

**APORTACIONES AL ESTUDIO DE LA DISTRIBUCIÓN DE PÓLYA:
CARACTERIZACIÓN Y PROPIEDADES DE SU FUNCIÓN
DE DISTRIBUCIÓN**

Memoria para optar al grado
de Doctor en Ciencias, Sección Matemáticas,
que presenta el Licenciado
David Almorza Gomar

Director de la Tesis
Prof. Dr. Héctor M. Ramos Romero

Vº Bº

FACULTAD DE CIENCIAS. UNIVERSIDAD DE CÁDIZ

APORTACIONES AL ESTUDIO DE
LA DISTRIBUCIÓN DE PÓLYA:
CARACTERIZACIÓN Y PROPIEDADES
DE SU FUNCIÓN DE DISTRIBUCIÓN

Doy las gracias

A mi amigo y maestro, Héctor Ramos Romero, que me dedica su magisterio y me honra con su amistad

En recuerdo de

Antonino García Rendón, a quien tanto habría alegrado este trabajo.

*Nombre de sueños,
de árboles escondiendo conejos
y juegos de reinas bailando en naipes.*

*Nombre de ilusión,
en mirada infinita de horizontes.*

*Nombre de nena
que registra carruseles
y viaja en carreras entusiastas
con gritos de alegrías.*

*Descubriendo las palabras
para conjugar sueños e ilusiones,
con el nombre sonoro de vocales abiertas,
con el nombre de Alicia.*

De Sara Molina para Alicia Almorza
La Plata, noviembre de 1997

A mis padres, por su ayuda siempre.

A Marisa y a Alicia, por todo el tiempo
que les robo.

A Amelia y a Elena, por estar conmigo.

INDICE

ÍNDICE

Introducción 5

Capítulo I

Introducción a los modelos de urna. La distribución de Pólya

Introducción a los modelos de urna	15
Modelos de urna	16
El modelo de urna de Pólya	18
Modelos derivados del modelo de Pólya	20
Modelo con reemplazamiento	20
Modelo sin reemplazamiento	22
Otros modelos de urna relacionados	24
Modelo de campaña de seguridad	24
Modelo de Ehrenfest	25
Modelo de Naor	25
Modelo de urna generalizado	26
Un nuevo modelo de urnas	27
Una extensión de los modelos de urna	28
Esquema-resumen	30
La distribución de Pólya	32
Casos particulares	34
Distribución inversa de Pólya	35
Casos particulares	35
Familias de distribuciones que incluyen a la distribución de Pólya	36
Sistema de Pearson discreto	36
Distribución binomial generalizada	37
Familias de Kemp	38
La distribución de Markov-Pólya generalizada	40

Momentos	41
Distribución de Pólya	41
Distribución inversa de Pólya	42
Función generatriz de probabilidad	43
Distribución de Pólya	43
Distribución inversa de Pólya	45
Distribución de Pólya multivariante	45

Capítulo II

Caracterizaciones de la distribución de Pólya

Introducción	48
Mixtura de distribuciones	49
Antecedentes: teorema de caracterización de Janardan	50
Mixtura de la distribución de Pólya	51
Teorema de caracterización	55

Capítulo III

Descripción de la distribución de Pólya como binomial generalizada. Aplicaciones

La distribución binomial generalizada	63
La distribución de Pólya como binomial generalizada	65
Algunos resultados y propiedades de la distribución de Pólya sin reemplazamiento derivadas de su descripción como binomial generalizada	67
Localización de la moda y de la mediana	67
Sobre la forma de la distribución	69
Acotaciones para la función de distribución	70
Intervalos de confianza y test de hipótesis	71

Suma y diferencia de distribuciones de Pólya sin reemplazamiento	72
Sobre la suma de distribuciones de Pólya sin reemplazamiento	72
Diferencia de distribuciones de Pólya sin reemplazamiento	73

Capítulo IV

La función de distribución de la distribución de Pólya

La función de distribución	76
Resultados	77

Capítulo V

Una acotación para la función de distribución de la distribución de Pólya

Introducción y definiciones previas	90
Resultados previos	92
Comparación entre las funciones de distribución de las distribuciones de Pólya y binomial de igual valor esperado y tamaño muestral	120
Conclusiones	125

Bibliografía	128
---------------------	-----

INTRODUCCIÓN

INTRODUCCIÓN

El título de esta memoria, **Aportaciones al estudio de la distribución de Pólya: caracterización y propiedades de su función de distribución**, sintetiza los resultados fundamentales que hemos incorporado a lo largo de los cinco capítulos que conforman el trabajo:

- Introducción a los modelos de urna. La distribución de Pólya.
- Caracterizaciones de la distribución de Pólya.
- Descripción de la distribución de Pólya como distribución binomial generalizada. Aplicaciones.
- La función de distribución de la distribución de Pólya.
- Una acotación para la función de distribución de la distribución de Pólya.

El primer capítulo tiene carácter introductorio. La primera parte está dedicada a los modelos de urna. Se parte del modelo de urna de Pólya y se recogen los modelos derivados así como la relación con otros modelos. Esta parte concluye con un esquema-resumen que hemos elaborado para esta memoria y que contiene veintinueve casos particulares.

La segunda parte de este primer capítulo sirve para introducir la distribución de Pólya a partir del modelo de urna de Pólya. Sobre la correcta denominación de esta distribución existe una cierta polémica, referida en este capítulo, que ha originado que autores distintos utilicen distintos nombres cuando se refieren a ella. A margen de estas discusiones y sin realizar consideraciones sobre lo adecuado de su nombre y sólo por utilizar una notación que además de breve es la más usual, en esta memoria se va a emplear la denominación de *Distribución de Pólya*.

El modelo de urna que da origen a la distribución de Pólya se va a plantear de la siguiente manera:

“Consideremos que una urna contiene M bolas blancas y $N-M$ bolas negras (o alternativamente, pN bolas blancas y qN bolas negras, siendo N el total de bolas de la urna, p la proporción de bolas blancas y $q=1-p$). Se extrae sucesivamente una bola de la urna y una vez observado su color es reemplazada junto a c bolas más del mismo color. El proceso se repite n veces.”

La variable aleatoria que indica el número de bolas blancas extraídas se dice que sigue una distribución de Pólya de parámetros N , M , n y c . Dicha distribución será notada $P(N, M, n, c)$ o, alternativamente, $P(N, p, n, c)$.

Del mismo modo que a partir del modelo de urna de Pólya se derivan otros modelos de urna, de la distribución de Pólya se derivan otras distribuciones relacionadas con estos nuevos modelos. Utilizando este paralelismo se introducen las distribuciones derivadas.

La distribución de Pólya se encuentra a su vez incluida como caso particular en otras familias de distribuciones y a ellas se hace referencia a continuación destacando, por el especial interés que tiene en este trabajo, la distribución binomial generalizada o distribución binomial de Poisson.

El capítulo concluye recogiendo algunas propiedades de la distribución de Pólya que hacen referencia a sus momentos y a su función generatriz de probabilidad. Una extensión al caso multivariante establece el punto final de un capítulo que se justifica en la presentación del objeto de estudio.

En el Capítulo II se recogen nuestras primeras aportaciones en esta memoria que se concretan en la obtención de dos teoremas de caracterización de la distribución de Pólya. Queremos destacar que, aunque las propiedades de la distribución de Pólya han sido ampliamente estudiadas, son escasas las caracterizaciones que existen de esta distribución. En la exhaustiva revisión bibliográfica que hemos efectuado sólo hemos encontrado la debida a Janardan (1984). Dicho autor afirma en su trabajo que hasta la fecha se desconoce la existencia de alguna caracterización para la distribución de Pólya.

Nuestras caracterizaciones van a estar referidas a distribuciones de Pólya $P(N, M, n, c)$ con parámetro reemplazamiento negativo. Como en Janardan (1984) exigiremos al parámetro c la condición $(-c)(n-1) < \min\{ M ; N-M \}$ que garantiza que el rango de la distribución es $[0, n]$, también llamado rango completo

Ambos teoremas de caracterización van a estar basados en el comportamiento de la distribución de Pólya frente a mixturas. En este sentido, se obtiene en el Teorema 2.1 un resultado previo referente a la mixtura de la distribución de Pólya con parámetro reemplazamiento negativo

con otra distribución de Pólya cualquiera. De este teorema se sigue el Corolario 2.2, que coincide con un resultado de Ramos (1989) y que aquí se obtiene como un caso particular del nuestro.

La primera caracterización obtenida, Teorema 2.3, se basa en la mixtura de la distribución de Pólya con una distribución binomial. Si se particulariza el resultado para el caso en que $c = -1$, se obtiene el teorema de caracterización de la distribución hipergeométrica dado en Skibinsky (1970).

La segunda caracterización obtenida, Teorema 2.4, se basa en la mixtura con otra distribución de Pólya. De nuevo, particularizando el resultado para $c = -1$ en el Corolario 2.5 se obtiene el teorema de caracterización para la distribución hipergeométrica dado en Ramos y Ollero (1989).

En el Capítulo III se hace referencia a la distribución de Pólya como distribución binomial generalizada de la que ya se ha hablado en el Capítulo I. Ollero y Ramos (1995) describen una subfamilia del sistema de Pearson discreto como distribuciones binomiales generalizadas o binomiales de Poisson y prueban que las distribuciones de Pólya con parámetro reemplazamiento negativo pertenecen a esta subfamilia.

Como consecuencia de la descripción de la distribución de Pólya con parámetro reemplazamiento negativo como distribución binomial generalizada, propiedades asociadas a esta última distribución son aplicadas en el citado trabajo para, en unos casos, obtener resultados acerca de la distribución de Pólya que de otra forma serían difíciles de probar y, en otros casos, escribir de una manera más simple propiedades ya conocidas de esta distribución.

Concretamente,

- se localiza la mediana y la moda,
- se da un resultado sobre la forma de la distribución del que se deriva. como caso particular, la unimodalidad fuerte de la distribución de Pólya,
- se obtiene un resultado que permite la comparación de las funciones de distribución de la distribución de Pólya y de la distribución binomial aproximante en todo punto salvo en los pertenecientes a un intervalo de indeterminación de amplitud unidad. Como consecuencia,
- se prueba la aplicabilidad del intervalo de confianza y test de hipótesis para el parámetro proporción binomial al caso del correspondiente parámetro en la distribución de Pólya.

Respecto del resultado que permite comparar las funciones de distribución de las distribuciones de Pólya $P(N,p,n,c)$, $c < 0$, y binomial $B(n,p)$, queremos destacar aquí que en el Capítulo V, como conclusión del estudio que ahí desarrollamos, vamos a mejorar este resultado consiguiendo que la comparación sea factible en todo punto salvo en un intervalo que está contenido estrictamente en el intervalo de indeterminación obtenido en Ollero y Ramos (1995).

