o/4 pasla

Revista

Tomo 5

de

Ciencias Económicas

Publicación mensual del Centro Estudiantes de Ciencias Económicas

Director:

Luciano Carrouché

Administrador:

Miguel G. Di Cio

Secretario de Redacciòn:

Italo Luis Grassi

Redactores:

Mario V. Ponisio - Mauricio E. Greffier - Agustín A. Forné

Jacobo Waisman - Dívico A. A. Fürnkorn

Año III | Julio y Agosto de 1915 | Núm. 25-26

Dirección y Administración

1835 - CALLE CHARCAS - 1835

BUENOS AIRES

La curva de los errores [1]

SUMARIO: I. — El teorema de Bernouilli y la curva de Gauss. II. — El promedio aritmético de los errores. Valores de h y de k que confieren a la probabilidad de los errores o de Gauss su valor máximo. III. — Las tablas de la integral de los errores. Su empleo.

I

I.—Vamos a deducir la ecuación de la curva de los errores o curva binomial o de Gauss, tomando como punto de partida un teorema conocido de cálculo de las probabilidades. Sabemos que si es p la probabilidad de un cierto acontecimiento y q la probabilidad del acontecimiento opuesto, los términos del desarrollo de:

$(p+q)^n$, es decir, p^n, $np^{n-1}q$, $\frac{n(n-1)}{1.2}p^{n-2}q^2, \ldots\ldots\ldots$

$, \frac{n(n-1)}{1.2}p^2q^{n-2}, npq^{n-1}, q^n$ expresan, respectivamente, las probabilidades de que, en n pruebas, el acontecimiento que se considera se presente n, n—1, n—2,..... 2, 1 ó ninguna vez. Así, por ejemplo, si tiramos 6 veces un dado, la probabilidad de que en las 6 veces una cara determinada se presente dos, resulta expresada por $\frac{6.5}{1.2}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^4$

II.—Hay un valor de r que confiere a la expresión

---

(1) Del curso de estadística de la Facultad de Ciencias Económicas. Versión taquigráfica de J. Ignacio Azpiazu.

$\binom{n}{r} p^r q^{n-r}$ un valor máximo. Este valor de r es, como sabemos, np. Es decir, el valor de r que confiere a la probabilidad un valor máximo, es igual al producto del número de pruebas por la probabilidad del acontecimiento considerado.

Luego pues, la expresión $\binom{n}{np} p^{np} q^{n-np} = \binom{n}{np} p^{np} q^{nq}$ es el término de valor máximo del desarrollo de $(p+q)^n$. Ese valor lo expreso por P, por manera que P es la probabilidad que corresponde al acontecimiento más probable. Expreso ahora por P_x la probabilidad de que el acontecimiento no se presente el número de veces más probable np, sinó $np+x$ (1).

III—Transformemos ahora las expresiones de P y P_x. Sabemos que $\binom{n}{np} = \frac{n(n-1)(n-2)\ldots\ldots(n-np+1)}{1.2.3\ldots\ldots np}$, y como $n-np=nq$ tendremos $\binom{n}{np} = \frac{n(n-1)(n-2)\ldots(nq+1)}{1.2.3\ldots np}$

Multiplicando numerador y denominador por $(nq)!$

$$\binom{n}{np} = \frac{n(n-1)(n-2)\ldots(nq+1)\,(nq)!}{1.2.3\ldots\ldots(np)\,(nq)!} = \frac{n!}{(np)!\,(nq)!} \therefore$$

$$P = \binom{n}{np} p^{np} q^{nq} = \frac{n!}{(nq)!\,(np)!} p^{np} q^{nq}$$

Por un procedimiento análogo encontraría para $\binom{n}{np+x}$ una forma semejante a la hallada para $\binom{n}{np}$.

Tendría $\binom{n}{np+x} = \frac{n!}{(np+x)!\,(nq-x)!} \therefore$

$$P_x = \binom{n}{np+x} p^{np+x} q^{nq-x} = \frac{n!}{(np+x)!\,(nq-x)!} p^{np+x} q^{nq-x}$$

IV—¿En qué se diferencian P y P_x? Observemos que el exponente de p en P_x es $np+x$ mientras que el exponente

---

(1) Por ejemplo, siendo $p=\frac{1}{6}$ y $n=6000$ tendré $np=1000$ y si supongo $x=50$ tendré $np+x=1050$: en este caso particular tendremos:

$$P = \binom{6000}{1000}\left(\frac{1}{6}\right)^{1000}\left(\frac{5}{6}\right)^{5000} \text{ y } P_{50} = \binom{6000}{1050}\left(\frac{1}{6}\right)^{1050}\left(\frac{5}{6}\right)^{4950}$$

en P es np, es decir, hay en la expresión P_x un factor p^x que no figura en la expresión P. Analogamente el exponente de q en P_x es nq—x y el exponente de q en P es nq: por lo tanto, en la expresión P_x figura un factor q^{-x} que no figura en la expresión P.

