

LA DISTRIBUCIÓN DE WARING Y SU APLICACIÓN A MAGNITUDES ECONÓMICAS.

José Rodríguez Avi

Departamento de Estadística e I.O. Universidad de Jaén.

M^a Victoria Reinoso Herrero.

Abstract

La Distribución Univariante Generalizada de Waring, $UGWD(a,k,p)$ es una distribución discreta perteneciente a la familia de distribuciones univariantes de Pearson. Debido a ello puede obtenerse a partir de la solución de una ecuación en diferencias con coeficientes polinomiales, en este caso de orden 2. Igualmente, la estimación de sus parámetros puede realizarse a partir de las relaciones de recurrencia entre sus momentos. Es también de interés el estudio de la descomposición de la varianza en tres factores: riesgo, predisposición y aleatoriedad, y, en determinadas situaciones, sirve para modelizar variables discretas fuertemente asimétricas.

En este trabajo presentamos una discusión de tal distribución así como la posibilidad de su aplicabilidad a magnitudes discretas fuertemente asimétricas, como, por ejemplo, el número de entidades bancarias por municipios en diversas provincias de Andalucía.

Palabras clave: Distribución de Waring, Entidades bancarias, Estudios Municipales

1. INTRODUCCIÓN.

La distribución Univariante Generalizada de Waring $UGWD(a,k,p)$ es una distribución discreta perteneciente a la familia de distribuciones de Pearson generadas por la función Hipergeométrica univariante de Gauss. Es una distribución con rango infinito y que converge a la distribución normal cuando los parámetros tienden a infinito. No obstante, si uno de los dos primeros parámetros -a los que llamaremos *parámetros del numerador*- es suficientemente pequeño, la distribución presenta una gran asimetría a la derecha, lo que la hace válida para modelar situaciones aleatorias reales en las que se presenta esta asimetría.

Esta distribución ha sido estudiada, entre otros por IRWING(1968, 1975) el cual la aplicó para

obtener modelos teóricos para distribuciones de accidentes, así como para obtener separadamente los elementos de riesgo y responsabilidad en el accidente. Así en IRWING(1968) ajusta una distribución de este tipo para al número de accidentes de trabajadores en una fábrica de jabón. Del mismo modo en XECALAKI(1984) aparece esta distribución ajustada a datos de accidentes de carretera en los que había involucrados conductores de autobús en Irlanda del Norte entre 1952-1953 y 1954-55. Un ejemplo de aplicación en otros campos está en GUTIÉRREZ-RODRÍGUEZ(1993) en el que se describe el nº de goles marcados por jugadores de primera división de la Liga española de fútbol en la temporada 89-90.

2.- LA DISTRIBUCIÓN UNIVARIANTE GENERALIZADA DE WARING.

La familia de distribuciones discretas univariantes de Pearson se obtiene a partir de la solución de la ecuación en diferencias

$$G(r) f_{r+1} - L(r) f_r = 0$$

en donde:

$$\begin{aligned} L: \mathbb{Z}^+ &\rightarrow \mathbb{R} \\ G: \mathbb{Z}^+ &\rightarrow \mathbb{R} - \{0\} \end{aligned}$$

son funciones dadas en principio cualesquiera.

Si consideramos como coeficientes de la anterior ecuación en diferencias los polinomios

$$G(r) = (a + k + r)(r + 1) \quad L(r) = (a + r)(k + r)$$

con a, m, p números reales positivos, la solución viene dada por:

$$f_r = f_0 \frac{(a)_r (k)_r}{(a + k + r)_r r!} \quad r \geq 0$$

en donde $(a)_r = a(a-1)\dots(a-r+1)$ es el símbolo de Pochhammer. La suma de la correspondiente serie es convergente ya que la función de masa de probabilidad es el término r-ésimo de una función hipergeométrica de Gauss ${}_2F_1(a, k; a+k+p; 1)$ y ésta es convergente siempre que $p > 0$.

