{(np)!(nq)!}np^{np}q^{nq}.$$

Derivemos ahora con respecto a p y tendremos :

$$0 + nq^{n-1} + n(n-1)pq^{n-2} + \frac{n(n-1)(n-2)}{2}p^2q^{n-3} +$$

$$+ \dots + \frac{n!}{(np)!(nq)!}np^{np}q^{nq}.$$

Fácil es ver, ahora, que al restar la segunda expresión de la primera, se anulan todos los términos, menos el que corresponde a la derivada con respecto a q del término máximo, es decir :

$$\frac{n!}{(np)!(nq)!}np^{np}q^{nq}.$$

Es evidente que si hacemos ahora los cálculos análogos que corresponden a los desvíos positivos, nos quedará sólo la derivada del término máximo con respecto a p. Como hicimos notar y como resulta de los cálculos ya hechos, las dos derivadas son iguales. Luego, el valor probable del desvío, o sea el *desvío medio*, será igual al duplo de ese desvío multiplicado por pq.

$$2pq\frac{n!}{(np)!(nq)!}np^{np}q^{nq} = 2npq\left[\frac{n!}{(np)!(nq)!}p^{np}q^{nq}\right] =$$

$$= 2npq\frac{1}{\sqrt{2\pi npq}} = \frac{\sqrt{2npq}}{\sqrt{\pi}}$$

toda vez que la cantidad que para mayor visibilidad hemos encerrado entre corchetes, no es sino el término máximo.

42. Los desvíos hasta aquí calculados son los que se llaman *absolutos*. Si los dividimos por el número de pruebas realizadas, obtendremos una relación entre el desvío absoluto y el número de experiencias, que se llama *desvío relativo*.

El *desvío medio absoluto* es $0{,}8\sqrt{npq}$.

El *desvío medio relativo* es, por lo tanto,

$$\frac{0{,}8\sqrt{npq}}{n} = 0{,}8\frac{\sqrt{pq}}{\sqrt{n}}.$$

El *desvío medio cuadrático absoluto* es $\sqrt{npq}$.

El *relativo* es

$$\frac{\sqrt{npq}}{n}=\frac{\sqrt{pq}}{\sqrt{n}}.$$

Vemos, pues, que al aumentar el número de experiencias, el desvío absoluto crece, al par que el relativo decrece. Y la razón es obvia. El desvío absoluto *crece*, pero lo hace proporcionalmente a la *raíz cuadrada* del número de experiencias, n, luego crece *mucho más lentamente* que dicho número n, y por lo tanto, su relación con él ha de disminuir a medida que crezca n.

Resultado que demuestra el teorema enunciado en el párrafo 31, bajo dos formas distintas.

De acuerdo con ello, sólo en el caso de ser n infinitamente grande, tendríamos exactamente :

$$\frac{m}{n}=p.$$

Esta observación precisa al alcance de la proposición demostrada. Nos permite determinar, con respecto a los *desvíos* que puedan producirse, un *límite superior* y otro *inferior*, cuya probabilidad de ser sobrepasados sea tan mínima que, *prácticamente*, pueda considerarse como *nula*, y el hecho, por lo tanto, como imposible.

43. Si en la expresión que nos da la probabilidad de que se produzca un desvío igual a $+h$ (o a $-h$),

$$\frac{1}{\sqrt{2\pi npq}}e^{-\frac{h^2}{2npq}}$$

hacemos

$$\varepsilon=\frac{1}{\sqrt{2npq}} \quad \text{y} \quad x=h\varepsilon$$

serán

$$\varepsilon^2=\frac{1}{2npq} \quad \text{y} \quad \frac{h^2}{2npq}=h^2\varepsilon^2=x^2$$

siendo ε una cantidad menor que uno, y tanto menor cuanto mayor sea n.

Representemos gráficamente la función

$$e^{-\frac{h^2}{2npq}}=e^{-x^2}$$

y veamos cómo varía para valores enteros de h. Ya sabemos que es simétrica con respecto al eje de las ordenadas, puesto que toma iguales valores para $h = \pm 1$; $h = \pm 2$; ...

Como dicha función depende de los valores de n, p y q, hagamos $n = 200$; $p = q = \frac{1}{2}$. Será, entonces,

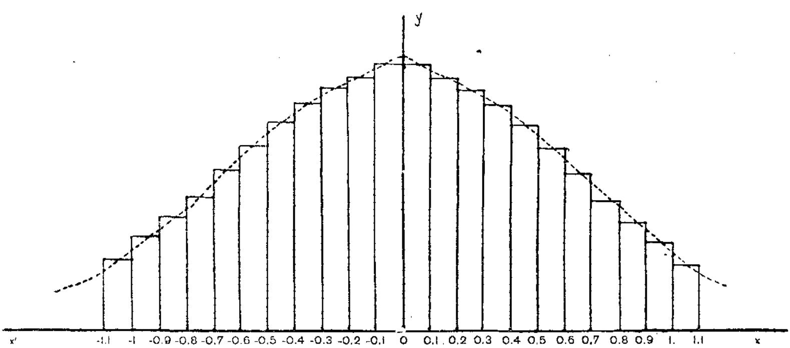
$$\varepsilon = \frac{1}{\sqrt{2 \times 200 \times \frac{1}{2} \times \frac{1}{2}}} = 0.1.$$

Y para $h = 1$; 2; 3; ..., tendremos $x = 0{,}1$; 0,2; 0,3...