V.—Consideremos ahora los dos coeficientes. Divid endo el coeficiente de P_x por el de P tendremos:

$$\frac{\binom{n}{np+x}}{\binom{n}{np}} = \frac{\frac{n!}{(np+x)!\,(nq-x)!}}{\frac{n!}{(np)!\,(nq)!}} = \frac{\frac{1}{(np+x)!\,(nq-x)!}}{\frac{1}{(np)!\,(nq)!}} =$$

$$= \frac{(np)!\,(nq)!}{(np+x)!\,(nq-x)!}$$

En el numerador tenemos el producto de los np primeros números naturales y en el denominador el producto de los np+x primeros números naturales. Luego: tenemos en el denominador sin que figure en el numerador el producto (np+1) (np+2) (np+3)......(np+x). Por otra parte tengo en el numerador (nq)! y en el denominador (nq—x)! es decir, falta en el denominador y figura en el numerador el producto (nq—x+1) (nq—x+2) (nq—x+3)........(nq).

Efectuando la simplificación, dividiendo numerador y denominador por (np)! (nq—x)! nos queda:

$$\frac{\binom{n}{np+x}}{\binom{n}{np}} = \frac{(nq-x+1)\,(nq-x+2)\,(nq-x+3)\ldots\ \ldots(nq)}{(np+1)\,(np+2)\,(np+3)\ldots(np+x)}$$

Este sería el cociente de los coeficientes binomiales de P_x y de P. El numerador y el denominador son productos de x factores: en otras palabras, si dividimos cada uno de los términos del numerador por nq quedará el numerador dividido por $(nq)^x$ y si dividimos cada uno de los términos del denominador por np el denominador resultará dividido por $(np)^x$. Entonces, pues, dividamos cada uno de los términos del numerador y denominador por nq y np respectivamente y, para que la relación no altere, multipliquemos por $(nq)^x$ el numerador y por $(np)^x$ el denominador.

Tendremos así:

$$\frac{\binom{n}{np+x}}{\binom{n}{np}} = \frac{(nq)^x \left(1 - \frac{x-1}{nq}\right)\left(1 - \frac{x-2}{nq}\right)\ldots\ldots\ldots\ldots 1}{(np)^x \left(1 + \frac{1}{np}\right)\left(1 + \frac{2}{np}\right)\ldots\left(1 + \frac{x}{np}\right)} =$$

$$= \frac{n^x q^x \left(1 - \frac{x-1}{nq}\right)\left(1 - \frac{x-2}{nq}\right)\ldots\ldots\ldots\ldots 1}{n^x p^x \left(1 + \frac{1}{np}\right)\left(+ \frac{2}{np}\right)\ldots\left(1 + \frac{x}{np}\right)} =$$

$$= \frac{q^x \left(1 - \frac{x-1}{nq}\right)\left(1 - \frac{x-2}{nq}\right)\ldots\ldots\ldots\ldots 1}{p^x \left(1 + \frac{1}{np}\right)\left(1 + \frac{2}{np}\right)\ldots\left(1 + \frac{x}{np}\right)}$$

VI.—De IV sacamos como consecuencia que P_x es igual al producto de P por $p^x q^{-x}$ y por el cociente de los dos coeficientes: el coeficiente de P_x dividido por el coeficiente de P. Luego $P_x = P \left(\frac{p}{q}\right)^x \frac{\binom{n}{np+x}}{\binom{n}{np}}$ $\therefore$

$$P_x = P \left(\frac{p}{q}\right)^x \frac{q^x \left(1 - \frac{x-1}{nq}\right)\left(1 - \frac{x-2}{nq}\right)\ldots\ldots\ldots\ldots 1}{p^x \left(1 + \frac{1}{np}\right)\left(1 + \frac{2}{np}\right)\ldots\left(1 + \frac{x}{np}\right)} \therefore$$

$$P_x = P \frac{\left(1 - \frac{x-1}{nq}\right)\left(1 - \frac{x-2}{nq}\right)\ldots\ldots\ldots\ldots 1}{\left(1 + \frac{1}{np}\right)\left(1 + \frac{2}{np}\right)\ldots\left(1 + \frac{x}{np}\right)}$$

P es mayor que P_x porque es el termino del desarrollo de valor máximo. Luego esa fracción será menor que 1.

