

# GENERACIÓN Y ESTUDIO DE LA DISTRIBUCIÓN DE POLYA MEDIANTE FUNCIONES HIPERGEOMETRICAS UNIVARIANTES.-

*Ramón Gutiérrez Jaímez (U. Granada)*

*Juan Cuadra Carreño (U. Almería)*

*Francisco Herrera Cuadra (U. Almería)*

## I.- INTRODUCCIÓN.-

Existen diversos métodos que permiten construir Distribuciones Discretas Univariantes que generalizan los Sistemas de Pearson. El más conocido consiste en variar los coeficientes de la siguiente Ecuación en Diferencias:

$$G(r).f_{r+1} - L(r).f_r = 0 \quad (r \in \mathbb{Z}^+) \quad (*)$$

siendo  $L: \mathbb{Z}^+ \longrightarrow \mathbf{R}$  ;  $G: \mathbb{Z}^+ \longrightarrow \mathbf{R} - \{0\}$

funciones cualesquiera.

Se sabe por diversos autores, que si  $L$  y  $G$  son determinados polinomios en  $r$ , entonces las Funciones Generatrices de Probabilidad asociadas a las soluciones de la ecuación (\*) vienen expresadas por Series Hipergeométricas.

Concretamente, si en la Ecuación en Diferencias (\*) tomamos las funciones  $L$  y  $G$  como los polinomios:

$$\begin{aligned} G(r) &= (g+r)(r+1) \\ L(r) &= (a+r)(b+r).1 \end{aligned} \quad (**)$$

se obtiene la siguiente Ecuación en Diferencias:

$$(g+r).(r+1).f_{r+1} - (a+r).(b+r).1.f_r = 0$$

con  $\alpha, \beta, \gamma, \lambda \in \mathbf{R} - \{0\}$  y  $\gamma \notin \mathbb{Z}^-$ .

Dicha Ecuación en Diferencias, da lugar a una Familia de Funciones que pertenecen al Sistema dado, y que es conocida como **Familia de Distribuciones Discretas Hipergeométricas**.

Para la obtención de las Funciones de Cuantía y de las Funciones Generatrices de Probabilidad de esta Familia, necesitaremos utilizar Funciones Hipergeométricas, concretamente la Función Hipergeométrica de Gauss:

$${}_2F_1(\mathbf{a}, \mathbf{b}, \mathbf{g}; z) = \sum_{n=0}^{\infty} \frac{[\mathbf{a}]_n \cdot [\mathbf{b}]_n}{[\mathbf{g}]_n} \cdot \frac{z^n}{n!}$$

con  $z \in \mathbf{C}; \quad \alpha, \beta, \gamma \in \mathbf{C} \quad / \quad \gamma \neq 0, -1, -2, \dots;$

siendo  $[a]_n$  el símbolo de Pochamer o coeficiente hipergeométrico, tal que

$$[a]_n = a \cdot (a+1) \cdot \dots \cdot (a+n-1) \quad \text{y} \quad [a]_0 = 1$$

La Función de Cuantía  $f_r$  Solución del Sistema viene dada para  $r = 0, 1, 2, \dots$  por:

$$f_r = f_0 \cdot \frac{[\mathbf{a}]_r \cdot [\mathbf{b}]_r \cdot l^r}{[\mathbf{g}]_r \cdot r!} \quad \text{con} \quad f_0 = \frac{1}{{}_2F_1(\mathbf{a}, \mathbf{b}, \mathbf{g}, l)}$$

Análogamente se tiene, que la Función Generatriz de Probabilidad viene dada por:

$$g(t) = \frac{{}_2F_1[\mathbf{a}, \mathbf{b}, \mathbf{g}, lt]}{{}_2F_1[\mathbf{a}, \mathbf{b}, \mathbf{g}, l]}$$

Un caso particularmente interesante para resolver el problema que nos ocupa, se obtiene haciendo  $\lambda = 1$  obteniéndose entonces una Subfamilia de la anterior Familia de Distribuciones Discretas Hipergeométricas, conocida como **Subfamilia de ORD**.