Por lo tanto, para desvíos ± 1; ± 2; ± 3; ... tendremos, sobre el eje de las abcisas, distancias iguales a $\pm 0{,}1$; $\pm 0{,}2$; $\pm 0{,}3$... a las que corresponderán las ordenadas que indica la siguiente tabla :

Valores de la función: $y = e^{-x^2}$

x	y	x	y	x	y
0	1	0,7	0,61263	1,5	0,10540
0,1	0,99005	0,8	0,52729	2	0,018316
0,2	0,96076	0,9	0,44486	2,5	0,001930
0,3	0,91393	1	0,36788	3,	0,000123
0,4	0,85214	1.1	0,29820	3,5	0,000004785
0.5	0,77880	1,2	0,23693	4	0,000000113
0,6	0,69768	1,3	0,18452	5	0,00000000001



Si ahora admitimos, como hicimos en otra ocasión, que el desvío puede variar de un modo continuo entre 0 y $\pm \infty$, no habremos hecho más que reemplazar la poligonal que representa la función discontinua, por la curva asintótica al eje de las abcisas que representa la función continua y que tenderá a confundirse con la poligonal a medida que aumente el valor de n.

En efecto, haciendo $n = 20000$; $p = q = \frac{1}{2}$; resulta :

$$\varepsilon = \frac{1}{\sqrt{2 \times 20000 \times \frac{1}{2} \times \frac{1}{2}}} = 0{,}01.$$

y para $h = 1;\ 2;\ 3;\ \ldots$ se hace $x = 0{,}01;\ 0{,}02;\ 0{,}03\ \ldots$ y la poligonal y la curva se confunden, prácticamente.

Esta curva se conoce con el nombre de curva de campana — debido a su forma —; o curva de los errores — por la aplicación que de ella se hace en la teoría de los errores —; o curva de Gauss — el fundador de esa teoría —; o curva de probabilidad.

Hemos visto ya que entre $-\infty$ y $+\infty$ encierra un área igual a $\sqrt{\pi}$.

Se han calculado tablas de la integral

$$\int_0^\lambda e^{-x^2}\,dx$$

que dan el área encerrada entre la curva, el eje de las abcisas, el de las ordenadas y una ordenada dada que corresponde al valor asignado a x en cada caso.

44. La función

$$\Theta(\lambda) = \frac{2}{\sqrt{\pi}} \int_0^\lambda e^{-x^2}\,dx$$

en la que, como hemos visto, es

$$x = \pm \frac{h}{\sqrt{2npq}} \quad \therefore \quad x^2 = \frac{h^2}{2npq}$$

nos da la suma de las probabilidades de los desvíos comprendidos entre $+h$ y $-h$, de conformidad con los valores particulares que toma h, en cada caso, cuando a x se le da un determinado valor λ.

Damos a continuación una tabla de los valores de la integral :

$$\Theta(\lambda) = \frac{2}{\sqrt{\pi}} \int_0^\lambda e^{-x^2}\,dx.$$

λ	$\Theta(\lambda)$	λ	$\Theta(\lambda)$	λ	$\Theta(\lambda)$
0	0	1.163	0.9	2.327	0.999
0.1	0.1124630	1.2	0.9103140	2.5	0.9995930
0.2	0.2227025	1.3	0.9340080	2.76	0.9999051
0.3	0.3286267	1.4	0.9522851	3	0.9999779
0.4	0.4283922	1.5	0.9661052	3.13	0.9999904
0.4769363	0.5	1.6	0.9763484	3.46	0.9999990
0.5	0.5204999	1.7	0.9837904	3.50	0.9999993
0.6	0.6038561	1.8	0.9890905	3.71	0.9999998
0.7	0.6778010	1.83	0.9903467	3.72	0.9999999
0.8	0.7421010	1.9	0.9927904	...	...
0.9	0.7969082	2	0.9953223	∞	1.
1.—	0.8427008	2.12	0.9972836		

Se advertirá que los valores de la función tienden rápidamente hacia *uno*.

Cuando es $\lambda = 0,4769363$, es $\Theta(\lambda) = 0,5$.

El desvío h, que resulta en este caso particular y que tiene la misma probabilidad de ser que de no ser sobrepasado, *en valor absoluto*, se llama *desvío probable*.

45. Ejemplos : I. ¿Cuántas veces se debería echar al aire una moneda para que la probabilidad de que una cualquiera de las caras saliese un millón de veces más que la otra fuese 0,99?

En este caso, lo que debemos determinar es n.

En la tabla hallamos

$$\Theta(\lambda) = 0,99$$

cuando $\lambda = 1,83$.