Tomando logaritmos, tenemos :

$$l\,P_x = l\,P + l\left(1 - \frac{x-1}{nq}\right) + l\left(1 - \frac{x-2}{nq}\right) + \ldots\ldots -$$

$$- l\left(1 + \frac{1}{np}\right) - l\left(1 + \frac{2}{np}\right) - \ldots - l\left(1 + \frac{x}{np}\right)$$

VII. — Ahora, aprovechamos el desarrollo en serie (1) que corresponde a $l\,(1+x)$ y $l\,(1-x)$ para transformar la expresión VI y, haciendo respectivamente:

$$x = \frac{1}{qn},\ x = \frac{1}{pn},\ x = \frac{2}{qn},\ x = \frac{2}{pn}$$

y así sucesivamente, tendremos.

$$l\,P_x = l\,P + \left(-\frac{1}{qn} - \frac{1}{2\,(qn)^2} - \ldots\right) +$$

$$+\left(-\frac{2}{qn} - \frac{4}{2\,(qn)^2} - \ldots\right) + \ldots\ldots +$$

$$+\left(-\frac{x-1}{qn} - \frac{(x-1)^2}{2\,(qn)^2} - \ldots\right) - \left(\frac{1}{pn} - \frac{1}{2\,(pn)^2} + \ldots\right) -$$

$$-\left(\frac{2}{pn} - \frac{4}{2\,(pn)^2} + \ldots\right) - \ldots - \left(\frac{x}{pn} - \frac{x^2}{2\,(pn)^2} + \ldots\right).$$

VIII.—Sumando todos los términos de primer grado y prescindiendo de los restantes que, como veremos, pueden despreciarse; tendremos:

$$l\,P_x = l\,P - \frac{1+2+3+\ldots+(x-1)}{qn} - \frac{1+2+3+\ldots\ldots+x}{pn} + R$$

Llamo R a la suma de los demás términos.

Los numeradores del segundo y tercer término del segundo miembro son sumas de términos que están en progresión aritmética de primer término 1 y de razón 1. Por lo tanto:

$$l\,P_x = l\,P - \frac{x\,(x-1)}{2\,qn} - \frac{x\,(x+1)}{2\,pn} + R$$

---

(1) Sabemos que si desarrollamos $l\,(1+x)$ en una serie infinita de potencias de x tendremos: $l\,(1+x) = \frac{x}{1} - \frac{x^2}{2} + \frac{x^3}{3} - \ldots\ldots$. Es la serie de Taylor que corresponde a $l\,(1+x)$. Si reemplazamos x por $-x$ encontramos: $l\,(1-x) = -\frac{x}{1} - \frac{x^2}{2} - \frac{x^3}{3} - \ldots$ Es decir que mientras la serie de Taylor que corresponde a $l\,(1+x)$ es una serie de signos alternados, la serie de $l\,(1-x)$ es una serie de términos negativos. Ambas representan efectivamente algo, si x es, en valor absoluto menor que 1.

IX.—He dicho en VIII que los términos que no fueran de primer grado los habríamos despreciado. Esos términos, o son cuadráticos, o tienen en el numerador la potencia tercera de x y en el denominador el cuadrado de n multiplicado por cierta cantidad, o tienen en el numerador la potencia cuarta de x y en el denominador la tercera de n y así sucesivamente. Ahora, para que efectivamente puedan despreciarse esos términos, debemos suponer que x es grande respecto á la unidad pero que es pequeña con respecto a n y tan pequeña que el cociente $\frac{x}{n}$ puede despreciarse.

Por otra parte, decir que $\frac{x}{n}$ tiene que ser una cantidad tan pequeña que se la puede despreciar, quiere decir que se consideran probabilidades que difieren del acontecimiento mas probable pero que son tales que el cuociente $\frac{pn+x}{n}$ —la frecuencia referida al número de pruebas—difiere de la probabilidad p en una cantidad tan pequeña que podemos no tomarla en cuenta.

X.—Efectuemos ahora la suma del segundo y tercer término, multiplicando el numerador y el denominador del segundo término por p y el numerador y denominador del tercero por q.

