UN ESTUDIO DE LAS ESTADISTICAS g y d

por GUIDO O. G. LISERRE, Rosario (Argentina)

I. Estas estadísticas, de las cuales en el presente damos las funciones generatrices, han sido utilizadas por la División de Asuntos Económicos de la Embajada Americana en Londres, al comienzo del año 1943 para estimar la producción de guerra alemana, mediante el análisis de numeraciones y series de números del equipo alemán capturado en combate.

II. Sea la población constituída por los elementos

$$1, 2, 3, \ldots, p. \tag{1}$$

De esa población extraemos al azar muestras de n elementos en block, siendo $n < p$, y decimos que una muestra es distinta de otra cuando se diferencia por lo menos en la naturaleza de un elemento. El número de muestras distintas que se podrá formar será de

$$\binom{p}{n}. \tag{2}$$

Si ordenamos los elementos que constituyen cada muestra en forma creciente, cada una de ellas sólo podrá terminar en alguno de los números $n, n+1, n+2, \ldots, n+k = p$.

Ha nacido así la variable *mayor elemento de la muestra,* que la indicamos con g;

$$n \leqq g \leqq n + k. \tag{3}$$

Haciendo el cambio de variable $g=n+v$, podemos expresar la variación anterior así:

$$0 \leqq v \leqq k. \tag{4}$$

La probabilidad de que una muestra termine en un valor determinado de g estará dada por

$$P\{g|n, p\} = \frac{\binom{g-1}{n-1}}{\binom{p}{n}}. \tag{5}$$

La expresión anterior se determina aplicando la definición de probabilidad como cociente del número de casos favorables sobre el número de casos posibles; el número de casos posibles está dado por (2). Determinamos a continuación el número de casos favorables. Si el mayor elemento de la muestra es g, significa que antes del mismo se pueden colocar $g-1$ números, de los cuales sólo hay $n-1$ pues el enésimo es g, que forma parte de la muestra; habrá entonces tantas muestras que terminan en g como $\binom{g-1}{n-1}$. Hemos determinado el número de casos favorables y por lo tanto queda justificada la (5).

Ha nacido así la variable aleatoria g. Podemos, por lo tanto, determinar su función generatriz

$$F(g, u) = \sum_{g=n}^{g=p} \frac{\binom{g-1}{n-1}}{\binom{p}{n}} e^{ug}$$

aplicando a la anterior la transformación $g=n+v$ tendremos:

$$F(g=n+v, u) = \frac{e^{un}}{\binom{p}{n}} \sum_{v=0}^{v=k} \binom{n+v-1}{n-1} e^{uv}. \tag{6}$$

Por la fórmula de Cauchy podemos escribir

$$\binom{n+v-1}{n-1} = \frac{1}{2\pi i} \oint_{\Gamma} \frac{z^{n+v-1}}{(z-1)^n} dz$$

donde Γ es una circunferencia con centro en la unidad. Substituyendo en (6) resulta

$$F(g,u)=\frac{e^{un}}{\binom{p}{n}2\pi i}\oint_{\Gamma}\frac{z^{n-1}}{(z-1)^{n}}\cdot\frac{(ze^{u})^{k+1}-1}{ze^{u}-1}\,dz. \qquad (7)$$

Derivando sucesivamente esta función generatriz y calculando la derivada en el punto $u=0$ obtenemos los momentos de la variable aleatoria g. Así tendremos:

$$E(g)=\frac{(p+1)n}{n+1}.$$

III. Si de cada muestra antes obtenida restamos al valor mayor el valor menor tendremos el ancho de la muestra, que representamos con la letra d. Ha nacido así la variable d que puede tomar los valores

$$n-1, n, n+1, \ldots, n+v, \ldots, p-1 \quad \text{o sea}$$

$$n-1 \leqq d \leqq p-1.$$

El primero se presenta cuando los n elementos de la muestra son los n primeros elementos de la población y el extremo superior se presenta cuando en los n elementos de la muestra aparecen los extremos de la población.

Para determinar la probabilidad de que se presente una muestra con $d=n+v$ comenzamos por averiguar cuántas veces se puede presentar ese valor en las $\binom{p}{n}$ muestras. Esa diferencia se presenta cuando se presentan muestras cuyo elemento mayor sea $n+v+1$ y el menor 1, o cuando el mayor sea $n+v+2$ y el menor 2, etc., o finalmente cuando el mayor sea p y el menor $p-(n+v)$. De lo visto se deduce que el valor $n+v$ se presenta en $p-(n+v)$ casos; dentro de cada caso hay que averiguar cuantas muestras comienzan con $p-(n+v)$ y terminan con p. Hay $n+v-1$ valores mayores que el menor y menores que el mayor; pero de ésos sólo hay que colocar

$n-2$ para formar la muestra; luego se pueden formar tantos dentro de cada caso como $\binom{n+v-1}{n-2}$.

Así hemos determinado que habrá tantas muestras cuya diferencia es $n+v$ como

$$\{p-(n+v)\}\binom{n+v-1}{n-2}.$$

Reemplazando en la anterior $n+v=d$ y aplicando la definición de probabilidad obtenemos

$$P\{d|n,p\}=\frac{(p-d)\binom{d-1}{n-2}}{\binom{p}{n}} \tag{8}$$

ley de distribución de la variable aleatoria d, por lo tanto, determinamos su función generatriz

$$F(d,u)=\sum_{d=n-1}^{d=p-1}\frac{(p-d)\binom{d-1}{n-2}}{\binom{p}{n}}e^{du}=$$

$$=\sum_{d=n-1}^{d=p-1}\frac{p}{\binom{p}{n}}\binom{d-1}{n-2}e^{du}-\frac{1}{\binom{p}{n}}\sum_{d=n-1}^{d=p-1}d\binom{d-1}{n-2}e^{du}$$

escribimos $\binom{d-1}{n-2}$ y $\binom{d}{n-1}$ por la fórmula de Cauchy, substituímos en la anterior y resulta

$$F(d,u)=\frac{e^{u(n-1)}}{2\pi i\binom{p}{n}}\oint_{\Gamma}\left\{p-(n-1)\frac{z}{z-1}\right\}$$

$$\frac{z^{n-2}}{(z-1)^{n-1}}\cdot\frac{(ze^{u})^{p-n+1}-1}{ze^{u}-1}dz \tag{9}$$

expresión de la función generatriz de la variable aleatoria d.

Derivando la función generatriz y calculando el valor de la derivada en el punto $u = 0$ obtenemos

$$E(d) = \frac{(p+1)(n-1)}{(n+1)}.$$

Las estadísticas g y d aquí estudiadas sirven para derivar los estimadores más eficientes del parámetro p.

BIBLIOGRAFIA

L. A. GOODMAN, *Serial Number Analysis*, Journnal of the American Statistical Association, vol. 47, Nº 260.

C. E. DIEULEFAIT, *Determinación de los momentos de las probabilidades hipergeométricas Ordinarias y en el caso de Contagio Polya*, Anales de la Sociedad Científica Argentina, t. CXXVII, págs. 108-117.