Para la Subfamilia de Ord la Función de Cuantía será:

$$f_r = f_0 \cdot \frac{[a]_r \cdot [b]_r}{[g]_r \cdot r!} \quad \text{con} \quad f_0 = \frac{1}{{}_2F_1(a, b, g, 1)}$$

y la Función Generatriz de Probabilidad viene dada por:

$$g(t) = \frac{{}_2F_1[a, b, g, t]}{{}_2F_1[a, b, g, 1]}$$

Es útil para el estudio de esta Subfamilia recordar que:

$${}_2F_1[a, b, g, 1] = \frac{\Gamma(g) \cdot \Gamma(g - a - b)}{\Gamma(g - a) \cdot \Gamma(g - b)}$$

y

$${}_2F_1[-n, b, g, 1] = \sum_{r=0}^n \frac{[-n]_r \cdot [b]_r}{[g]_r \cdot r!} = \frac{[g - b]_n}{[g]_n}$$

## II.- GENERACIÓN.-

**A continuación vamos a generar la conocida Distribución de Polya a partir de la Ecuación (\*) como un caso particular de la Subfamilia de ORD.**

Vamos a sustituir en los polinomios  $G$  y  $L$  (\*\*)

$$a = -n \quad b = \frac{N_1}{c} \quad g = 1 - n - \frac{N_2}{c} \quad \text{con} \quad N = N_1 + N_2$$

es inmediato que tendrá rango finito porque  $\forall r > n \quad f_r = 0$  luego

$${}_2F_1(a, b, g, 1) = {}_2F_1\left(-n, \frac{N_1}{c}, 1 - n - \frac{N_2}{c}, 1\right) =$$

$$= \sum_{r=0}^n \frac{[-n]_r \cdot \left[ \frac{N_1}{c} \right]_r}{\left[ 1 - n - \frac{N_2}{c} \right]_r \cdot r!} = \frac{\left[ 1 - n - \frac{N_2}{c} - \frac{N_1}{c} \right]_n}{\left[ 1 - n - \frac{N_2}{c} \right]_n} = \frac{\left[ 1 - n - \frac{N}{c} \right]_n}{\left[ 1 - n - \frac{N_2}{c} \right]_n}$$

pero como  $f_0 = \frac{1}{{}_2F_1(\mathbf{a}, \mathbf{b}, \mathbf{g}, 1)}$  entonces  $f_0 = \frac{\left[ 1 - n - \frac{N_2}{c} \right]_n}{\left[ 1 - n - \frac{N}{c} \right]_n}$

y por tanto la Función de Cuantía se podrá expresar como:

$$\begin{aligned} f_r &= f_0 \cdot \frac{[\mathbf{a}]_r \cdot [\mathbf{b}]_r}{[\mathbf{g}]_r \cdot r!} = \frac{\left[ 1 - n - \frac{N_2}{c} \right]_n}{\left[ 1 - n - \frac{N}{c} \right]_n} \cdot \frac{[-n]_r \cdot \left[ \frac{N_1}{c} \right]_r}{\left[ 1 - n - \frac{N_2}{c} \right]_r \cdot r!} = \\ &= \binom{n}{r} \frac{(-1)^r \left[ 1 - n - \frac{N_2}{c} + r \right]_{n-r} \cdot \left[ \frac{N_1}{c} \right]_r}{\left[ 1 - n - \frac{N}{c} \right]_n} = \\ &= \binom{n}{r} \frac{(-1)^r \left( \prod_{k=0}^{n-r-1} 1 + r - n - \frac{N_2}{c} + k \right) \left( \prod_{j=0}^{r-1} \frac{N_1}{c} + j \right)}{\left( \prod_{i=0}^{n-1} 1 - n - \frac{N}{c} + i \right)} = \end{aligned}$$