Tendremos así:

$$l\,P_x = l\,P - \left(\frac{p\,.\,x\,(x-1)}{2\,.\,p\,.\,q\,.\,n} + \frac{q\,.\,x\,(x+1)}{2\,.\,p\,.\,q\,.\,n}\right) =$$

$$= l\,P - \frac{p\,(x^2 - x) + q\,(x^2 + x)}{2\,.\,p\,.\,q\,.\,n} = l\,P - \frac{x^2 + x\,(q-p)}{2\,.\,p\,.\,q\,.\,n}$$

Tenemos un término $\frac{x\,(q-p)}{2\,.\,p\,.\,q\,.\,n} = \frac{x}{n}\;\frac{(q-p)}{2\,.\,p\,.\,q}$ que se puede despreciar si $\frac{x}{n}$ es despreciable. Nos queda entonces: $l\,P_x = l\,P - \frac{x^2}{2\,.\,p\,.\,q\,.\,n}$

Y pasando de los logaritmos a los números, tenemos que:

$$P_x = \frac{P}{e^{\frac{x^2}{2\,pq\,n}}} = P\,e^{-\frac{x^2}{2\,np\,q}}$$

pn

XI.—Es decir, que si es muy grande el número n de pruebas la probabilidad P_x de que el acontecimiento que se considera se presente un número de veces igual a $pn + x$ —donde x es grande respecto a 1 pero chico respecto a n—tiene por expresión aproximada $P_x = Pe^{-\frac{x^2}{2pqn}}$

Y como $P = \frac{1}{\sqrt{2.\pi.n.p.q}}$; resulta

$$P_x = \frac{1}{\sqrt{2.\pi.n.p.q}}\, e^{-\frac{x^2}{2pnq}} \therefore$$

$$P_x = \frac{1}{\sqrt{\pi}\sqrt{2.n.p.q}}\, e^{-\frac{x^2}{2.n.p.q}}$$ (1) Llamaremos «precisión» el número h definido por la relación $\frac{1}{\sqrt{2.p.q.n}} = h$

Reemplazando en la expresión de P_x, tendremos:

$$P_x = \frac{h}{\sqrt{\pi}}\, e^{-x^2 h^2}$$

---

(1) Supongamos que se hacen 15000 pruebas de un acontecimiento de probabilidad 2/3. El número de veces, más probable, en que el acontecimiento se presentará, es: np=1000. ¿Cuál, es la probabilidad de que el acontecimiento se presente 10149 veces?, es decir, ¿cuál es la probabilidad de que se produzca un desvío de 149?

Haciendo $x=149$, $n=15000$, $p=\frac{2}{3}$, $q=\frac{1}{3}$ y sustituyendo en la fórmula hallada, tenemos:

$$P_{149} = \frac{1}{\sqrt{2\pi\, 15000\, \frac{2}{3}\, \frac{1}{3}}}\, e^{-149^2 \frac{1}{2.15000.\frac{2}{3}.\frac{1}{3}}} =$$

$$= \frac{1}{\sqrt{\pi\, .\, 15000\, .\, \frac{4}{9}}}\, e^{-149^2 \frac{1}{6666,66}}$$

$$P_{149} = \frac{1}{\sqrt{\pi}\sqrt{15000\,\frac{2}{3}}}\, e^{-149^2 \frac{1}{6666,66}} = \frac{1}{\frac{246}{3}\sqrt{\pi}}\, e^{-149^2 \frac{1}{6666.66}}$$

Esa expresión nos daría la probabilidad buscada.

XII. — Vamos a representar gráficamente la función $P_x = \frac{h}{\sqrt{\pi}} e^{-x^2 h^2} = y$. A $x = o$ corresponde $y =$ máximo; a $x = \pm \infty$, $y = o$. El valor de y disminuye con el crecer de x.

Como el valor de P_x no varia si se pone $-x$ en lugar de x porque x figura en la expresión con exponente 2, tendremos a la izquierda ordenadas iguales a las correspondientes ordenadas de la derecha.

Si reunimos por medio de una linea los puntos encontrados, tenemos la curva de los errores que expresa gráficamente el valor de las probabilidades P. (fig. 1)

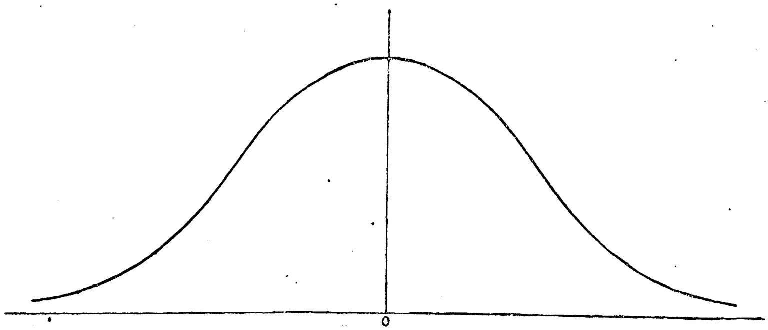

Fig 1

La curva será mas o menos abierta según que h sea mayor o menor.