Nuestra aportación en el Capítulo III se concreta en unos resultados sobre la distribución de la suma y diferencia de variables aleatorias independientes que siguen distribuciones de Pólya con parámetro reemplazamiento negativo. Probamos que en ambos casos la distribución resultante es binomial generalizada, si bien en el caso de la diferencia se trata de una distribución desplazada. Estos resultados los obtenemos teniendo en consideración, de una parte, un teorema de caracterización dado por Ollero (1988) para la distribución binomial generalizada. Los resultados obtenidos son presentados en los Corolarios 3.1 y 3.3, respectivamente. Por otra parte, probamos en el Corolario 3.2 que si la suma de dos variables aleatorias independientes y no negativas sigue una distribución de Pólya con parámetro reemplazamiento negativo, ambas variables siguen distribuciones binomiales generalizadas.

El Capítulo IV lo dedicamos al estudio de la función de distribución de la distribución de Pólya. Aquí no vamos a hacer ninguna restricción sobre el parámetro c , por lo que los resultados que obtenemos serán válidos para toda la familia de distribuciones de Pólya $P(N,M,n,c)$.

En el Teorema 4.1, obtenemos el valor de la diferencia entre las funciones de distribución de las distribuciones de Pólya $P(N,M,n,c)$ y $P(N,M+c,n,c)$, llegándose a una expresión sencilla. De este teorema se derivan cuatro corolarios.

El primero de ellos, Corolario 4.2, proporciona la correspondiente expresión de la diferencia de las funciones de distribución de dos hipergeométricas negativas y es consecuencia de la particularización del Teorema 4.1 al caso $c = 1$.

En el Corolario 4.3 se recoge la expresión de la diferencia entre las funciones de distribución de las distribuciones de Pólya $P(N,M-c,n,c)$ y $P(N,M,n,c)$. Los dos siguientes corolarios obtenidos, Corolarios 4.4 y 4.5, hacen referencia a la expresión de las diferencias de las funciones de distribución de variables aleatorias hipergeométricas cuando aumentamos o disminuimos en una unidad el número de bolas blancas de la urna, es decir, el parámetro M .

Obtenemos respectivamente la expresión de la diferencia de las funciones de distribución de las distribuciones hipergeométricas $H(N, M-1, n)$ y $H(N, M, n)$ así como de $H(N, M, n)$ y $H(N, M+1, n)$ con objeto de calcular la segunda diferencia.

Estos resultados que aquí presentamos como particularizaciones al caso $c = -1$ de un resultado más general, fueron obtenidos por Uhlmann (1966) en un estudio acerca de la distribución hipergeométrica, siguiendo una metodología distinta.

Si en el Teorema 4.1 conseguimos una expresión para la primera diferencia, en el Teorema 4.6 desarrollamos la expresión de la segunda diferencia de la función de distribución de la distribución de Pólya. Al igual que antes, no se establecen restricciones sobre el parámetro reemplazamiento. Se observa que la segunda diferencia se anula para $x=p(n-1)$.

De la particularización del Teorema 4.6 al caso $c = -1$, se obtiene el Corolario 4.7 sobre la expresión de la segunda diferencia de la función de distribución hipergeométrica. Este resultado, que aquí es un caso particular, también fue obtenido por Uhlmann (1966) siguiendo una metodología distinta.

En el Teorema 4.8 se comparan las funciones de distribución de dos distribuciones de Pólya $P(N, M, n, c)$ y $P(N, M, n, c')$. No se hacen restricciones sobre el signo del parámetro reemplazamiento, lo que hace posible comparar, de forma inmediata, las funciones de distribución de las distribuciones hipergeométrica negativa, binomial e hipergeométrica en dichos puntos, como queda recogido en el Corolario 4.9.

Todos los resultados obtenidos en este capítulo son originales y van encaminados a ser aplicados en la búsqueda de una acotación para la función de distribución de la distribución de Pólya con parámetro reemplazamiento negativo en términos de la distribución binomial aproximante, objetivo al que dedicamos el Capítulo V.

En efecto, en el Capítulo V nuestro objetivo es llegar a poder comparar la función de distribución de la distribución de Pólya $P(N, p, n, c)$, $c < 0$, con la de la binomial $B(n, p)$. El resultado al que llegamos permitirá la comparación de ambas funciones de distribución en todo \mathbb{R} salvo en un intervalo de amplitud menor que la unidad. Con este resultado se mejora el obtenido por Ollero y Ramos (1995) ya que ahora se prueba que el intervalo de indeterminación está contenido estrictamente en el allí obtenido.

En este capítulo seguiremos la metodología sugerida en Uhlmann (1966) en un estudio acerca de la distribución hipergeométrica. En esta línea, utilizaremos en el desarrollo del resultado dos funciones auxiliares $G_{N,x,n,c}(p)$ y $H_{N,x,n,c}(p)$ relacionadas ambas con la función de distribución $F_{N,p,n,c}(x)$ de la distribución de Pólya. Concretamente,

$$F_{N,p,n,c}(x) = \Pr[X \leq x] = \begin{cases} 0 & ; x < 0 \\ \sum_{m=0}^k \binom{n}{m} \frac{(pN)^{(m,c)}(qN)^{(n-m,c)}}{N^{(n,c)}} & ; k \leq x < k+1 \quad (k:0,1,\dots,n) \\ 1 & ; x \geq n \end{cases}$$

$$G_{N,x,n,c}(p) = \begin{cases} 0 & ; 1 + \frac{c(n-x-1)}{N} < p \leq 1 \\ \sum_{m=0}^x \binom{n}{m} \frac{(pN)^{(m,c)}(qN)^{(n-m,c)}}{N^{(n,c)}} & ; \frac{-cx}{N} \leq p \leq 1 + \frac{c(n-x-1)}{N} \\ 1 & ; 0 \leq p < \frac{-cx}{N} \end{cases}$$

donde

$$A^{(k,c)} = A(A+c)(A+2c)\dots(A+(k-1)c) : A^{(0,c)} = 1$$

y

$$H_{N,x,n,c}(p) = \sum_{m=0}^x \binom{n}{m} \frac{(pN)^{(m,c)}(qN)^{(n-m,c)}}{N^{(n,c)}} \quad ; \quad 0 \leq p \leq 1$$

que coincide con $G_{N,x,n,c}(p)$ en el intervalo $\left[\frac{-cx}{N} ; 1 + \frac{c(n-x-1)}{N} \right]$.

Si consideramos fijos los valores N, n y c , se tiene que, para todo $x:0,1,\dots,n$, y para todo valor de p compatible con el modelo de urna asociado a la distribución de Pólya con parámetro reemplazamiento negativo,

$$G_{N,x,n,c}(p) = H_{N,x,n,c}(p) = F_{N,p,n,c}(x).$$

Obtendremos en primer lugar una serie de propiedades de estas funciones auxiliares que conforman el apartado titulado Resultados Previos y que serán necesarias para nuestro objetivo final. Dichos resultados conducen a conseguir en primer lugar (Teorema 5.16), la comparación de $G_{N,x,n,c}(p)$ y $G_{N-c,x,n,c}(p)$.

Seguidamente, teniendo en cuenta que para $x:0,1,\dots,n$ $\lim_{N \rightarrow \infty} G_{N,x,n,c}(p) = G_{x,n}(p)$ donde,

$$G_{x,n}(p) = \sum_{m=0}^x \binom{n}{m} p^m (1-p)^{n-m} ; 0 \leq p \leq 1 .$$

se obtiene (Teorema 5.17) que, para $0 < x < n-1 < N$,

$$G_{N,x,n,c}(p) - G_{x,n}(p) = \begin{cases} 0 & \text{si } p = 0 \\ > 0 & \text{si } 0 < p \leq \frac{xN}{(n-1)(N-c)} \\ < 0 & \text{si } \frac{x}{n-1} - \frac{c(n-x-1)}{(n-1)(N-c)} \leq p < 1 \\ 0 & \text{si } p = 1 \end{cases}$$

A partir de aquí, sin más que tener en consideración que para todo $x:0,1,\dots,n-1$ y para cada valor de p fijo compatible con el resto de parámetros de la distribución de Pólya,

$$F_{N,p,n,c}(x) = G_{N,x,n,c}(p)$$

y

$$F_{p,n}(x) = G_{x,n}(p)$$

se obtiene (Teorema 5.19) que, para todo $x:0,1,\dots,n-1$ ($n > 1$),

$$F_{N,p,n,c}(x) - F_{p,n}(x) \begin{cases} < 0 & \text{si } x \leq \frac{p(n-1)(N-c)+c(n-1)}{N} \\ > 0 & \text{si } x \geq \frac{p(n-1)(N-c)}{N} \end{cases}$$

donde con $F_{p,n}$ estamos notando la función de distribución de la binomial $B(n,p)$.

A la vista de este resultado se concluye que como máximo existe un único valor de x entero para el que la comparación de las funciones de distribución de la distribución de Pólya con parámetro reemplazamiento negativo y binomial de igual tamaño muestral y valor esperado, queda sin determinar. Concretamente, esto ocurre si x pertenece al intervalo

$$I = \left(\frac{p(n-1)(N-c)+c(n-1)}{N} ; \frac{p(n-1)(N-c)}{N} \right).$$

Finalmente, se comprueba que I está contenido estrictamente en el intervalo $(pn-1, pn)$ siendo este último el intervalo de indeterminación que se obtiene en Ollero y Ramos (1995) por lo que, como avanzamos, conseguimos mejorar dicho resultado.

CAPÍTULO I

INTRODUCCIÓN A LOS MODELOS DE URNA. LA DISTRIBUCIÓN DE PÓLYA

CAPÍTULO I

INTRODUCCIÓN A LOS MODELOS DE URNA. LA DISTRIBUCIÓN DE PÓLYA

Este capítulo tiene carácter introductorio. Se presentan los modelos de urna especificando su utilidad e importancia y describiendo una amplia serie de modelos hasta concretar el modelo de urna de Pólya que da origen a la distribución de Pólya objeto de esta memoria. Concluye el capítulo con una relación de algunas de las propiedades de esta distribución.

INTRODUCCIÓN A LOS MODELOS DE URNA

Los modelos de urna son muy utilizados en probabilidad, entre otras razones, como recoge Berg (1989), por ser fácilmente visualizables, por su flexibilidad y por su adaptabilidad a un amplio rango de situaciones. De hecho, numerosos resultados de teoría de probabilidad para variables aleatorias discretas, pueden reducirse a un modelo de urnas

Alfred Rényi (1970) recuerda el caso de un colega suyo que estuvo enseñando Estadística en Etiopía, cuando los juegos de azar estaban prohibidos. La reducción a modelos de urnas le permitió sustituir estos juegos e introducir la Teoría de Probabilidad a sus alumnos

Pólya (1954) hace la siguiente observación al respecto:

“Cualquier problema en probabilidad se puede hacer comparable a un problema acerca de bolsas que contengan bolas, y cualquier fenómeno aleatorio puede asimilarse, en sus aspectos esenciales, a extracciones sucesivas de bolas de un sistema combinado de bolsas”.