$$\begin{aligned}
&= \binom{n}{r} \frac{(-1)^r (-1)^{n-r} \left( \prod_{k=0}^{n-r-1} \frac{N_2}{c} + n - r - 1 - k \right) \left( \prod_{j=0}^{r-1} \frac{N_1}{c} + j \right)}{(-1)^n \left( \prod_{i=0}^{n-1} \frac{N}{c} + n - 1 - i \right)} = \\
&= \binom{n}{r} \frac{\left( \prod_{k=0}^{n-r-1} \frac{N_2}{c} + k \right) \left( \prod_{j=0}^{r-1} \frac{N_1}{c} + j \right)}{\left( \prod_{i=0}^{n-1} \frac{N}{c} + i \right)} = \\
&= \binom{n}{r} \frac{\left( \prod_{k=0}^{n-r-1} N_2 + c.k \right) \left( \prod_{j=0}^{r-1} N_1 + c.j \right)}{c^{n-r} \cdot c^r} = \\
&= \binom{n}{r} \frac{\left( \prod_{k=0}^{n-r-1} N_2 + c.k \right) \left( \prod_{j=0}^{r-1} N_1 + c.j \right)}{\underbrace{\left( \prod_{i=0}^{n-1} N + c.i \right)}_{c^n}} = \\
&= \binom{n}{r} \frac{\left( \prod_{k=0}^{n-r-1} N_2 + c.k \right) \left( \prod_{j=0}^{r-1} N_1 + c.j \right)}{\left( \prod_{i=0}^{n-1} N + c.i \right)} = \\
&= \binom{n}{r} \frac{N_1.(N_1 + c). \dots .(N_1 + (r-1)c). N_2.(N_2 + c). \dots .(N_2 + (n-r-1)c)}{N.(N + c). \dots .(N + (n-1)c)} =
\end{aligned}$$

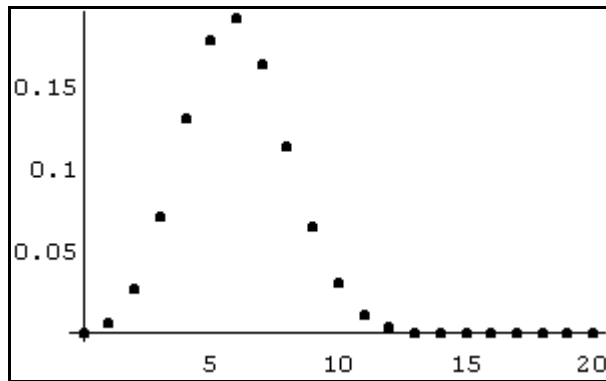
que como se puede apreciar fácilmente es la Función de Cuantía de una variable aleatoria de Polya con composición inicial  $N_1$  y  $N_2$ , con  $N = N_1 + N_2$ , y siendo  $n$  las extracciones realizadas,  $c$  la cantidad a reponer del mismo tipo que el extraído y  $r$  la cantidad que queremos que sea del tipo estudiado.

### III.- GRÁFICAS.-

Si hacemos  $c = 0$  obtendremos la **Distribución Binomial** siendo  $p = \frac{N_1}{N}$ .

$$f_r = \binom{n}{r} \frac{N_1^r \cdot N_2^{n-r}}{N^n} = \binom{n}{r} p^r \cdot (1-p)^{n-r}$$

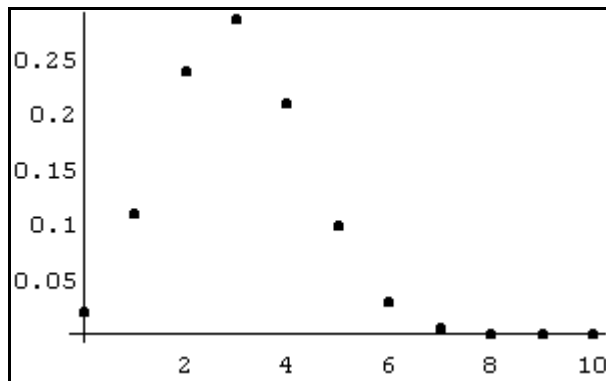
Dibujemos su gráfica para  $p = 0.3$  y  $n = 20$