La constante normalizadora f_0 se obtiene a partir del teorema de sumación de Gauss:

$$f_0 = \frac{1}{{}_2F_1(a, k; a + k + r; 1)} = \frac{\Gamma(k + r) \Gamma(a + r)}{\Gamma(a + k + r) \Gamma(r)}$$

Si calculamos el resto de sus características obtenemos: (FAJARDO(1986), GUTIÉRREZ-

RODRÍGUEZ(1993)):

a) Su función generatriz de probabilidad es:

$$g(t) = \frac{{}_2F_1(a, k; a+k+r; t)}{{}_2F_1(a, k; a+k+r; 1)}$$

y por tanto es una distribución que pertenece a la familia de distribuciones discretas generadas por la función hipergeométrica de Gauss.

La ecuación diferencial que verifica la f.g.p.

$$\left[(1-t)q + [a+k+r-1-t(a+k)]q - tak \right] g = 0$$

A partir de aquí podemos calcular la relación de recurrencias entre momentos, la cual es la base para la aplicación del método de los momentos a la hora de estimar los parámetros. Tal relación es:

$$\begin{aligned} a_1(q_1 - q_2) - q_3 &= 0 \\ a_2q_1 - (a_2 + a_1)q_2 - (a_1 + 1)q_3 &= a_2 \\ a_3q_1 - (a_3 + 2a_2 + a_1)q_2 - (a_2 + 2a_1 + 1)q_3 &= 2a_3 + a_2 \end{aligned}$$

en donde

$$q_1 = a + k + r - 1 ; \quad q_2 = a + k ; \quad q_3 = ak$$

y α_i es el momento respecto al origen de orden i .

3.- APLICACIÓN A LA MODELIZACIÓN DEL Nº DE ENTIDADES BANCARIAS POR MUNICIPIO EN ANDALUCÍA.

Vamos a presentar un ejemplo de utilización de esta distribución para describir la variable **X = número de entidades bancarias por municipio en Andalucía (excluyendo las capitales de provincia)**¹. Los datos de partida aparecen reflejados en la Tabla I, en la que se muestra el conteo de

¹Fuente: Anuario Banesto, 1992.

frecuencias para los 759 municipios de Andalucía, una vez excluidas las ocho capitales de provincia.

Analizando la Tabla I podemos observar como los datos presentan una fuerte asimetría a la derecha, con la presencia de varios valores muy superiores a 22, cuando la moda y la mediana es el valor 0.

X	n	f_i	F_i^*
0	411	0.5415	0.5415
1	121	0.1594	0.7009
2	74	0.0975	0.7984
3	40	0.0527	0.8511
4	32	0.0422	0.8933
5	14	0.0184	0.9117
6	23	0.0303	0.9420
7	16	0.0211	0.9631
8	3	0.0039	0.9670
9	5	0.0066	0.9736
10	4	0.0053	0.9789
11	2	0.0026	0.9815
12	0	0.0000	0.9815
13	2	0.0026	0.9841
14	2	0.0026	0.9867
15	1	0.0013	0.9880
16	1	0.0013	0.9893
17	0	0.0000	0.9893
18	1	0.0013	0.9906
19	0	0.0000	0.9906
20	2	0.0026	0.9932
21	1	0.0013	0.9945
22 +	4	0.0055	1

TABLA 1

Vamos a proceder a la estimación de parámetros para ver si se puede ajustar a este modelo una distribución de Waring². Para ello vamos a emplear la ecuación (7), reemplazando los momentos por sus correspondientes estimaciones muestrales.

²Todos los cálculos necesarios se han efectuado mediante la elaboración de un programa en MATLAB.