Posteriormente otros autores como Heitele (1975), apoyan este punto de vista. Por último, en el libro de Johnson y Kotz (1977) dedicado exclusivamente a los modelos de urna, se ratifica la importancia de estos modelos en la Teoría de Probabilidad.

MODELOS DE URNA

Un modelo de urnas se construye a partir de un conjunto de urnas que contengan bolas de diferentes colores y un conjunto de reglas que determinen el procedimiento a seguir según el color de la bola extraída en cada urna.

Estas reglas hacen referencia al número de bolas que se añaden o se retiran de la urna indicándose en qué momento se efectúa esta operación, a cuántas bolas y a qué urnas afecta. Estas reglas dan origen a una gran variedad de modelos de urnas.

Dentro de los modelos de urna, van a ocupar un lugar preferente en este trabajo los llamados *Modelos por Contagio*, esto es, modelos donde la ocurrencia de un suceso tiene el efecto de cambiar la probabilidad de las posteriores ocurrencias de ese mismo suceso.

Se atribuye la idea de usar modelos de urnas para describir *efectos secundarios* a Pólya, que describe el proceso en un artículo junto con Eggenberger [Eggenberger y Pólya (1923)].

Pólya y Eggenberger usan la expresión *Chancevermehrung durch Erfolg* para indicar el aumento de la probabilidad en un suceso debido a su ocurrencia anterior. Pólya sugiere el término de *contagio*, haciendo referencia a la similitud con las enfermedades contagiosas, donde cada ocurrencia incrementa la posibilidad de ocurrencias posteriores.

En Feller (1993) se generaliza el concepto y lo define *Efecto Secundario (aftereffect)*, proponiéndose el siguiente paralelismo para describirlo:

“Consideremos una planta industrial expuesta a accidentes. Se puede pensar en la ocurrencia de un accidente como resultado de un juego de azar sobrehumano: el destino tiene almacenada una urna que contiene bolas negras (rojas en el texto original) y blancas; a intervalos de tiempo regulares,

se extrae al azar una bola; si es negra, representa un accidente. Si la posibilidad de un accidente permanece constante en el tiempo, la composición de la urna es siempre la misma. Pero se piensa que cada accidente tiene un *efecto secundario*, con el cual se incrementa o decrece la probabilidad de nuevos accidentes”.

Este capítulo se va a iniciar a partir del siguiente Modelo de Urna:

“Una urna contiene N bolas, a blancas y b negras; se extrae al azar una bola, se reemplaza y se añaden c bolas del mismo color y d bolas del color contrario. Se hace una nueva extracción aleatoria de la urna (que ahora contiene $a-b-c+d$ bolas) y se repite el mismo procedimiento sucesivamente”

Se utilizan, de esta manera, cuatro parámetros para describir el modelo: a , b , c y d . El número de bolas que la urna contiene inicialmente se deduce a partir de los dos primeros, $N=a+b$.

El procedimiento que lleva al cese del experimento puede proporcionar un nuevo parámetro, cuando se establece un número máximo de repeticiones n . En cualquier caso, debe entenderse que se trata de una regla más del modelo, precisamente la que establece el final del proceso, y que también sirve para diferenciar los modelos de urna entre sí.

Cuando se establece que el número de repeticiones del experimento sea una cantidad n , el modelo se denomina *Modelo a - Urna de Bernard Friedman*, en honor a quien lo propuso, Friedman (1949), y se trata de un modelo general a partir del cual son desarrollados modelos posteriores.

Este valor n determina el número máximo de repeticiones. El experimento puede pararse antes de llegar a esta cantidad si la composición de la urna da origen a un experimento determinístico.

Se tiene así un primer modelo que viene definido por cinco parámetros, y que en general se va a notar en lo sucesivo:

$$(a, b, c, d, n)$$

A este modelo se referirá como *Modelo Directo*.

Otro procedimiento de parada puede plantearse si en lugar de imponer un número de repeticiones del experimento, se establece parar el experimento después de haber obtenido un número k de bolas de un determinado color. En este caso, el modelo que se plantea también tiene cinco parámetros:

$$[a, b, c, d, k],$$

y se va a denominar *Modelo Inverso*, para distinguirlo del anterior.

EL MODELO DE URNA DE PÓLYA

Se va a definir el Modelo de Urna de Pólya como una particularización del Modelo de Urna de Bernard Friedman definido en el apartado anterior. Aunque, como se ha dicho, la idea de usar modelos de urnas se atribuye a Pólya, en el artículo que da origen al tratamiento de la distribución de Pólya [Eggenberger y Pólya (1923)], se utiliza un modelo más reducido que tiene un parámetro menos, precisamente el parámetro d . En este sentido, la única restricción que se impone es considerar $d = 0$ en el modelo de Bernard Friedman.

La notación que se utilizará en este modelo, para mantener la consonancia con el anterior, será:

$$(a, b, c, 0, n)$$

La distribución a que conduce este modelo es conocida como *Distribución de Pólya* o *Distribución de Pólya-Eggenberger*. Sobre la correcta denominación de esta distribución existe una cierta polémica que ha dado lugar a distintos nombres para la misma distribución. Autores como Hald (1960), Bosch (1963), Coleman (1964), Patil y Joshi (1968) y Feller (1993), entre otros, utilizan la denominación de *Distribución de Pólya*. Johnson y Kotz (1969) y Johnson, Kotz y Kemp (1992) utilizan el nombre de *Distribución de Pólya-Eggenberger*. Kryukov (1973) la llama *Distribución de Markov*. Janardan y Schaeffer (1977) y Janardan (1984) mantienen que la denominación correcta es la sugerida por Stancu (1968) de *Distribución de Markov-Pólya*, ya que fue inicialmente propuesta por Markov (1917). Otros autores la denominan *Distribución de Skellam*, en referencia al trabajo de Skellam (1948). La polémica se traduce en referencias, justificaciones y notas parentéticas en algunos de los trabajos que se publican sobre esta distribución.

Al margen de estas discusiones y por utilizar una denominación que además de breve es la más empleada, sin realizar consideraciones sobre lo correcto de su nombre, en este trabajo se va a utilizar la primera denominación: *Distribución de Pólya*.

En relación con el denominado *contagio* en la literatura original, Eggenberger y Pólya (1923) utilizan para distinguir distintos casos derivados del modelo que proponen, el parámetro denominado *factor de contagio* γ , definido como sigue:

$$\gamma = \frac{c}{N},$$

ya que es el parámetro c el que va a determinar las variaciones sucesivas en la composición de la urna.

Cabe hacer la indicación de que en el caso del Modelo de Urna de Bernard Friedman, el *contagio* vendría determinado por los parámetros c y d . Este último no aparece en el Modelo de Pólya, como ya se ha referido, y por esta razón se obvia en este estudio.

También puede ser utilizado un parámetro p que describe la composición inicial de la urna, la proporción de bolas blancas que hay al principio del experimento

$$p = \frac{a}{N}$$

Si se plantea la misma restricción en el Modelo Inverso especificado anteriormente, se obtiene:

$$[a, b, c, 0, k]$$

La distribución a que conduce este modelo se denomina *Distribución Inversa de Pólya*, ya que ha sido obtenida a partir del Modelo Inverso, o también *Distribución Negativa de Pólya*. En lo sucesivo se va a utilizar la primera de estas denominaciones: *Distribución Inversa de Pólya*

MODELOS DERIVADOS DEL MODELO DE PÓLYA

En función de los parámetros antes mencionados, γ y p , Eggenberger y Pólya (1928) distinguen los siguientes casos derivados del esquema de urna inicial

(I) En función del parámetro p :

- (i) Sucesos usuales (*événements usuels*): Es el caso en que p se mantiene constante cuando n tiende a infinito
- (ii) Sucesos raros (*événements rares*): Cuando p tiende a cero, pero de forma que el producto np tiende a una constante λ cuando n tiende a infinito.

(II) En función del Factor de Contagio (parámetro γ)

- (i) Sucesos independientes (*événements indépendants*): Cuando $\gamma=0$ (por tanto no hay factor de contagio).
- (ii) Contagio débil (*faible contagion*): Cuando γ tiende a cero pero de forma que el producto $n\gamma$ tiende a una constante a cuando n tiende a infinito
- (iii) Contagio fuerte (*forte contagion*): Cuando γ es una constante estrictamente mayor que cero. En este caso $n\gamma$ tenderá a infinito con n

Se distinguen dos modelos principales que se derivan del modelo de urna de Pólya, y que se obtienen por la simplificación de alguno de sus parámetros (a, b, c, θ, n), de la siguiente forma:

1.- MODELO CON REEMPLAZAMIENTO

Se trata del caso en que el parámetro c toma el valor cero. Es decir, no hay factor de contagio ($\gamma=0$) y cada bola extraída se devuelve a la urna junto con *ninguna otra bola del mismo color*. El esquema, siguiendo la notación usada hasta el momento, sería el siguiente:

$$(a, b, 0, 0, n)$$

La composición de la urna no va a cambiar a lo largo de las n extracciones. El parámetro p va a permanecer, por tanto, constante durante el proceso, siendo su valor siempre

$$p = \frac{a}{N}$$

La probabilidad de obtener una bola blanca no varía, proporcionando extracciones independientes.

En relación con la clasificación clásica, se trataría del caso *événements usuels* por la constancia de p , y del caso *événements indépendants*, ya que el factor de contagio siempre es cero.

Este modelo dará origen, como se verá más adelante, a la distribución binomial.

En general puede considerarse que cuando el parámetro c toma un valor mayor o igual que cero, siempre se añaden bolas a la urna, convirtiéndose en un modelo con reemplazamiento.

En términos del factor de contagio, se diría que cuando su valor fuera mayor o igual que cero, se estaría tratando de un modelo con reemplazamiento.

Una situación especial se tiene cuando el número, n , de repeticiones es 1. Este modelo da lugar a la distribución de Bernoulli.

$$(a, b, 0, 0, 1)$$

Si se estudia este modelo a partir del modelo inverso, se llega a

$$[a, b, 0, 0, k],$$

tratándose también de un modelo con reemplazamiento que va a dar origen a la distribución binomial negativa.

2.- MODELO SIN REEMPLAZAMIENTO

Se trata del caso en que el parámetro c toma el valor -1 , es decir, las bolas extraídas no son devueltas a la urna (o se devuelven a la urna junto con -1 bola del mismo color). El esquema, siguiendo la notación utilizada hasta el momento, es el siguiente:

$$(a, b, -1, 0, n)$$

La composición de la urna va a cambiar durante el proceso, quedando cada vez menos bolas en su interior. El factor de contagio será, por tanto, negativo e igual a

$$\gamma = \frac{-1}{N}$$

Este modelo dará origen a la distribución hipergeométrica.

En general, puede considerarse que cuando el valor del parámetro c es estrictamente negativo, la urna va perdiendo bolas que no se reemplazan, originando, en un sentido más amplio, un modelo sin reemplazamiento que también podríamos denominar modelo con reemplazamiento negativo.