Si hacemos  $c = -1$  obtendremos la **Hipergeométrica** de parámetros  $N$ ,  $N_1$ ,  $n$

$$f_r = \binom{n}{r} \frac{V_{N_1}^r \cdot V_{N_2}^{n-r}}{V_N^n} = \frac{n!}{r! \cdot (n-r)!} \cdot \frac{N_1!}{(N_1-r)!} \cdot \frac{N_2!}{(N_2-n+r)!} \cdot \frac{N!}{(N-n)!} = \frac{\binom{N_1}{r} \binom{N-N_1}{n-r}}{\binom{N}{n}}$$

Dibujemos su gráfica para  $N = 80$ ,  $N_1 = 24$  y  $n = 10$ .

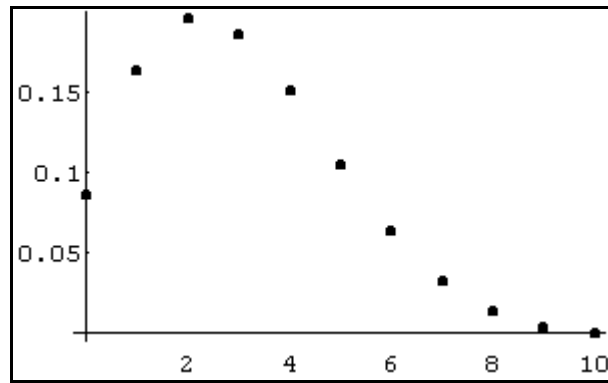


Si hacemos  $c = 1$  obtendremos la **Distribución Hipergeométrica Negativa** de parámetros  $N, N_1, n$ .

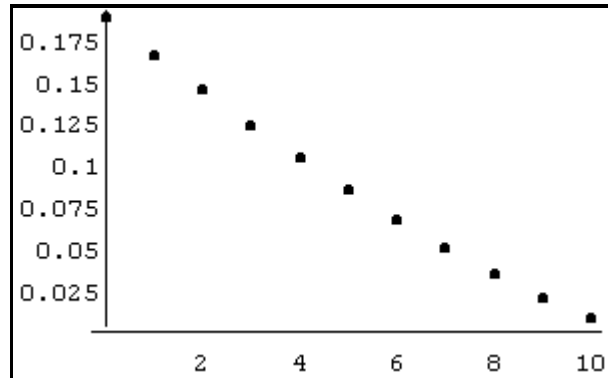
$$f_r = \binom{n}{r} \frac{V_{N_1+r-1}^r \cdot V_{N_2+n-r-1}^{n-r}}{V_{N+n-1}^n} = \frac{n!}{r! \cdot (n-r)!} \cdot \frac{\frac{(N_1+r-1)!}{(N_1-1)!} \cdot \frac{(N_2+n-r-1)!}{(N_2-1)!}}{\frac{(N+n-1)!}{(N-1)!}};$$

$$f_r = \frac{\binom{N_1+r-1}{r} \binom{N-N_1+n-r-1}{n-r}}{\binom{N+n-1}{n}}$$

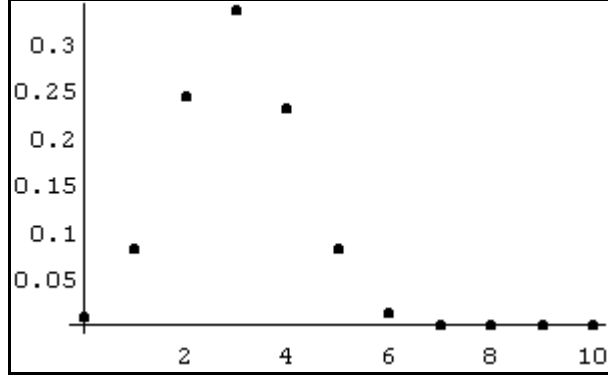
Dibujemos su gráfica para  $N = 10, N_1 = 3$  y  $n = 10$ .