Esa curva se llama «curva de los errores», «binomial» o «curva de Gauss», o en estadística, «curva de Quételet».

XIII.—Sabemos (II) que si se hacen n pruebas de un acontecimiento dado, lo más probable es que él se presente np veces, pero la probabilidad de que se presente np veces tiende a cero con el crecer de n y es expresada por

$$\frac{1}{\sqrt{2 . \pi . n . p . q}}$$

Hemos encontrado (XI) que la probabilidad de que el acontecimiento no se presente np, sino $np + x$ veces nos dá como expresión aproximada: $\frac{1}{\sqrt{2 . \pi . n . p . q}} e^{-\frac{x^2}{2 n . p . q}}$

Esa probabilidad también tiende a cero a medida que crece n. Pero, ¿qué quiere decir esa expresión? Quiere decir que si n es muy grande, la probabilidad de que el acontecimiento se presente un número de veces contenido entre $pn + x$ y $pn + x + dx$ resulta expresada por $P_x . dx$, en donde dx es la amplitud del intervalo. En la figura 2 a $P_x . dx$ corresponde el espacio sombreado.

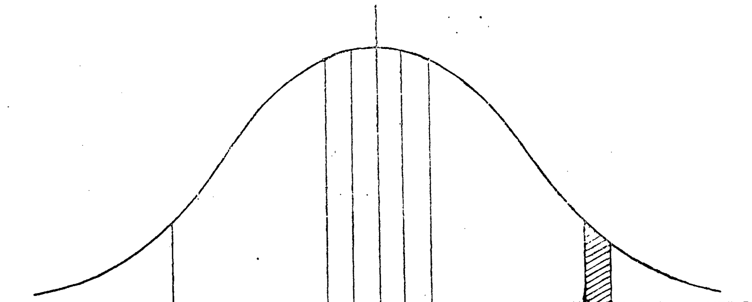

Fig 2

Si suponemos dx muy pequeño podemos considerar como iguales las ordenadas que corresponden a x y a $x + dx$, es decir, podemos suponer que ese espacio es un rectángulo. Es lo que se hace cuando se reemplazan las expresiones verdaderas por los infinitésimos del cálculo diferencial.

XIV.—¿Cuál será la probabilidad de que el acontecimiento se presente un número de veces contenido entre $pn - x$ y $pn + x$? La probabilidad de que se presente un número de veces comprendido entre pn y $pn + x$ será la superficie comprendida entre o, x y la curva, y la de que se presente entre pn y $pn - x$, la superficie comprendida entro $o, - x$ y la curva. Analiticamente, la superficie resulta expresada por una integral. La probabilidad buscada será la suma de todos los rectángulos tomados entre los límites $- x$ y $+ x$ es decir $\int_{-x}^{+x} P_x . dx$.

Y reemplazando el valor de P_x tendremos:

$$\frac{1}{\sqrt{2 . \pi . n . p . q}} \int_{-x}^{+x} e^{-\frac{x^2}{2npq}} dx$$

La expresión que figura dentro del signo de integral no cambia si se reemplaza x por $-x$. Es decir:

$$\frac{1}{\sqrt{2.\pi.n.p.q}}\int_{o}^{x} e^{-\frac{x^2}{2\,npq}}\,dx =$$

$$= \frac{1}{\sqrt{2.\pi.n.p.q}}\int_{-x}^{o} e^{-\frac{x^2}{2\,npq}}\,dx \;\therefore$$

$$\frac{1}{\sqrt{2.\pi.n.p.q}}\int_{-x}^{+x} e^{-\frac{x^2}{2\,npq}}\,dx =$$

$$= \frac{2}{\sqrt{2.\pi.n.p.q}}\int_{o}^{x} e^{-\frac{x^2}{2\,npq}}\,dx \quad (1)$$