En términos del factor de contagio, se diría que cuando su valor fuera menor que cero, se estaría tratando de un modelo sin reemplazamiento.

Si se estudia este modelo a partir del modelo Inverso, se llega a

$$[a, b, -1, 0, k],$$

tratándose también de un modelo con reemplazamiento que va a dar origen a la distribución hipergeométrica negativa o distribución beta binomial.

Algunos tipos especiales de modelos con reemplazamiento, que se obtienen según el valor que toma el parámetro c , son los siguientes:

a) $c = 1$

En este caso el modelo obtenido tiene la forma

$$(a, b, 1, 0, n),$$

es decir, siempre que se extrae una bola se reemplaza junto con otra bola del mismo color. El factor de contagio es:

$$Y = \frac{1}{N},$$

que es una cantidad positiva.

El proceso sería el contrario al caso que se ha descrito como modelo sin reemplazamiento donde cada vez se extrae una bola de la urna. Por esta razón la distribución a que da origen este modelo tendrá relación con aquella y recibe el nombre de distribución hipergeométrica negativa o distribución beta binomial.

Se llega de esta manera a una distribución que coincide con la que se obtuvo a partir del modelo inverso dado por

$$[a, b, -1, 0, k],$$

como se vio antes.

Dentro del modelo inverso cuando el parámetro c es igual a 1, se tiene el esquema siguiente:

$$[a, b, 1, 0, k],$$

que va a generar la distribución beta Pascal.

Si además se da la situación especial en que el número de bolas blancas es igual a 1 ($a=1$):

$$(1, b, 1, 0, k),$$

se llegará a la distribución geométrica.

$$b) c = a = b$$

En este caso el modelo obtenido tendrá la forma:

$$(c, c, c, 0, n)$$

La urna constará inicialmente de un mismo número c de bolas blancas y negras (el número total de bolas será $N=2c$), y en este caso la probabilidad de extraer una bola blanca coincide con el factor de contagio:

$$p = \frac{c}{2c} = \frac{1}{2} = \gamma$$

Este modelo dará lugar a la distribución uniforme sobre $(0, 1, \dots, n)$

OTROS MODELOS DE URNA RELACIONADOS Y SUS APLICACIONES

A partir del modelo de urna de Friedman expuesto al inicio de este capítulo, y siguiendo la notación utilizada hasta ahora, se han podido describir otros modelos de urna con distintas aplicaciones que se presentan a continuación

1.- MODELO DE CAMPAÑA DE SEGURIDAD

Propuesto por el mismo Friedman como un caso particular de su modelo y relacionándolo con el ejemplo de los accidentes de Feller. El nombre se debe al planteamiento de que una campaña de seguridad es impulsada cuando ocurre un accidente (se extrae una bola negra), mientras que si no ocurre ningún accidente la campaña de seguridad afloja y la probabilidad de un accidente se incrementa.

En el modelo de urna de Friedman este hecho se traduce en que al obtener una bola (negra si hay un accidente o blanca si no lo hay), la probabilidad de que ocurra lo contrario aumenta (si ha ocurrido un accidente se impulsa la campaña de seguridad y cuando no ocurre se relaja), es decir, se incrementa el valor del parámetro d . La forma del modelo es la siguiente

$$(a, b, 0, d, n),$$

con la restricción de que el parámetro d toma siempre valores positivos

Un caso particular de este modelo, cuando $d=1$, ha sido estudiado por Freedman (1965),

$$(a, b, 0, 1, n).$$

2.- MODELO DE EHRENFEST

También llamado Modelo de Intercambio de Calor entre dos cuerpos aislados, debido al origen del problema planteado por Ehrenfest y Ehrenfest (1907) Considera una urna con bolas blancas y negras. En cada extracción se reemplaza la bola obtenida por una bola del color opuesto. En este caso los parámetros del modelo general toman los siguientes valores:

$$(a, b, -1, 1, n).$$

3.- MODELO DE NAOR

Se trata de un modelo que se construye a partir del modelo inverso. Naor (1956) planteó el siguiente modelo de urna con N bolas de las que una es negra (roja en el artículo original) ($b=1$), y el resto blancas ($a=N-1$). La sustitución de una bola blanca (cuando es extraída) por una bola negra continúa hasta que se extrae una bola negra. Naor estudia la variable *número de extracciones requeridas*. Después de $N-1$ extracciones la urna contiene solamente bolas negras, no siendo necesarias más extracciones.

Consiste, por tanto, en un modelo donde se penaliza la extracción de una bola blanca (hasta ahora éxito), y viene expresado por el esquema

$$[N-1, 1, -1, 1, k]$$

A partir de este modelo se obtiene la distribución de Naor. El interés que tuvo esta distribución para Naor (1957) fue su aplicación en la resolución de problemas de reparación de máquinas.

MODELO DE URNA GENERALIZADO

Bagchi y Pal (1985) presentan un modelo de urna más general y con una notación adaptada al desarrollo computacional de árboles. Gouet (1989) hace uso de este modelo y de la notación cuando realiza sus estudios de convergencia.

Se considera una urna con la composición inicial utilizada hasta ahora de N bolas, M de las cuales son blancas. El experimento consiste en extraer una bola y, si resulta ser blanca, se devuelve a la urna junto con α bolas blancas y β bolas negras. Si la bola extraída resulta ser negra, se devuelve a la urna junto con γ bolas blancas y δ bolas negras más. La variable aleatoria de interés en el estudio es el número de bolas blancas en la urna después de n extracciones.

La notación que se utiliza para representar el proceso es la siguiente:

$$A = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$$

Se hacen las siguientes suposiciones sobre los parámetros:

1) $\alpha + \beta = \gamma + \delta = s \geq 1$

Es decir, se añade el mismo número de bolas en la urna en cada etapa. De esta forma, tras la n -ésima extracción en la urna habrá $N + ns$ bolas.

2) $M \geq 1$

3) $\alpha \neq \gamma$

4) $\beta > 0; \gamma > 0$

5) Si $\alpha < 0$, entonces α divide a γ y a M . Si $\delta < 0$, entonces δ divide a β y a $N - M$.

A partir de este modelo se obtienen otros casos particulares entre los que se destacan los siguientes:

a) Cuando: $\alpha = \delta = s \neq 0$ y $\beta = \gamma = 0$, se tiene el esquema de urna de Pólya. En este caso se verifica que el parámetro $c = s$. La notación es

$$A = \begin{pmatrix} c & 0 \\ 0 & c \end{pmatrix}$$

b) Cuando: $\alpha = \delta \neq 0$ y $\beta = \gamma \neq 0$, se tiene el esquema de urna de Friedman. Los parámetros correspondientes vienen dados por: $\alpha = \delta = c$ y $\beta = \gamma = d$. La notación es

$$A = \begin{pmatrix} c & d \\ d & c \end{pmatrix}$$

c) Cuando: $\alpha = \delta = -1$ y $\beta = \gamma = 1$, se tiene el modelo de Ehrenfest

d) Cuando: $\alpha = \gamma = 0$ y $\delta = -\beta > 0$, se tiene un modelo construido por Woodbury (1949), donde plantea que la extracción de una bola blanca (no infectada) no tiene efecto en la composición de la urna, pero que la extracción de una bola negra (infectada), provoca que δ bolas no infectadas se infecten. Es un caso en que el número de bolas de la urna permanece constante, y a largo plazo todas las bolas serán negras

e) Cuando: $\alpha = \delta = -1$ y $\beta = \gamma = s$, se tiene un modelo construido por Wei (1976) para determinar un asignamiento secuencial de tratamientos en una prueba clínica

UN NUEVO MODELO DE URNAS

En forma de un nuevo modelo de urnas Janardan y Schaeffer (1977) presentaron un esquema consistente en tres urnas: U_1 , U_2 y U_3 . La urna U_1 contiene a bolas blancas (verdes en el trabajo original), la urna U_3 contiene b bolas negras (rojas en el trabajo original) y la urna U_2 contiene a bolas blancas y b bolas negras. Se determinan dos enteros positivos N y t y un entero c de tal forma que verifiquen que

$$\min \{ a, b \} + Nt + Nc > 0$$

A continuación se elige un número entero k , $0 \leq k \leq N$. Se realiza una extracción de U_2 , después una extracción de U_1 o de U_3 según el resultado de la extracción anterior, y por último N extracciones más de U_2 bajo las siguientes condiciones:

- i) Se extrae una bola de U_2 y se devuelve a esa urna después de anotar su color.
- ii) Se añaden kt bolas negras a U_1 , $(N-k)t$ bolas blancas a U_1 y kt bolas blancas junto con $(N-k)t$ bolas negras a U_2 .
- iii) Si la bola que se extrajo de U_2 era negra, se extrae una de U_1 . Si ésta también es negra, no se realizan más extracciones y no hay posibilidad de éxito.
- iv) Si la bola que se extrajo de U_2 era blanca, se extrae una de U_3 . Si ésta también es blanca, no se realizan más extracciones y no hay posibilidad de éxito.
- v) Si la bola que se extrajo de U_1 era negra y la que se extrajo de U_1 era blanca, o si la bola que se extrajo de U_2 era blanca, y la que se extrajo de U_3 era negra, entonces se realizan N extracciones de U_2 con reemplazamiento como se especifica en vi).
- vi) Después de cada extracción de U_2 la bola extraída se reemplaza junto con c bolas más del mismo color. Antes de la siguiente extracción, las bolas de la urna U_2 se mezclan. Se considera un éxito cuando exactamente k de estas N bolas son blancas.

En el caso particular en que $t=0$, se tiene la misma situación que en el modelo de urna de Pólya.

UNA EXTENSIÓN DE LOS MODELOS DE URNA

El modelo de urna de Bernard Friedman puede extenderse al caso de N bolas de $m+1$ colores distintos, de tal forma que haya M_i bolas del color i ($i=1, \dots, m$) y $N - \sum_{i=1}^m M_i$ bolas del $(m+1)$ -ésimo color.

Si se extrae una bola del color i , esa bola junto con c más del mismo color y d bolas de cada uno de los m colores restantes se devuelven a la urna.

De esta manera se consigue un modelo más amplio que se va a notar con la siguiente estructura:

$$(N, M_1, M_2, \dots, M_m, c, d, n),$$

cuando la condición de parada es realizar un total de n extracciones, es decir, en el caso del modelo directo

Por otra parte, si se plantea el modelo inverso, es decir, se establece como procedimiento de parada el conseguir una configuración de colores determinada que esté formada por k_1 bolas del primer color, k_2 bolas del segundo, y así sucesivamente, se habrá obtenido un modelo análogo al presentado como modelo inverso. La notación será:

$$[M_1, M_2, \dots, M_m, c, d, k_1, k_2, \dots, k_m]$$

o, simplificada:

$$[N, M_1, M_2, \dots, M_m, c, d, k],$$

donde k representa el vector (k_1, k_2, \dots, k_m) .