Dibujemos una gráfica general de la **Distribución de Polya** para  $c > 0$ , con los siguientes valores de los parámetros:  $c = 3, N = 10, N_1 = 3$  y  $n = 10$ .



Finalmente, dibujemos una gráfica general de la **Distribución de Polya** para  $c < 0$ , con los siguientes valores de los parámetros:  $c = -3$ ,  $N = 80$ ,  $N_1 = 24$  y  $n = 10$ .



#### IV.- MOMENTOS.-

A partir de la Función Generatriz de Probabilidad obtendremos los Momentos más importantes. Dada

$$g(t) = \frac{{}_2F_1[a, b, g, t]}{{}_2F_1[a, b, g, 1]} = \frac{\sum_{r=0}^{\infty} \frac{[a]_r \cdot [b]_r}{[g]_r \cdot r!} \cdot t^r}{{}_2F_1[a, b, g, 1]}$$

tendremos que

$$g'(t) = \frac{\sum_{r=1}^{\infty} \frac{[a]_r \cdot [b]_r}{[g]_r \cdot (r-1)!} \cdot t^{r-1}}{{}_2F_1[a, b, g, 1]} = \frac{\sum_{r=0}^{\infty} \frac{[a]_{r+1} \cdot [b]_{r+1}}{[g]_{r+1} \cdot r!} \cdot t^r}{\frac{a \cdot b}{g} \sum_{r=0}^{\infty} \frac{[a+1]_r \cdot [b+1]_r}{[g+1]_r \cdot r!} \cdot t^r} = \frac{\sum_{r=0}^{\infty} \frac{[a+1]_r \cdot [b+1]_r}{[g+1]_r \cdot r!} \cdot t^r}{{}_2F_1[a, b, g, 1]}$$

de donde

$$g'(1) = \frac{\frac{a \cdot b}{g} \sum_{r=0}^{\infty} \frac{[a+1]_r \cdot [b+1]_r}{[g+1]_r \cdot r!}}{{}_2F_1[a, b, g, 1]} = \frac{\frac{a \cdot b}{g} {}_2F_1[a+1, b+1, g+1, 1]}{{}_2F_1[a, b, g, 1]} =$$

$$= \frac{\frac{a \cdot b}{g} \cdot \frac{\Gamma(g+1) \cdot \Gamma(g-a-b-1)}{\Gamma(g-a) \cdot \Gamma(g-b)}}{\frac{\Gamma(g) \cdot \Gamma(g-a-b)}{\Gamma(g-a) \cdot \Gamma(g-b)}} = \frac{a \cdot b}{g-a-b-1} = \frac{a \cdot b}{d}; \quad d = g-a-b-1$$

análogamente

$$g''(t) = \frac{\frac{a.b}{g} \cdot \sum_{r=1}^{\infty} \frac{[a+1]_r \cdot [b+1]_r}{[g+1]_r \cdot (r-1)!} \cdot t^{r-1}}{{}_2F_1[a, b, g, 1]} = \frac{\frac{a.b}{g} \cdot \sum_{r=0}^{\infty} \frac{[a+1]_{r+1} \cdot [b+1]_{r+1}}{[g+1]_{r+1} \cdot r!} \cdot t^r}{{}_2F_1[a, b, g, 1]} =$$

$$= \frac{\frac{a.b}{g} \cdot \frac{(a+1) \cdot (b+1)}{g+1} \sum_{r=0}^{\infty} \frac{[a+2]_r \cdot [b+2]_r}{[g+2]_r \cdot r!} \cdot t^r}{{}_2F_1[a, b, g, 1]}$$