XV.—¿Qué relación hay entre esta formula y el teorema de Bernouilli? Si hacemos n pruebas de un acontecimiento de probabilidad p lo más probable es que el acontecimiento se presente np veces: si el número de las pruebas es muy grande, la probabilidad de ese acontecimiento es muy pequeña. Por otra parte, cuando se hace una previsión, no interesa prever con toda exactitud lo que sucederá, lo que interesa es poder indicar ciertos límites entre los cuales estará contenido el acontecimiento a verificarse (2). Ahora la probabilidad de que el acontecimiento, no se presente exactamente np veces, sino que se presente un número de veces contenido entre $np - x$ y $np + x$ es la suma de las probabilidades de que se presente $np - x$, $np - x + 1$, $np - x + 2$, ... np, $np + 1$, $np + 2$, ... $np + x - 1$,

(1) Si hacemos $x = \infty$ y tomamos la integral entre $-\infty$ y $+\infty$ el valor de la integral será 1. Tenemos la certeza de que el acontecimiento se presentará un número de veces comprendido entre o y el número total de las pruebas.

(2) Si yo soy asegurador, no me interesa saber sí de 10000 asegurados morirán exactamente 1000, sinó, saber que morirá un número de asegurados contenido, por ejemplo, entre 950 y 1050.

np + x. Esta probabilidad resulta expresada (fig. 3) por la superficie comprendida entre las ordenadas que corresponden á — x y + x y la curva de Gauss. Si suponemos que la ordenada A expresa la probabilidad de que el acontecimiento se presente pn + x veces y la ordenada B la de que se presente pn — x, la superficie comprendida entre las dos ordenadas y la curva, expresa la probabilidad de que el acontecimiento se presente un número de veces comprendido entre pn — x y pn + x y el teorema de Bernouilli nos dice que esa probabilidad es casi igual a 1.

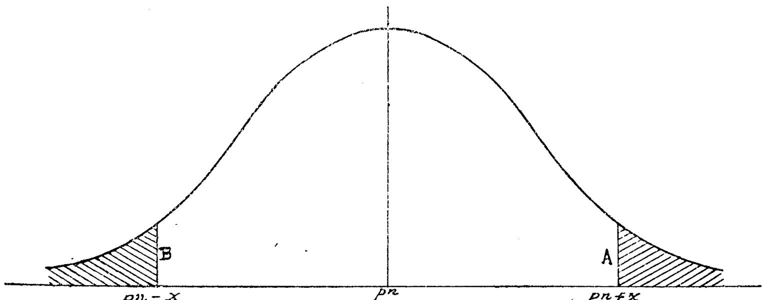

Fig. 3

Luego, la probabilidad de que el acontecimiento se presente más de pn + x veces o menos de pn — x, que nos es dada por la superficie sombreada, puede despreciarse respecto de la probabilidad de que efectivamente se presente un número de veces contenido entre esos límites, la que nos es dada por la superficie comprendida entre las ordenadas A y B y la curva. Si crece n, dice el teorema de Bernouilli, puede crecer también x puede ser el error en la previsión muy grande, pero, por grande que sea el error en la previsión, podemos estar seguros de que tendrá que ser pequeño el cociente $\frac{x}{n}$, es decir: error contenido en la previsión dividido por el número de pruebas verificadas (1).

Así mismo el teorema de Bernouilli nos dice que si es p la probabilidad de un acontecimiento y el número de prue-

(1) Si tengo 10000 asegurados y sé que el número de los muertos en el año estará contenido entre 950 y 1050, x no es chico en sí (x=50); lo que es chico es $\frac{50}{10.000} = 0,005$.

bas, n, es muy grande, podemos estar casi seguros de que el número de veces en que el acontecimiento se presentará: $pn \pm x$, dividido por el total de pruebas verificadas, n, : $\frac{pn \pm x}{n}$ diferirá muy poco de p (1).

En efecto, si divido el número de veces en los cuales el acontecimiento se presenta, por n tendré como límite inferior a $p - \frac{x}{n}$ y como límite superior a $p + \frac{x}{n}$. Es decir, que yo estoy casi seguro que el cociente del número de veces en los cuales el acontecimiento se presenta dividido por el número de pruebas, se encuentra contenido entre los límites $p - \frac{x}{n}$ que es pequeño y $p + \frac{x}{n}$ que también es pequeño, es decir, que se encuentra contenido entre límites muy reducidos.

Hugo BROGGI.

(Continuará).

(1) Por ejemplo: Tratándose de probabilidades de muerte y siendo el nùmero de individuos, considerado, muy grande, podemos estar casi seguros de que el número total de los que mueran, dividido por el número total de los individuos considerados diferirá muy poco de la probabilidad de que uno de dichos individuos muera.