Estos modelos darán lugar a las distribuciones Multivariantes de Pólya Directa e Inversa respectivamente, si se considera el parámetro $d=0$, obteniéndose, para los distintos valores del parámetro c , las distribuciones multivariantes particularizadas correspondientes. Este modelo, en esta versión multivariante, fue propuesto por Steyn (1957).

Los valores que tome el parámetro c , cuando $d=0$, van a determinar si se trata de un modelo con o sin reemplazamiento, del mismo modo que se ha visto en apartados anteriores ($c \geq 0$ ó $c < 0$).

A partir del modelo directo se van a generar las distribuciones multinomial y Dirichlet-multinomial, para el caso de modelo con reemplazamiento, y la distribución hipergeométrica múltiple cuando el modelo es sin reemplazamiento.

El modelo inverso va a dar lugar a las distribuciones multinomial negativa y Dirichlet-Pascal cuando haya reemplazamiento y a la distribución hipergeométrica negativa múltiple cuando no lo haya.

En uno y otro caso se comprueba que se trata, como ya se indicó, de las generalizaciones al caso multivariante de las correspondientes distribuciones obtenidas a partir de los modelos de urna con dos colores.

Merece una mención especial el estudio de la distribución bivariante de Pólya-Eggenberger desarrollado por Marshall y Olkin (1990), donde se consideran tres colores: rojo, negro y blanco. Siguiendo la notación introducida, se tratará del modelo:

$$(M_1, M_2, M_3, c, 0, n).$$

En el esquema-resumen que sigue, se incluyen todos los casos presentados en este capítulo.

ESQUEMA-RESUMEN

El siguiente esquema-resumen de los modelos planteados hasta el momento ha sido construido especialmente para este capítulo.

I).- (a, b, c, d, n) : Modelo de Urna de Bernard Friedman.

*** Si $d=0$:**

$(a, b, c, 0, n)$: Modelo de Urna de Pólya Directo

$(a, b, c > 0, 0, n)$: Modelo Con Reemplazamiento

$(a, b, 0, 0, n)$: Caso especial (distribución binomial).

$(a, b, 0, 0, 1)$: Caso especial (distribución de Bernoulli).

$(a, b, 1, 0, n)$: Caso especial (distribución hipergeométrica negativa).

$(a=c, b=c, c, 0, n)$: Caso especial (distribución uniforme).

$(a, b, c < 0, 0, n)$: Modelo Sin Reemplazamiento.

$(a, b, -1, 0, n)$: Caso especial (distribución hipergeométrica).

* Si $c=0$:

$(a, b, 0, d, n)$: Modelo de Campaña de Seguridad.

$(a, b, 0, 1, n)$: Modelo teórico de Freedman.

* Si $c \neq 0$; $d \neq 0$:

$(a, b, -1, 1, n)$: Modelo de Ehrenfest.

II.- $[a, b, c, d, k]$: Modelo Inverso.* Si $d=0$:

$[a, b, c, 0, k]$: Modelo de Urna de Pólya Inverso.

$[a, b, c \geq 0, 0, k]$: Modelo Con Reemplazamiento.

$[a, b, 0, 0, k]$: Caso especial (distribución binomial negativa).

$[a, b, 1, 0, k]$: Caso especial (distribución beta Pascal).

$[1, b, 0, 0, k]$: Caso especial (distribución geométrica).

$[a, b, c < 0, 0, k]$: Modelo Sin Reemplazamiento.

$[a, b, -1, 0, k]$: Caso especial (distribución hipergeométrica negativa).

* Si $c \neq 0$; $d \neq 0$:

$[N-1, 1, -1, 1, k]$: Modelo de Naor.

III.- $(M_1, M_2, \dots, M_m, c, d, n)$: Modelo Directo Multivariante.* Si $d=0$:

$(M_1, M_2, \dots, M_m, c \geq 0, 0, n)$: Modelo Con Reemplazamiento.

$(M_1, M_2, \dots, M_m, 0, 0, n)$: Caso especial (distribución multinomial).

$(M_1, M_2, \dots, M_m, 1, 0, n)$: Caso especial (distribución Dirichlet-multinomial).

$(M_1, M_2, \dots, M_m, c < 0, 0, n)$: Modelo Sin Reemplazamiento.

$(M_1, M_2, \dots, M_m, -1, 0, n)$: Caso especial (distribución hipergeométrica múltiple).

IV.- $[M_1, M_2, \dots, M_m, c, d, k]$: Modelo Inverso Multivariante.

* Si $d=0$:

$[M_1, M_2, \dots, M_m, c \geq 0, 0, k]$: Modelo Con Reemplazamiento

$[M_1, M_2, \dots, M_m, 0, 0, k]$: Caso especial (distribución multinomial negativa)

$[M_1, M_2, \dots, M_m, 1, 0, k]$: Caso especial (distribución Dirichlet-Pascal)

$[M_1, M_2, \dots, M_m, c < 0, 0, k]$: Modelo Sin Reemplazamiento.

$[M_1, M_2, \dots, M_m, -1, 0, k]$: Caso especial (distribución hipergeométrica negativa múltiple)

LA DISTRIBUCIÓN DE PÓLYA

En lo sucesivo se va a partir del esquema de urna de Pólya. Con el fin de distinguir los aspectos referentes a la distribución de Pólya, de los ya comentados en los modelos de urna, y para utilizar una notación que ha sido comúnmente aceptada en la literatura al respecto, se va a plantear el esquema de urna que da origen a la distribución de Pólya de la manera que se muestra a continuación

Consideremos que una urna contiene inicialmente M bolas blancas y $N-M$ bolas negras (o alternatively, pN bolas blancas y qN bolas negras, siendo N el total de bolas de la urna, $p=M/N$ y $q=1-p$). Se extrae una bola, se anota su color y se introduce de nuevo en la urna junto a c bolas del mismo color que la extraída. El procedimiento se repite n veces. La probabilidad de extraer k bolas blancas en las n extracciones es

$$P[X=k] = \frac{\binom{-\frac{pN}{c}}{k} \binom{-\frac{qN}{c}}{n-k}}{\binom{-\frac{N}{c}}{n}}$$

donde

$$\binom{-A}{B} = (-1)^B \binom{A+B-1}{B}$$

La forma habitual de expresar esta probabilidad, y a la que se hará referencia en este trabajo, es la siguiente:

$$P[X=k] = \binom{n}{k} \frac{A^{(k,c)} M^{(n-k,c)}}{N^{(n,c)}}$$

donde

$$A^{(k,c)} = A(A+c)(A+2c) \dots (A+(k-1)c), \quad A^{(0,c)} = 1$$

Bosch (1963) presenta otras formas distintas de expresar esta misma probabilidad. A continuación se recogen dos de ellas, donde la expresión de la función de probabilidad de la distribución de Pólya es, en términos de las funciones Gamma y Beta respectivamente, siempre tomando $c \neq 0$:

$$P[X=k] = \binom{n}{k} \frac{\Gamma\left(\frac{M}{c}+k\right) \Gamma\left(\frac{N-M}{c}+n-k\right) \Gamma\left(\frac{N}{c}\right)}{\Gamma\left(\frac{M}{c}\right) \Gamma\left(\frac{N-M}{c}\right) \Gamma\left(\frac{N}{c}+n\right)}$$

$$P[X=k] = \binom{n}{k} \frac{\beta\left(\frac{M}{c}+k, \frac{N-M}{c}+n-k\right)}{\beta\left(\frac{M}{c}, \frac{N-M}{c}\right)}$$

El parámetro c puede tomar valores negativos, pero de ser así debe verificar la condición

$$N + c(n-1) > 1.$$

También se utilizan como parámetros los valores

$$p = \frac{M}{N} ; q = \frac{(N-M)}{N}$$

CASOS PARTICULARES

La distribución de Pólya puede definirse, tal como se ha planteado en este capítulo, en función de los cuatro parámetros N , M , n , y c , que determinan el proceso. El parámetro M puede sustituirse por p , ya que ambos parámetros proporcionan la misma información.

En función de los valores que reciba el parámetro c , Pólya (1954) interpretó los siguientes casos particulares:

$c > 0$

Interpreta que el éxito y el fracaso son contagiosos en el sentido de que un éxito o un fracaso aumenta la probabilidad de éxito o de fracaso, respectivamente

$c = 0$

Entonces los sucesos son independientes.

$c < 0$

Interpreta que cada extracción va a originar un revés de la fortuna, en el sentido en que el éxito disminuye a su vez la probabilidad de obtener un nuevo éxito (del mismo modo el fracaso disminuye también la probabilidad de un nuevo fracaso).

En función de los valores particulares de los parámetros se obtienen distribuciones notables como casos particulares de la distribución de Pólya. Así,

$c = -1$: Distribución hipergeométrica

$c = 0$: Distribución binomial

$c = 0$; $n = 1$; $X = 0, 1$: Distribución de Bernoulli

$c = 1$: Distribución beta binomial o distribución hipergeométrica negativa

$c = M = N - M$: Distribución uniforme discreta

DISTRIBUCIÓN INVERSA DE PÓLYA

También puede realizarse el planteamiento que se ha llamado *inverso*, si en lugar de repetir el proceso n veces se establece la regla de parar después de haber obtenido k bolas blancas. De esta forma, en lugar de considerar la probabilidad de obtener k bolas blancas en n extracciones, se consideran cuántas extracciones son necesarias para obtener k bolas blancas. La probabilidad de que se necesiten $k+r$ extracciones, viene indicada por la distribución *Inversa de Pólya*. La probabilidad de que sean necesarias r extracciones para obtener k bolas blancas vale

$$P\{X=r\} = \frac{k}{r} \binom{-\frac{pN}{c}}{k} \binom{-\frac{qN}{c}}{r-k} \binom{-\frac{N}{c}}{r}$$

Con más frecuencia se utiliza la siguiente expresión:

$$P\{X=r\} = \binom{k+r-1}{r} \frac{M^{(k,c)}(N-M)^{(r,c)}}{N^{(k+r,c)}}$$

CASOS PARTICULARES

La distribución Inversa de Pólya puede definirse, tal como se ha planteado en este capítulo, como función de los parámetros N , M , k y c , que determinan el proceso. En este caso también puede sustituirse el parámetro M por el parámetro p .

En función de los valores particulares de los parámetros se obtienen distribuciones notables como casos particulares de la distribución de Pólya. Así, tenemos que:

$c = -1$: Distribución beta binomial o distribución hipergeométrica negativa

$c = 0$: Distribución binomial negativa

$c = 1$: Distribución beta Pascal

$c = 1; M = 1$: Distribución geométrica

FAMILIAS DE DISTRIBUCIONES QUE INCLUYEN A LA DISTRIBUCIÓN DE PÓLYA

1.- SISTEMA DE PEARSON DISCRETO

Pearson (1895) desarrolló este Sistema, también llamado Sistema de Ord, a partir de las distribuciones discretas de probabilidad cuyas funciones de probabilidad verifican la ecuación en diferencias finitas

$$\Delta f_{k-1} = f_k - f_{k-1} = \frac{a-k}{b_0 + b_1 k + b_2 k(k-1)} f_{k-1}, \quad k \in T;$$

donde f_k es el valor de la función de masa de probabilidad en el punto k ; a , b_0 , b_1 y b_2 son parámetros y T es un retículo regular de anchura unidad.