de donde

$$g''(1) = \frac{\frac{a.b}{g} \cdot \frac{(a+1) \cdot (b+1)}{g+1} \sum_{r=0}^{\infty} \frac{[a+2]_r \cdot [b+2]_r}{[g+2]_r \cdot r!}}{{}_2F_1[a, b, g, 1]} =$$

$$= \frac{\frac{a.b}{g} \cdot \frac{(a+1) \cdot (b+1)}{g+1} {}_2F_1[a+2, b+2, g+2, 1]}{{}_2F_1[a, b, g, 1]} =$$

$$= \frac{a.b \cdot (a+1) \cdot (b+1)}{(g-a-b-1) \cdot (g-a-b-2)} = \frac{a.b \cdot (a+1) \cdot (b+1)}{d(d-1)}$$

Como

$$E[X] = g'(1); \quad E[X^2] = g''(1) + g'(1) \quad \text{Var}[X] = g''(1) + g'(1) - [g'(1)]^2$$

La Esperanza Matemática y la Varianza de las Distribuciones pertenecientes a este Sistema vienen dadas por:

$$E[X] = \frac{a.b}{d} \quad ; \quad \text{Var}(X) = \frac{a.b \cdot (a+d) \cdot (b+d)}{d^2 \cdot (d-1)}$$

con  $d = g - a - b - 1$ , siempre que se verifique que  $\gamma > \alpha + \beta + 3$

Así pues, como estamos ante un caso particular de la subfamilia de Ord, tenemos que su Esperanza y Varianza vendrán dadas por:

$$E[X] = \frac{a \cdot b}{d} = \frac{(-n) \cdot \frac{N_1}{c}}{1 - n - \frac{N_2}{c} - (-n) - \frac{N_1}{c} - 1}; \quad \boxed{E[X] = n \cdot \frac{N_1}{N}}$$

$$Var(X) = \frac{(-n) \cdot \frac{N_1}{c} \cdot \left(1 - n - \frac{N_2}{c} - \frac{N_1}{c} - 1\right) \left(1 - n - \frac{N_2}{c} - (-n) - 1\right)}{\left(1 - n - \frac{N_2}{c} - (-n) - \frac{N_1}{c} - 1\right)^2 \cdot \left(1 - n - \frac{N_2}{c} - (-n) - \frac{N_1}{c} - 1 - 1\right)} =$$

$$= \frac{(-n) \cdot \frac{N_1}{c} \cdot \left(-n - \frac{N}{c}\right) \left(-\frac{N_2}{c}\right)}{\left(-\frac{N}{c}\right)^2 \cdot \left(-\frac{N}{c} - 1\right)}; \quad \boxed{Var(X) = \frac{n \cdot N_1 \cdot N_2 (N + nc)}{N^2 \cdot (N + c)}}$$

que evidentemente coincide con los resultados que conocemos para esta distribución.

### **BIBLIOGRAFÍA.-**

- FAJARDO CALDERA, M.A. (1985). **Generalización de los sistemas de Pearson discretos**. Universidad de Extremadura.
- HERRERÍAS PLEGUEZUELO, R. (1976). **Extensión del sistema de distribuciones discretas de Pearson**. Cuadernos Departamento Estadística Matemática. Serie A. nº 3. Facultad de Ciencias. Granada.
- JOHNSON, N.L. , KOTZ, S. y KEMP, A.W. (1992). **Univariate Discrete Distributions..** Wiley. New York
- ORD, J.M. (1967). **On a system of discrete distributions**. Biometrika nº 54, 649 - 656.
- PATIL, G.P. , KOTZ, S. , ORD, J.K. (1974). **Statistical distributions in scientific work**. Vol. I, II, III. D. Reidel.
- SRIVASTAVA, H.M. (1964). **Hypergeometric functions of three variables**. Ganita 15, 97-108.