Los tres primeros momentos con respecto al origen son:

$$\hat{\alpha}_1 = 1.6772; \quad \hat{\alpha}_2 = 16.2016; \quad \hat{\alpha}_3 = 339.8037$$

En consecuencia, el sistema de ecuaciones (7) necesario para la estimación es

$$\begin{array}{rrrrr} 1.6772\theta_1 & -1.6772\theta_2 & -\theta_3 & = & 0 \\ 16.2016\theta_1 & -17.8788\theta_2 & -2.6772\theta_3 & = & 16.2016 \\ 339.8037\theta_1 & -373.8841\theta_2 & -20.5560\theta_3 & = & 695.8090 \end{array}$$

cuya solución es:

$$\hat{\theta}_1 = 33.7874; \quad \hat{\theta}_2 = 28.3447; \quad \hat{\theta}_3 = 9.1285$$

Para obtener los estimadores de los parámetros, resolvemos el sistema (8). Por tanto, los parámetros estimados son:

$$\hat{a} = 28.0189$$

$$\hat{k} = 0.3258$$

$$\hat{\rho} = 6.4427$$

Esto indica que el modelo propuesto es una distribución UGWD(28.0819, 0.3252; 6.4427). Para calcular la bondad del ajuste vamos a realizar tanto el test χ^2 como el de Kolmogorov-Smirnov.

En la Tabla II se refleja la función de masa de probabilidad y la función de densidad de la distribución anteriormente citada. Podemos comprobar como, en efecto, esta distribución presenta igualmente una fuerte aimetría, ya que los valores por encima de 22 engloban aproximadamente un 5% de la probabilidad total . Igualmente el valor mediano es 0 así como el valor modal.

X	f_i	F_i
0	0.5710	0.5710
1	0.1498	0.7209
2	0.0805	0.8014
3	0.0510	0.8524
4	0.0348	0.8871
5	0.0248	0.9120
6	0.0183	0.9303
7	0.0138	0.9441
8	0.0106	0.9546
9	0.0082	0.9629
10	0.0065	0.9694
11	0.0052	0.9746
12	0.0042	0.9787
13	0.0034	0.9821
14	0.0028	0.9849
15	0.0023	0.9871
16	0.0019	0.9890
17	0.0016	0.9906
18	0.0013	0.9919
19	0.0011	0.9930
20	0.0009	0.9939
21	0.0008	0.9947
22 +	0.0053	1

TABLA 2

a) Contrastes de Bondad de Ajuste.

Para realizar este test agrupamos de manera que ninguna casilla de frecuencias esperadas sea menor de 2. Los datos se recogen en la Tabla III. Como podemos observar, el valor proporcionado por el estadístico del test es 24.61 para una distribución χ^2 con 15 grados de libertad, lo que proporciona un valor de p superior a 0.05, concretamente $p=0.055$. Si bien este valor está próximo a 0.05 es de destacar por una parte el gran tamaño de n (759) así como el hecho de que la mayor carga de los residuos se concentra en $X=6$, comportándose muy bien en la mayoría de los casos, y en especial en las colas.

X	O _i	E _i	χ_i^*
0	411	433.4	1.15
1	121	113.8	0.45
2	74	61.1	2.72
3	40	38.7	0.04
4	32	26.4	1.18
5	14	18.8	1.22
6	23	13.9	5.95
7	16	10.5	2.88
8	3	8.1	3.20
9	5	6.2	0.23
10	4	4.9	0.16
11	2	3.9	0.92
12	0	3.1	3.10
13	2	2.5	0.10
14	2	2.1	0.00
15-16	2	3.2	0.45
17-18	1	2.1	0.57
19-21	3	2.2	0.29
21- +	4	4.02	0.00

TABLA 3

Del mismo modo, en la Tabla 5 se recoge la realización del test de Bondad de Ajuste de Kolmogorov-Smirnov. En ella

$$D_i = | F(x_i) - F_n^*(x_i) |$$

y el estadístico del test, $D_n = 0.0295$. Si calculamos la probabilidad de rechazar la hipótesis de ajuste, $p > 0.20$, por tanto podemos aceptar que los datos sigan esta distribución.