La distribución de Pólya $P(N, M, n, c)$ pertenece al Sistema de Pearson y los valores que la caracterizan dentro del Sistema son

$$a = \frac{(M-c)(n+1)}{N-2c}, \quad b_0 = 0$$

$$b_1 = \frac{N-M+c(n-1)}{N-2c}; \quad b_2 = \frac{-c}{N-2c}$$

Por otra parte, las distribuciones de Pólya con parámetro reemplazamiento negativo pueden concretarse dentro de una subfamilia del Sistema de Pearson. En efecto, Ollero y Ramos (1995) definen la subfamilia \mathcal{P}_H como la constituida por aquellas distribuciones que tienen rango finito $T=[s, S]$, $s \geq 0$, y satisfacen

$$b_0=0 \quad \text{y} \quad b_1/b_2 + 2s \in \mathbb{R}^+.$$

Estas condiciones son verificadas por las distribuciones de Pólya $P(N, M, n, c)$ con $c < 0$.

2.- DISTRIBUCIÓN BINOMIAL GENERALIZADA

La distribución binomial generalizada, o distribución binomial de Poisson, aparece al eliminar del modelo binomial la exigencia de que cada éxito tenga la misma probabilidad en todas las pruebas. De esta manera, dadas n pruebas independientes, cada una con probabilidad de éxito p_i , $i:1, \dots, n$, la distribución binomial generalizada $B(p)$ queda determinada por el vector de parámetros

$$p = (p_1, p_2, \dots, p_n)$$

Ollero y Ramos (1995) demostraron que cualquier distribución perteneciente a la familia ρ_H , y en particular la distribución de Pólya con $c \leq 0$, puede escribirse como una distribución binomial generalizada como afirma el siguiente resultado

Teorema [Ollero y Ramos (1995)]

Toda distribución de Pólya, $P(N, M, n, c)$, con $c < 0$ y rango $[m, s]$, puede ser descrita como una distribución binomial de Poisson cuyo vector de probabilidad $p = (p_1, p_2, \dots, p_s)$ queda determinado por $p_i = \frac{1}{1-z_i}$, $i:1, \dots, s$, siendo z_i ($i:1, \dots, s$) las s raíces de la función generatriz de

probabilidad de la distribución de Pólya.

Obsérvese que si $c = -1$ se obtiene la correspondiente descripción de la distribución hipergeométrica como binomial generalizada.

Dedicaremos el Capítulo III para tratar las propiedades de la distribución de Pólya que van a ser derivadas de la aplicación de la anterior descripción de esta distribución como binomial generalizada o de Poisson.

3.- FAMILIAS DE KEMP

Para el estudio de las distribuciones pertenecientes a las Familias de Kemp, desarrolladas por Kemp (1968 a y b) a partir del trabajo de Kemp y Kemp (1956), se toma como base la función hipergeométrica gaussiana que tiene la forma:

$${}_pF_q[a_1, \dots, a_p; b_1, \dots, b_q; x] = {}_pF_q \left[\begin{matrix} a_1, \dots, a_p \\ b_1, \dots, b_q \end{matrix} \middle| x \right] = \sum_{j=0}^{\infty} \frac{(a_1)_j \dots (a_p)_j}{(b_1)_j \dots (b_q)_j} \frac{x^j}{j!}$$

con la restricción: $b_i \neq 0, -1, -2, \dots; i: 1, \dots, q$;
siendo $(a)_j$ el símbolo de Pochhammer, definido como

$$(a)_j = a(a-1)\dots(a-j+1).$$

El grupo principal de estas distribuciones es el denominado GHRD, Distribuciones Hipergeométricas Generalizadas Recast, que incluye a aquellas distribuciones cuya función generatriz de probabilidad tiene la forma

$$\frac{{}_pF_q[\lambda z + \xi]}{{}_pF_q[\lambda + \xi]}$$

Esta familia contiene a su vez a otras dos: GHPD, Distribuciones Hipergeométricas Generalizadas, y GHFD, Distribuciones Hipergeométricas Generalizadas de Momento Factorial.

La familia de distribuciones GHFD constituye el caso particular en que $\xi = -\lambda$, con lo que la función generatriz de probabilidad tiene la forma

$$G(z) = {}_pF_q[\lambda(z-1)].$$

La familia de distribuciones GHPD aparece cuando se considera $\xi = 0$. Su función generatriz de probabilidad tiene la forma

$$\frac{{}_pF_q[\lambda z]}{{}_pF_q[\lambda]}$$

La distribución de Pólya pertenece a esta última familia, ya que su función generatriz de probabilidad es:

$${}_pF_q[\lambda z] = C' {}_2F_1[-n, M/c; -n-1-(N-M)/c; z]$$

donde C' es una constante.

El estudio realizado por Kemp y Kemp (1956), donde proponen el término de Distribuciones Hipergeométricas Generalizadas y desarrollan este contenido, se basa en los trabajos de Davies (1933 y 1934) y Noack (1950), donde estudian las condiciones bajo las que la expresión

$$P[x=k] = \frac{\binom{a}{k} \binom{b}{n-k}}{\binom{a+b}{n}}$$

proporciona una verdadera distribución de probabilidad.

Se distinguen así cuatro tipos principales de distribuciones, divididas a su vez en subtipos, de acuerdo con el siguiente esquema tomado del libro de Johnson, Kotz y Kemp (1992).

Condiciones para la existencia de los tipos de Distribuciones Hipergeométricas Generalizadas I, II, III y IV

Tipo		Condiciones	Soporte
I	IA(i)	$n-b-1 < 0$; n entero; $0 \leq n-1 < a$	$x=0, 1, \dots, n$
	IA(ii)	$n-b-1 < 0$; a entero; $0 \leq a-1 < n$	$x=0, 1, \dots, a$
	IB	$n-b-1 < 0$; $J < a < J+1$; $J < n < J+1$	$x=0, 1, \dots$
II	IIA	$a < 0 < n$; n entero; $b < 0$; $b \neq -1$	$x=0, 1, \dots, n$
	IIB	$a < 0 < a+b+1$; $J < n < J+1$; $J < n-b-1 < J+1$	$x=0, 1, \dots$
III	IIIA	$n < 0 < a$; a entero; $b < n-a$; $b \neq n-a-1$	$x=0, 1, \dots, a$
	IIIB	$n < 0 < a+b+1$; $J < a < J+1$; $J < n-b-1 < J+1$	$x=0, 1, \dots$
IV	IV	$a < 0$; $n < 0$; $0 < a+b+1$	$x=0, 1, \dots$

Nota: J es un entero no negativo (el mismo para cualquier tipo de distribución).

La distribución de Pólya pertenece al tipo II, subtipo IIA.

LA DISTRIBUCIÓN DE MARKOV-PÓLYA GENERALIZADA

A partir del modelo de urnas que desarrollan Janardan y Schaeffer (1977) y que ha sido presentado en este capítulo, se deriva la distribución de Markov-Pólya generalizada. El nombre se debe a Janardan y Schaeffer que, inmersos en la polémica de la denominación de la distribución de Pólya, reivindican la idea de Markov. De esta forma se le llamará aquí, ya que se trata del único nombre conocido dado a esta distribución.

Utilizando la misma notación dada a los parámetros en el modelo de urna de referencia, la función de masa de probabilidad de la distribución de Markov-Pólya generalizada es

$$P\{X=k\} = \frac{\frac{a}{a+kt} (a+kt)^{(k,c)} \frac{b}{b+(N-k)t} (b+(N-k)t)^{(N-k,c)}}{\frac{a+b}{a+b+Nt} (a+b+Nt)^{(N,c)}}$$

Como se especificó en el modelo de urna, para el caso particular en que $t=0$ se tiene directamente la distribución de Pólya.

Casos especiales de la distribución de Markov-Pólya generalizada son:

* $c=-1$: Distribución mixtura cuasi-hipergeométrica.

$t=0$: Distribución hipergeométrica.

$t=1$: Distribución hipergeométrica negativa.

* $c=0$: Distribución mixtura cuasi-binomial.

$t=0$: Distribución binomial.

* $c=1$: Distribución mixtura cuasi-hipergeométrica negativa.

$t=0$: Distribución hipergeométrica negativa.

MOMENTOS

1.- DISTRIBUCIÓN DE PÓLYA

El primero en realizar un estudio acerca de los momentos correspondientes a la distribución de Pólya fue Jordan (1927). El procedimiento que se expone a continuación se basa en la caracterización dada de la distribución de Pólya dentro de las distribuciones hipergeométricas generalizadas.

Para las distribuciones hipergeométricas generalizadas existe una forma general del r -ésimo momento factorial, en el caso de existir. Para las distribuciones del tipo IIA, y particularizando la forma a los parámetros dados para la distribución de Pólya, la expresión es

$$\mu_{(r)}' = \frac{n! \left(\frac{-M}{c} \right)! \left(\frac{-N}{c} - r \right)!}{(n-r)! \left(\frac{-M}{c} - r \right)! \left(\frac{-N}{c} \right)!}$$

Para las distribuciones del tipo IIA, estos momentos existen siempre, y toman el valor cero cuando $r > n$. Además, son finitos para todo r .

A partir de esta expresión se obtienen la esperanza y la varianza de la distribución de Pólya:

$$E[X] = np$$

$$Var[X] = \mu_2 = \frac{npq(N+nc)}{(N+c)}$$

Del mismo modo, se tiene que el momento de orden 3 es

$$\mu_3 = \frac{\mu_2(q-p)(N+2nc)}{(N+2c)}$$

Bosch (1963) también tiene estos momentos sin utilizar la propiedad de las distribuciones hipergeométricas generalizadas.

2.- DISTRIBUCIÓN INVERSA DE PÓLYA

Patil y Joshi (1968) recogen los momentos media y varianza de la distribución inversa de Pólya:

$$E[X] = \frac{kqN}{(pN-c)}$$

para el caso en que $M > c$, no existiendo la esperanza para otros valores de M .

$$Var[x] = \left(\frac{k(N-M)}{M-c} \right) \left(\frac{N-c}{M-c} \right) \left(\frac{M+(k-1)c}{M-2c} \right)$$

La varianza tiene sentido cuando $M > 2c$, y no existe para otros valores.

