Divulgaciones Matemáticas v. 6, No. 1 (1998), 31–36

Momentos no Centrados de la Distribución Hipergeométrica

Uncorrected Moments of the Hypergeometric Distribution

Juan Antonio García Ramos
Departamento de Matemáticas.
Escuela Universitaria de Estudios Empresariales.
Universidad de Cádiz. Porvera 54.
11403 Jerez de la Frontera. Cádiz. España.

Resumen

Se dedican las líneas que siguen al estudio de los momentos no centrados de la gran desconocida entre las distribuciones de probabilidad elementales: la distribución hipergeométrica. En *De Ratiociniis in Ludo Aleae*, Huygens propone cinco problemas de gran importancia en el desarrollo de la teoría de la Probabilidad. De Moivre, en 1711, resuelve el cuarto utilizando lo que hoy conocemos como función de probabilidad hipergeométrica, siendo ésta la primera referencia histórica que sobre la citada distribución aparece en la literatura probabilística (ver [2] y [6]). Los momentos han sido estudiados, principalmente, por Karl Pearson [4] y V. Romanovsky [5], en sucesivos artículos publicados entre 1899 y 1925. En este último año Pearson encuentra la relación entre sus $k+1$ primeros momentos centrados. Siguiendo la idea usada por Pearson, presentamos en este trabajo una expresión similar para los momentos no centrados de la distribución hipergeométrica.
Palabras y frases clave: distribución hipergeométrica, función generatriz de probabilidad, momentos no centrados (momentos respecto al origen), función generatriz de momentos ordinarios.

Abstract

In this paper we deal with the study of the uncorrected moments of the great unknown between the elementary probability

distributions: the hypergeometric distribution. In *De Ratiociniis in Ludo Aleae*, Huygens proposes five problems of great importance in the development of Probability Theory. De Moivre, in 1711, solves the fourth one by using what we know today as hypergeometric probability function, being this the first historical reference to this distribution in the probabilisty literature (see [2] and [6]). The moments have been studied mainly by Karl Pearson [4] and V. Romanovsky [5], in succesive papers published between 1899 and 1925. In this last year Pearson found the relation between its $k+1$ first corrected moments. Following the idea used by Pearson, we present a similar expresion for the uncorrected moments of the hypergeometric distribution.
Key words and phrases: hypergeometric distribution, probability generating function, uncorrected moments (moments about zero), uncorrected moments generating function.

1 Introducción

Aparece la distribución hipergeométrica al considerar una población de N elementos, dividida en dos clases exhaustivas formadas por A y B puntos muestrales, respectivamente. Considerados los elementos de la primera clase como "éxitos" y los de la segunda como "fracasos", se define la variable aleatoria $\mathbf{X}$ como el número de éxitos o puntos de la primera clase obtenidos al realizar n pruebas o extracciones sin reemplazamiento. La función de probabilidad de la variable $\mathbf{X}$ es

$$h_x = \mathbf{P}(\mathbf{X} = x) = \frac{\binom{A}{x}\binom{B}{n-x}}{\binom{A+B}{n}}, \quad \text{para } \max\{0, n-B\} \le x \le \min\{n, A\}$$

Si llamamos N al tamaño de la población, es decir $N = A + B$, y p y q a las probabilidades de éxito y fracaso iniciales, respectivamente $p = A/N$ y $q = B/N$, podemos escribir

$$h_x = \mathbf{P}(\mathbf{X} = x) = \frac{\binom{Np}{x}\binom{Nq}{n-x}}{\binom{N}{n}} = \frac{\binom{n}{x}\binom{N-n}{Np-x}}{\binom{N}{Np}}$$

para $\max\{0, n-B\} \le x \le \min\{n, A\}$.

Sea k un número entero positivo. Se define el *momento no centrado* (o *momento respecto al origen*) de orden k de la variable $\mathbf{X}$, como el valor esperado de $\mathbf{X}^k$. Lo denotaremos

$$\mu'_k(\mathbf{X}) = \mu'_k = \mathbf{E}(\mathbf{X}^k)$$

Se verifica que $\mu'_0 = 1$ y $\mu'_1 = \mathbf{E}(\mathbf{X}) = \mu$. A $\mathbf{E}(e^{t\mathbf{X}})$, para $t \in \mathbb{R}$, se le llama *función generatriz de momentos* de la variable $\mathbf{X}$, y se representa por $\mathbf{M}_{\mathbf{X}}(t)$ cuando exista, es decir cuando sea finita.