Esta es la distribución que mejor se le ajusta. Así, si intentamos el ajuste por medio de otras distribuciones discretas, como la Poisson o la geométrica, los estadísticos del test y la significación aparecen en la Tabla siguiente

DISTRIBUCIÓN	χ^2	p	K-S	p
Geométrica	38.83	3.6×10^{-9}	0.367199	0
Poisson	-	-	0.363029	0

TABLA 4

X	F(x)	F _n [*]	D _n
0	0.5710	0.5415	0.0295
1	0.7209	0.7009	0.0200
2	0.8014	0.7984	0.0030
3	0.8524	0.8511	0.0013
4	0.8871	0.8933	0.0062
5	0.9120	0.9117	0.0003
6	0.9303	0.9420	0.0117
7	0.9441	0.9631	0.0190
8	0.9546	0.9670	0.0124
9	0.9629	0.9736	0.0107
10	0.9694	0.9789	0.0095
11	0.9746	0.9815	0.0069
12	0.9787	0.9815	0.0028
13	0.9821	0.9841	0.0020
14	0.9849	0.9867	0.0018
15	0.9871	0.9880	0.0009
16	0.9890	0.9893	0.0003
17	0.9906	0.9893	0.0013
18	0.9919	0.9906	0.0013
19	0.9930	0.9906	0.0024
20	0.9939	0.9932	0.0007
21	0.9947	0.9945	0.0002
22 +			< 0.0055

TABLA 5

Dado que podemos suponer adecuado el ajuste, vamos a descomponer la varianza en sus tres factores de aleatoriedad, riesgo y predisposición, de la siguiente manera:

$$\mathbf{s}^2 = \mathbf{s}_I^2 + k^2 \mathbf{s}_n^2 + \mathbf{s}_R^2$$

en donde el primer término es el riesgo, el segundo la predisposición y el tercero la aleatoriedad. Si los estimamos:

$$\hat{s}_1^2 = \frac{\hat{a}\hat{k}(\hat{a}+1)}{(\hat{r}-1)(\hat{r}-2)} = 10.9552$$

$$\hat{k}^2 \hat{s}_n^2 = \frac{\hat{a}(\hat{a}+\hat{r}-1)}{(\hat{r}-1)^2(\hat{r}-2)} = 2.3209$$

$$\hat{s}_R^2 = \frac{\hat{a}\hat{k}}{\hat{r}-1} = 1.4169$$

$$\hat{s}^2 = \frac{\hat{a}\hat{k}(\hat{a}+\hat{r}-1)(\hat{k}+\hat{r}-1)}{(\hat{r}-1)^2(\hat{r}-2)} = 14.693$$

lo que indica que es muy pequeño el grado de aleatoriedad en la distribución. Es decir, la variabilidad en el número de entidades bancarias por municipio presenta un fuerte componente no aleatorio debido a dos factores que pueden ser la población y la actividad económica ("riesgo" y "predisposición"). Lo que no podemos determinar, debido al carácter simétrico de a y k en la distribución es qué cantidad corresponde a cada uno de ellos. Para esto es necesario un análisis mediante una variable bidimensional.

BIBLIOGRAFÍA

FAJARDO CALERA, M.A. (1985). **Generalizaciones de los sistemas de Pearson discretos**. Universidad de Extremadura.

GUELFOND, A.O. (1963). **Calcul des différences finites**. Dunod.

GUTIÉRREZ, R y RODRÍGUEZ, J (1993): **Inclusion of The Univariate Generalized Waring Distribution To The Pearson Family. Application To Situations In The Sport**. *Communications in Statistics (En espera de publicación)*

IRWING, J.O. (1968). **The generalized Waring distribution applied to accident theory**. J. R. Statist. Soc. A, 131, pp 205-225.

IRWING, J.O. (1975). **The generalized Waring Distribution**. J. R. Statist. Soc. A, 138 pp 18-31 (part I), 204-227 (part II), 374-378 (part III).

JOHNSON, N.L.; KOTZ, S; KEMP, A.W. (1993). **Univariate Discrete Distributions. (Second ed.)** Ed. Wiley, New York.

JORDAN, C. (1968). **Calculus on finite differences.** Chelsea.

KEMP, A.W. & KEMP, C.D. (1975). **Models for Gaussian hypergeometric distributions.** G. P. Patil (ed.). Statistical distributions in scientific work. Vol 1 pp 31-40 by D. Reidel.

XECALAKI, E (1983). **The Univariate Generalized Waring Distribution in relation to accident theory: proneness, spells or contagion?**. Biometrika 39(3) pp 887-895.