FUNCIÓN GENERATRIZ DE PROBABILIDAD

1.- DISTRIBUCIÓN DE PÓLYA

La distribución de Pólya pertenece a la familia de distribuciones GHPD, por lo que su función generatriz de probabilidad se adapta a la expresión general de esta familia, y una vez particularizados los parámetros, será

$$G(z) = \frac{{}_2F_1\left[-n, \frac{M}{c}; -n+1 - \frac{(N-M)}{c}, z\right]}{{}_2F_1\left[-n, \frac{M}{c}; -n+1 - \frac{(N-M)}{c}, 1\right]}$$

Como $-n < 0$, el numerador de la expresión dada se puede escribir, vease Abramowitz y Stegun (1972), como

$${}_2F_1\left[-n, \frac{M}{c}; -n+1 - \frac{(N-M)}{c}, z\right] = \sum_{j=0}^{\infty} \frac{(-n)_j \left(\frac{M}{c}\right)_j z^j}{\left(-n+1 - \frac{(N-M)}{c}\right)_j j!}$$

Para obtener la correspondiente expresión del denominador basta hacer $z=1$

Ollero y Ramos (1995), tomando como base la inclusión de la distribución de Pólya en el sistema de Pearson discreto, establecen la siguiente expresión de la función generatriz de probabilidad en el caso en que $c < 0$:

$$G(z) = f_m z^m {}_2F_1\left(-n+m, \frac{M}{c}+m, -\frac{N-M+c(n-1)}{c}+2m, z\right)$$

donde el rango de la variable viene dado por el intervalo $[m,s]$ y f_m es el valor de la función de masa de probabilidad en el entero menor m .

El interés de esta última expresión radica en que se contempla el caso en que el rango no sea completo, es decir $m > 0$. Por otra parte, en el trabajo antes citado, Ollero y Ramos obtienen asimismo la expresión de la función generatriz de probabilidad para las distribuciones de la familia ρ_H .

Patil y Joshi (1968), indican el valor de la función generatriz de probabilidad en los siguientes términos:

$$G(z) = \frac{\left(\frac{-qN}{c}\right)^{(n)}}{\left(\frac{-N}{c}\right)^{(n)}} {}_2F_1\left(-n, \frac{pN}{c}, \frac{-qN}{c}-n+1, z\right)$$

donde

$$a^{(n)} = 1(a-1)(a-2)\dots(a-n+1) ; a^{(0)} = 1.$$

Acerca de las raíces de la función generatriz de probabilidad de la distribución de Pólya, Ollero y Ramos (1995) probaron el siguiente resultado

Teorema

Si $G_p(z)$ es la función generatriz de probabilidad de una distribución de Pólya $P(N,M,n,c)$, con $c < 0$ y rango $[m,s]$, entonces $G_p(z)$ tiene s raíces reales, de las cuales $s-m$ son simples negativas y las restantes son la raíz $z=0$, con orden de multiplicidad m .

En el Capítulo III se analizará también este resultado

2.- DISTRIBUCIÓN INVERSA DE PÓLYA

Patil y Joshi (1968) dan la siguiente expresión para la función generatriz de probabilidad de la distribución inversa de Pólya:

$$G(z) = \frac{\left(\frac{-qN}{c}\right)^{(k)}}{\left(\frac{-N}{c}\right)^{(k)}} {}_2F_1\left(k, \frac{qN}{c}; \frac{N}{c} - k; z\right)$$

DISTRIBUCIÓN DE PÓLYA MULTIVARIANTE

A partir de las extensiones de los modelos de urna que se han desarrollado, puede obtenerse la *distribución multivariante de Pólya*. Esta distribución fue desarrollada por Steyn (1957) como perteneciente a una familia de distribuciones multivariantes discretas que él mismo definió.

En este caso, utilizando el esquema de urna

$$(c, M_1, M_2, \dots, M_m, c, d, n),$$

se plantea buscar la probabilidad de obtener una configuración de colores $K=(K_1, K_2, \dots, K_m)$ en n extracciones.

La función de probabilidad de la distribución multivariante de Pólya viene dada por

$$P[X=K] = \frac{\prod_{i=1}^{m+1} \binom{-p_i \frac{N}{c}}{K_i}}{\binom{-N}{c, n}}$$

donde $p_i = \frac{M_i}{N}$, y por tanto $p_{m+1} = 1 - \sum_{i=1}^m p_i$

Cuando m vale 1 esta distribución se reduce a la distribución de Pólya

En el trabajo de Patil y Joshi (1968) se recogen los principales momentos correspondientes a la distribución multivariante de Pólya

Vector de medias $\mu_i = np_i$

Matriz de covarianzas

$$\sigma_{ij} = \begin{cases} n \left(\frac{a_i}{N} \right) \left(\frac{N-a_i}{N} \right) \left(\frac{N+nc}{n+c} \right) & , i=j \\ -n \left(\frac{a_i}{N} \right) \left(\frac{a_j}{N} \right) \left(\frac{N+nc}{n+c} \right) & , i \neq j \end{cases}$$

Asimismo se recoge en el citado trabajo la correspondiente función generatriz de probabilidad cuyo valor es

$$G(z_1, z_2, \dots, z_m) = \frac{\left(\frac{I_{m+1}N}{c} + n - 1 \right)^{(n)}}{\left(\frac{N}{c} + n - 1 \right)^{(n)}} \cdot A$$

siendo

$$A = \left[{}_{m+1}F_1 \left(-n, \frac{p_1 N}{c}, \frac{p_2 N}{c}, \dots, \frac{p_m N}{c}, -\frac{p_{m+1} N}{c} ; -n+1, z_1, z_2, \dots, z_m \right) \right]$$

CAPÍTULO II

CARACTERIZACIONES DE LA DISTRIBUCIÓN DE PÓLYA

CAPÍTULO II

CARACTERIZACIONES DE LA DISTRIBUCIÓN DE PÓLYA

En este capítulo se presentan dos caracterizaciones de la distribución de Pólya con parámetro reemplazamiento negativo, basadas en mixturas con la distribución binomial y la distribución de Pólya respectivamente. El único antecedente de caracterización de la distribución de Pólya que hemos encontrado corresponde a Janardan (1984) donde afirma que hasta esa fecha se desconoce la existencia de alguna caracterización para la distribución de Pólya.

INTRODUCCIÓN

Aunque las propiedades de la distribución de Pólya han sido ampliamente estudiadas, son escasas las caracterizaciones que existen de esta distribución. Janardan (1984) obtiene una caracterización basada en el comportamiento de la distribución de Pólya frente a mixturas respecto del parámetro n cuando éste sigue una distribución binomial negativa.

El propósito de este capítulo es considerar distribuciones de Pólya $P(N, M, n, c)$ con parámetro reemplazamiento o de contagio c negativo, y obtener caracterizaciones de esta familia basándonos en su comportamiento frente a mixturas respecto del parámetro m ($m = M/c$) cuando éste sigue una distribución de Pólya cualquiera.

MIXTURA DE DISTRIBUCIONES

Definición

Dada una familia de funciones de distribución $F_j(x_1, x_2, \dots, x_n)$, con $j = -1, 0, 1, 2, \dots$, y una sucesión de parámetros a_j tales que $\sum_{j=-\infty}^{\infty} a_j = 1$ con $a_j > 0, \forall j$, se llama *composición o mixtura* de F_j respecto de a_j a la distribución obtenida como

$$F(x_1, x_2, \dots, x_n) = \sum_{j=-\infty}^{\infty} a_j F_j(x_1, x_2, \dots, x_n)$$

En el caso particular de que las funciones $F_j(x_1, x_2, \dots, x_n)$ son de la forma $F(x; \theta_j)$, y los a_j representan las probabilidades de la distribución del parámetro θ

En este capítulo van a tratarse casos de mixtura de la distribución de Pólya $P(N, M, n, c)$ respecto del parámetro m ($m = M/c$). En primer lugar consideraremos que m es una variable aleatoria binomial y ello dará al Teorema 2.3 de caracterización de la distribución de Pólya. Dicho teorema será necesario como paso previo para abordar el resultado principal de este capítulo, Teorema 2.4, en el que obtenemos una caracterización más general de esa distribución atendiendo al comportamiento de la misma ante una mixtura respecto de m cuando m sigue una distribución de Pólya cualquiera.

La notación que se va a utilizar para la mixtura respecto del parámetro m será

$$P(N, M, n, c) \hat{=} {}_m F,$$

donde F representa la distribución que sigue el parámetro m

ANTECEDENTES: TEOREMA DE CARACTERIZACIÓN DE JANARDAN

Janardan (1984) obtiene la siguiente caracterización de la distribución de Pólya $P(N, M, n, c)$ basada en su comportamiento frente a mixturas respecto del parámetro n cuando este parámetro sigue una distribución binomial negativa.

Teorema

Consideremos la familia de distribuciones $f(k, n)$ indexadas por el parámetro $n: 0, 1, 2, \dots$, definida cada una sobre un subconjunto de $\{0, 1, \dots, n\}$, e independientes de θ . Si n sigue una distribución binomial negativa, $BN(n; N/c, \theta)$, entonces la mixtura resultante con $f(j, n)$ es una distribución binomial negativa $BN(k; M/c, \theta)$ si y sólo si la familia de distribuciones $f(k, n)$ es la familia de distribuciones de Pólya, $P(N, M, n, c)$.

A partir de este teorema Janardan obtuvo el siguiente corolario.

Corolario

Las variables aleatorias independientes X e Y siguen distribuciones $BN(x; M/c, \theta)$ y $BN(y; (N-M)/c, \theta)$ respectivamente si y sólo si la distribución condicionada de X dado $X+Y=n$ sigue una distribución de Pólya $P(N, M, n, c)$.

Ramos y Ollero (1989) obtuvieron caracterizaciones de la distribución hipergeométrica basada en su comportamiento frente a mixturas con otras distribuciones. Una de las mixturas estudiada fue con la distribución de Pólya, derivándose de ella el siguiente teorema de caracterización.

Teorema

Sea $\{F_M\}$, $M=0, 1, \dots, N$, una familia de funciones de distribución que asignan probabilidades no nulas sólo a los enteros $\{0, 1, \dots, n\}$. Condición necesaria y suficiente para que esta familia sea la formada por la familia de distribuciones hipergeométricas $H(N, M, n)$, es que se verifique:

- 1) Su composición cuando M sigue una distribución de Pólya $P(N', M', N, c)$, es la distribución de Pólya $P(N', M', n, c)$
- 2) Las distribuciones F_M no dependen del valor N' .

En este capítulo vamos a considerar distribuciones de Pólya $P(N, M, n, c)$ con parámetro de reemplazamiento c negativo, y vamos a obtener una caracterización de esta familia basada en su comportamiento frente a mixturas respecto del parámetro m ($m = M/c$) cuando dicho parámetro sigue una distribución de Pólya cualquiera. En Ramos (1989) se hace referencia a la posibilidad de establecer el teorema de caracterización de la distribución de Polya que aquí precisamente se ha desarrollado.