Por otra parte, según el enunciado del problema, tenemos

$$h = 1000000 \, ; \quad p = q = \frac{1}{2}.$$

Luego, reemplazando valores en la ecuación

$$\lambda = \frac{h}{\sqrt{2npq}}$$

tenemos

$$1,83 = \frac{1000000}{\sqrt{2n \times \frac{1}{2} \times \frac{1}{2}}} \quad \therefore$$

$$\sqrt{n} = \frac{1000000\sqrt{2}}{1,83} \quad \therefore$$

$$n = \frac{2 \times 1000000^2}{1,83^2}$$

o sea 597211 millones, aproximadamente.

II. Se lanza al aire una moneda 200 veces, y se observa que sale cara m veces. ¿Cuál es la probabilidad de que la frecuencia observada $\frac{m}{20}$ no difiera de la probabilidad teórica, $p = \frac{1}{2}$, en más de un 4 por ciento?

El desvío absoluto no debe pasar de 8 en las 200 experiencias. Luego

$$\lambda = \frac{8}{\sqrt{2 \times 200 \times \frac{1}{2} \times \frac{1}{2}}} = 0.8.$$

Según la tabla

$$\Theta\ (0{,}8) = 0{,}742,$$

luego esa es la probabilidad de que el número de caras que salgan esté comprendido entre 92 y 108.

III. Dos jugadores, A y B, juegan a cara o cruz un determinado número de partidas. A se compromete a pagar a B una suma s si el desvío que se produzca no pasa de 8 en valor absoluto. Por su parte, B se compromete a pagar a A esa misma suma s si el desvío llega a pasar de dicho límite. ¿Cuántas partidas deberán jugar para que el juego sea equitativo?

Para ello las probabilidades de uno y otro han de ser iguales. Hemos de tener, por lo tanto,

$$\Theta\ (\lambda) = \frac{1}{2}$$

lo que ocurre cuando es

$$\lambda = 0{,}4769363$$

Tomemos 0,4769 : resulta entonces

$$0.4769 = \frac{8}{\sqrt{2n\frac{1}{2}\cdot\frac{1}{2}}} \quad \therefore \quad n = \frac{8^2 \times 2}{0.4769^2} = 563.$$

IV. La relación entre el número de nacimientos masculinos y femeninos guarda una cierta constancia que no ha escapado a los estadígrafos. Durante el siglo XVIII se admitía que era de $\frac{25}{24}$; para Laplace era de $\frac{22}{21}$, y en nuestros días se estima en $\frac{18}{17}$. Aceptando como cierta esta última razón, se pregunta cuál es la probabilidad de que sobre un total de 83.300 nacimientos se produzca un desvío *superior* a 475.

Según la hipótesis admitida, las probabilidades respectivas de que nazca un varón o una niña son :

$$p = \frac{18}{35} \quad ; \quad q = \frac{17}{35}.$$

Luego tenemos :

$$\lambda = \frac{475}{\sqrt{2 \times 83300 \times \frac{18}{35} \times \frac{17}{35}}} = \frac{475}{204} = 2{,}328.$$

Para

$$\lambda = 2.327$$

tenemos :

$$\Theta(\lambda) = 0{,}999$$

Esa es la probabilidad de que el desvío *no exceda* de 475. La de que *exceda* dicho límite es, por lo tanto,

$$1 - \Theta(\lambda) = 0{,}001$$

V. Una urna contiene 100 bolillas blancas y 100 negras y se hacen 100 extracciones, reponiendo en cada caso la bolilla extraída. ¿Entre qué límites estará comprendido el desvío cuya probabilidad es 0,997?

Siendo

$$\Theta(\lambda) = 0{,}997 \quad \text{es} \quad \lambda = 2{,}12$$

Luego

$$2.12 = \frac{h}{\sqrt{2npq}} \quad \therefore \quad h = 2.12 \,.\, \sqrt{2}\,\sqrt{npq} = 3\,\sqrt{npq} \quad \therefore$$

$$h = 3\,\mu$$

puesto que, como sabemos, el *desvío medio cuadrático* — que hemos simbolizado por μ — es igual a $\sqrt{npq}$.

En el caso concreto que tenemos es

$$h = 3\sqrt{100 \times \frac{1}{2} \times \frac{1}{2}} = 3 \times 5 = 15.$$

Luego, el número máximo de bolillas de un color dado que, dentro de esos límites, podrán salir es 65. Y, por lo tanto, el menor número de bolillas que podrán salir del otro color es 35. Tales son los límites pedidos.

VI. ¿Qué probabilidad hay de que tirando 200 veces una moneda a cara o cruz, se produzca un desvío superior a $+15$?

Para $h = 15$, tenemos :

$$\lambda = \frac{15}{\sqrt{2 \times 200 \times \frac{1}{2} \times \frac{1}{2}}} = 1.5 \quad \therefore \quad \Theta(1.5) = 0.966.$$

Por lo tanto, la probabilidad de que el desvío *no exceda* de 15 *en valor absoluto*, es 0,966.

La de que exceda de 15 (sin atender al signo) es

$$1 - \Theta(1,5) = 0,034$$

Y la de que exceda, *en un sentido determinado* (el positivo, en este caso) es, evidentemente, la mitad de la anterior. Es decir

$$\frac{1}{2}[1 - \Theta(1.5)] = 0.017.$$

José González Galé.

(Continuará.)