Si $\mathbf{M}_{\mathbf{X}}(t)$ existe en un intervalo $|t| < T$, con $T > 0$, se verifica que μ'_k es el coeficiente de $t^k/k!$ en el desarrollo de Taylor de $\mathbf{M}_{\mathbf{X}}(t)$.

$$\mathbf{M}_{\mathbf{X}}(t) = \mathbf{E}(e^{t\mathbf{X}}) = \sum_{k=0}^{\infty} \mu'_k \frac{t^k}{k!}$$

Se conoce como *serie hipergeométrica* o *serie hipergeométrica de Gauss* a la siguiente:

$${}_2F_1(a, b; c; x) = \sum_{i=0}^{\infty} \frac{a_{(i)} b_{(i)}}{c_{(i)}} \frac{c^i}{i!}$$

donde $c \neq 0, -1, -2, \ldots$ y los $m_{(i)}$ son los símbolos de Pochhammer:

$$m_{(i)} = m(m+1)\ldots(m+i-1), \quad m_o = 1.$$

Si a o b son enteros negativos la serie tiene un número finito de términos no nulos.

Es absolutamente convergente para $|x| < 1$ y divergente para $|x| > 1$. Si $|x| = 1$ es:

a) absolutamente convergente si $c - a - b > 0$,

b) condicionalmente convergente si $-1 < c - a - b \leq 0$ y $x = -1$,

c) divergente si $c - a - b \leq -1$.

2 Exposición

La función generatriz de probabilidad de la distribución hipergeométrica es (ver [3]):

$$\begin{aligned} G_n(z) &= \sum_{x=\max\{0,n-Nq\}}^{\min\{n,Np\}} \mathbf{P}(\mathbf{X} = x) z^x \\ &= h_m z^m {}_2F_1(-n+m, -Np+m; Nq-n+1+2m; z), \end{aligned}$$

siendo $m = \max\{0, n - Nq\}$.
Consideremos la expresión de $\mathbf{M_X}(t)$ (que en lo sucesivo denotaremos simplemente como $M(t)$) para la distribución hipergeométrica

$$\begin{aligned} M(t) &= \mathbf{E}(e^{t\mathbf{X}}) = \sum_{x=\max\{0,n-Nq\}}^{\min\{n,Np\}} \mathbf{P}(\mathbf{X} = x)e^{tx} \\ &= h_m e^{tm} {}_2F_1(-n+m, -Np+m; Nq-n+1+2m; e^t), \end{aligned}$$

siendo $m = \max\{0, n - Nq\}$.

La serie ${}_2F_1(a, b; c; x)$ es una solución de la "ecuación de Gauss" (ver [1]):

$$x(1-x)\frac{d^2{}_2F_1(x)}{dx^2} + [c - (a+b+1)x]\frac{d\,{}_2F_1(x)}{dx} - ab\,{}_2F_1(x) = 0$$

Si hacemos

$$x = e^t, \quad \frac{d\,{}_2F_1(x)}{dx} = \frac{dF}{dt}e^{-t}, \quad \frac{d^2{}_2F_1(x)}{dx^2} = \frac{d^2F}{dt^2}e^{-2t} - \frac{dF}{dt}e^{-2t}$$

como $M(t) = h_m e^{tm} F$ se tiene que $F = (1/h_m)e^{-tm}M$ y además

$$F' = \frac{e^{-tm}}{h_m}(M' - mM)$$

$$F'' = \frac{e^{-tm}}{h_m}(M'' - 2mM' + m^2M).$$

Luego, sustituyendo en la ecuación

$$e^t(1-e^t)[\frac{e^{-tm}}{h_m}(M'' - 2mM' + m^2Me^{-2t} - \frac{e^{-tm}}{h_m}(M' - mM)e^{-2t}]$$

$$+[Nq - n + 1 + 2m - (-n + m - Np + m + 1)e^t]\frac{e^{-tm}}{h_m}(M' - mM)e^{-t}$$

$$-(-n+m)(-Np+m)\frac{e^{-tm}}{h_m}M = 0.$$

y simplificando convenientemente, se obtiene

$$(1-e^t)(M'' - 2mM' + m^2M - M' + mM)$$

$+[Nq - n + 1 + 2m - (-n + 2m - Np + 1)e^t](M' - mM) - (n - m)(Np - m)e^t M = 0.$

o incluso

$M''(1 - e^t) + M'[Nq - n + e^t(n + Np)] - M(m^2 + Nqm - nm + nNpe^t) = 0$

pero si tenemos en cuenta que $m^2 + Nqm - nm = m[m - (n - Nq)] = 0$, ya que $m = \max\{0, n - Nq\}$,

$$M''(1 - e^t) + M'[Nq - n + e^t(n + Np)] - MnNpe^t = 0$$

y recordando que

$$\begin{aligned} M(t) &= 1 + \mu_1' \frac{t}{1!} + \mu_2' \frac{t^2}{2!} + \cdots \\ M'(t) &= \mu_1' + \mu_2' \frac{t}{1!} + \mu_3' \frac{t^2}{2!} + \cdots \\ M''(t) &= \mu_2' + \mu_3' \frac{t}{1!} + \mu_4' \frac{t^2}{2!} + \cdots \\ e^t &= 1 + \frac{t}{1!} + \frac{t^2}{2!} + \cdots \end{aligned}$$

resulta

$$(\mu_2' + \mu_3' \frac{t}{1!} + \mu_4' \frac{t^2}{2!} + \cdots)(-\frac{t}{1!} - \frac{t^2}{2!} - \frac{3}{3!} - \cdots)$$

$$+(\mu_1' + \mu_2' \frac{t}{1!} + \mu_3' \frac{t^2}{2!} + \cdots)[Nq - n + (1 + \frac{t}{1!} + \frac{t^2}{2!} + \cdots)(n + Np)]$$

$$-nNp(1 + \mu_1' \frac{t}{1!} + \mu_2' \frac{t^2}{2!} + \cdots)(1 + \frac{t}{1!} + \frac{t^2}{2!} + \cdots) = 0.$$

Igualando ahora a cero el coeficiente de t^k resulta

$$\sum_{i=1}^{k} \binom{k}{i} [\mu_{k+2-i}' - (n + Np)\mu_{k+1-i}' + nNp\mu_{k-i}'] = N\mu_{k+1}' - nNp\mu_k' \qquad (1)$$

y utilizando el operador desplazamiento E, $E^p(\mu_s') = \mu_{s+p}'$, tenemos

$$[(1 + E)^k - E^k][\mu_2' - (n + Np)\mu_1' + nNp\mu_0'] = N\mu_{k+1}' - nNp\mu_k' \qquad (2)$$

resultado válido para $k \leq 0$.

La expresión (1) puede escribirse también en la forma

$$\sum_{i=1}^{k} \binom{k}{i-1} [\mu'_{i+1} - (n+Np)\mu'_i + nNp\mu'_{i-1}] = N\mu'_{k+1} - nNp\mu'_k$$

y de aquí, si convenimos que $\binom{k}{-2} = \binom{k}{-1} = 0$ para todo k, podemos obtener el valor del $(k+1)$-ésimo momento en función de los $k+1$ precedentes:

$$\mu'_{k+1} = \frac{\sum_{i=0}^{k} \mu'_i [\binom{k}{i-2} - (n+Np)\binom{k}{i-1} + nNp\binom{k}{i}]}{N-k} \tag{3}$$

válido para $k \geq 0$.

Para finalizar, si aplicamos los resultados (1), (2) o (3) para los primeros valores de k se obtiene:

$$\begin{aligned}
\mu'_0 &= 1 \\
\mu'_1 &= np \\
\mu'_2 &= \frac{np(nNp - n - Np + N)}{N-1} \\
\mu'_3 &= \frac{np(2n^2+N^2+3nNp+3nN^2p+n^2N^2p^2+2N^2p^2-3n^2Np-3nN^2p^2-3nN-3N^2p)}{(N-1)(N-2)}
\end{aligned}$$

Referencias

[1] Bronson, Richard, *Ecuaciones Diferenciales Modernas*, Mc-Graw-Hill, Mexico, 1985.

[2] Hald, A. *A History of Probability and Statistics and their Applications before 1750*, John Wiley&Sons, New York, 1990.

[3] Ollero Hinojosa, J., Ramos Romero, H. M. *La Distribución Hipergeométrica como Binomial de Poisson*, Trabajos de Estadística vol. 6 No. 1 (1991), C.S.I.I., Sociedad de Estadística e Investigación Operativa, 35–43.

[4] Pearson, K., *On the Moments of the Hypergeometrical Series*, Biometrika, **16** (1924), 157–162.

[5] Romanovsky, V.,*On the Moments of the Hypergeometrical Series*, Biometrika, **17** (1925), 57–60.

[6] Todhunter, I. *A History of the Mathematical Theory of Probability from the time of Pascal to that of Laplace*, Chelsea Pub. Co., New York, 1965.