MIXTURA DE LA DISTRIBUCIÓN DE PÓLYA

Teorema 2.1

Si $c < 0$, con $N = n/c$ y $M = m/c$, se verifica

$$P(N, M, s, c) \wedge m P(N', M', n, c') = P(N', M', s, c')$$

Demostración

Por las restricciones impuestas, la expresión anterior puede expresarse como

$$P(n|c, m|c', s, c) \wedge m P(N', M', n, c') = P(N', M', s, c')$$

La función de masa de probabilidad de la distribución resultante de la mixtura será

$$P\{\xi = k\} = \sum_j p(k, n|c', j|c', s, c) \cdot p(j, N', M', n, c')$$

Utilizando la notación de la función de probabilidad de la distribución de Polya, la función de masa de probabilidad de la distribución resultante de la mixtura también puede expresarse de la forma

$$P\{\xi = k\} = \sum_j \binom{s}{k} \frac{(j|c')^{(k, c')} ((n-j)|c')^{(s-k, c')}}{(n|c')^{(s, c')}} \binom{n}{j} \frac{M'^{(j, c')} (N' - M')^{(n-j, c')}}{N'^{(n, c')}}.$$

Asimismo, teniendo en cuenta que

$$\begin{aligned} (A|c|)^{(B,c)} &= (A|c|)(A|c|-|c|) \dots (A|c|-(B-1)|c|) = \\ &= A|c|(A-1)|c| \dots (A-B+1)|c| = c^B A!/(A-B)!. \end{aligned}$$

puede escribirse

$$\begin{aligned} P\{\xi=k\} &= \sum_j \binom{s}{k} \frac{\binom{j}{k} k! \binom{n-j}{s-k} (s-k)!}{\binom{n}{s} s!} \binom{n}{j} \frac{M^{j,c} (N-M)^{n-j,c}}{N^{n,c}} \\ &= \sum_j \frac{\binom{j}{k} \binom{n-j}{s-k}}{\binom{n}{s}} \binom{n}{j} \frac{M^{j,c} (N-M)^{n-j,c}}{N^{n,c}} \end{aligned}$$

Por la conocida propiedad de los números combinatorios,

$$\binom{j}{k} \binom{n}{j} = \binom{n}{k} \binom{n-k}{j-k},$$

la expresión de la función de masa de probabilidad de la distribución resultante de la mixtura quedará

$$P\{\xi=k\} = \sum_j \frac{\binom{n}{k} \binom{n-k}{j-k} \binom{n-j}{s-k} \binom{s}{k}}{\binom{n}{s} \binom{s}{k}} \frac{M^{j,c} (N-M)^{n-j,c}}{N^{n,c}}$$

Haciendo ahora uso de las relaciones:

$$\begin{aligned} M^{(j,c)} &= M^{(k,c)} (M+kc')^{(j-k,c)}, \\ (N'-M')^{(n-j,c)} &= (N'-M')^{(s-k,c)} (N'-M'+(s-k)c')^{(n-j-s+k,c)}, \\ N^{(n,c)} &= N^{(s,c)} (N'+sc')^{(n-s,c)}, \end{aligned}$$

se tiene que

$$P[\xi=k] = \binom{s}{k} \frac{M^{(k,c)}(N'-M')^{(s-k,c)}}{N^{(s,c)}} \sum_j \frac{\binom{n}{k} \binom{n-k}{j-k} \binom{n-j}{s-k}}{\binom{n}{k} \binom{n-k}{s-k}} A$$

donde

$$A = \frac{(M'+kc')^{(j-k,c)}(N'-M'+(s-k)c')^{(n-j-s+k,c)}}{(N'+sc')^{(n-s,c)}}$$

y como

$$\frac{\binom{n-k}{j-k} \binom{n-j}{s-k}}{\binom{n-k}{s-k}} = \binom{n-s}{j-k},$$

la expresión de la función de masa de probabilidad de la distribución resultante de la mezcla se puede escribir como sigue

$$P[\xi=k] = \binom{s}{k} \frac{M^{(k,c)}(N'-M')^{(s-k,c)}}{N^{(s,c)}} \sum_j \binom{n-s}{j-k} \frac{(M'+kc')^{(j-k,c)}(N'-M'+(s-k)c')^{(n-j-s+k,c)}}{(N'+sc')^{(n-s,c)}}$$

donde el sumatorio que aparece es igual a la unidad por ser la suma de las probabilidades puntuales de una distribución de Pólya $P(N'+sc',M'+kc',n-s,c')$. Por tanto,

$$P[\xi=k] = \binom{s}{k} \frac{M^{(k,c)}(N'-M')^{(s-k,c)}}{N^{(s,c)}}$$

Vemos, finalmente, que esta expresión se corresponde con la de la función de probabilidad de la distribución de Pólya $P(N',M',s,c)$, con lo que queda probado el resultado ▮

Para el caso particular de la distribución hipergeométrica se obtiene el siguiente corolario

Corolario 2.2

$$H(N,M,n) \overset{M}{\wedge} H(N',M',N) = H(N',M',n)$$

Demostración

Se obtiene directamente al particularizar $c = -1$, ya que

$$P(N,M,n,-1) = H(N,M,n)$$
▮

En Ramos (1989) se recoge el resultado del anterior corolario y se obtiene, asimismo, el siguiente resultado acerca de la mixtura de una distribución hipergeométrica con una distribución de Pólya, mixtura que está sugerida en el libro de Johnson y Kotz (1969)

Teorema

$$H(N,M,n) \overset{M}{\wedge} P(N',M',N,c) = P(N',M',n,c)$$

TEOREMA DE CARACTERIZACIÓN

A continuación se desarrolla para esta memoria un teorema de caracterización de la distribución de Pólya, basado en el comportamiento de esta distribución frente a la mixtura respecto del parámetro m ($m=M/|c|$) en el caso en que éste también siga una distribución de Pólya.

Como paso previo para la obtención de nuestro resultado fundamental, Teorema 2.4, de caracterización, necesitamos obtener un resultado que se va a enunciar en forma de teorema (Teorema 2.3) en lugar de en forma de lema ya que, como veremos, va a constituir por sí mismo un teorema de caracterización, aunque finalmente quede englobado en el más general Teorema 2.4.

Vamos a necesitar el concepto de *completitud*. Decimos que una familia de distribuciones $\{p_\theta, \theta \in \Omega\}$ de una variable aleatoria X indexada por el parámetro θ en el conjunto Ω es completa si, para cualquier función $u(x)$ independiente de θ , $E[u(x)] = 0$ para cada $\theta \in \Omega$ implica que $u(x) = 0$ para todo x (excepto posiblemente para un conjunto de x con medida de probabilidad cero para todo $\theta \in \Omega$).

Teorema 2.3

Sea P_m ($m: 0, 1, \dots, n$) una familia de distribuciones de probabilidad con función de probabilidad definida en un conjunto \mathfrak{S} de enteros consecutivos contenido en $\{0, 1, \dots, s\}$ con $s \leq n$. Condición necesaria y suficiente para que esta familia sea la constituida por las distribuciones de Pólya $P(N, M, s, c)$ con $c < 0$, siendo $N = n|c|$ y $M = m|c|$, es que se verifique

- (i) Su expresión cuando m sigue una distribución binomial $B(n, \theta)$ es la distribución $B(s, \theta)$
- (ii) Las distribuciones P_m no dependen de θ .

Demostración

Admitamos que la familia P_m es la constituida por las distribuciones de Pólya $P(N, M, s, c)$ con $c < 0$, $N = n|c|$ y $M = m|c|$. La condición (i) queda probada por el Teorema 2.1 sin más que considerar $c' = 0$ y $\theta = M'/N'$, y la condición (ii) se verifica trivialmente. En consecuencia, la condición necesaria del teorema queda probada.

Admitamos ahora que la familia P_m verifica las condiciones (i) y (ii) del teorema. Vamos a probar que esta familia es la constituida por las distribuciones de Pólya $P(N, M, s, c)$ con $c < 0$, $N = n|c|$ y $M = m|c|$. En efecto, por (i) se verificará que

$$\sum_{m=0}^n p_m(k) b(m; n, \theta) = b(k; s, \theta) \tag{2.1}$$

para todo $k: 0, 1, \dots, s$.

Una vez sustituidas en [2.1] las probabilidades binomiales por su correspondiente expresión, se obtiene:

$$\sum_{m=0}^n p_m(k) \binom{n}{m} \theta^m (1-\theta)^{n-m} = \binom{s}{k} \theta^k (1-\theta)^{s-k} \tag{2.2}$$

Esta expresión, para un valor k fijo, es una ecuación lineal con $n+1$ incógnitas p_0, p_1, \dots, p_n . Sabemos por el Teorema 2.1 que

$$p_m(k) = p(k; N, M, s, c) \text{ con } c < 0, N = n|c| \text{ y } M = m|c|, m = 0, 1, \dots, n,$$

es solución de [2.2].

Debemos probar que ésta es la única solución de [2.2]. Para ello consideremos que existe alguna otra solución $f_m(k)$ tal que no dependa de θ . Entonces,

$$\sum_{m=0}^n f_m(k) \binom{n}{m} \theta^m (1-\theta)^{n-m} = \binom{s}{k} \theta^k (1-\theta)^{s-k} \tag{2.3}$$

para todo $k: 0, 1, \dots, s$.

Restando [2.2]-[2.3], se tiene que

$$\sum_{m=0}^n [p_m(k) - f_m(k)] \binom{n}{m} \theta^m (1-\theta)^{n-m} = E[u_m(k)]$$

para todo $k: 0, 1, \dots, s$, siendo $u_m(k) = p_m(k) - f_m(k)$.

Puesto que $u_m(k)$ no depende de θ y la familia de distribuciones binomial $B_\theta(n, \theta)$ es completa [Patel, Kapadia y Owen (1976)] se sigue necesariamente que $u_m(k) = 0$, y de aquí que $p_m(k) = f_m(k)$ para todo $k: 0, 1, \dots, s$, quedando probado el teorema. ■

De la particularización del Teorema 2.3 para el caso $c = -1$, se obtiene el teorema de caracterización de la distribución hipergeométrica dado por Skibinsky (1970) como un caso particular del nuestro.

Teorema

Si h_0, h_1, \dots, h_N son funciones que toman valores en $\{0, 1, \dots, n\}$, y por $b(i, r, \theta)$ notamos la probabilidad de que la variable binomial de parámetros r y θ tome el valor i , entonces:

$$h_j(i) = \frac{\binom{i}{j} \binom{N-j}{n-i}}{\binom{N}{n}}$$

con: $i: 0, 1, \dots, n$; $j: 0, 1, \dots, N$; $n \leq N$, si y sólo si:

- a) los h_j son independientes de θ
- b) $\sum_j h_j(i) b(j; N, \theta) = b(i; n, \theta)$.

El siguiente teorema de caracterización engloba al Teorema 2.3 y constituye el resultado principal de este capítulo.

Teorema 2.4

Sea P_m ($m:0,1,\dots,n$) una familia de distribuciones de probabilidad con función de masa de probabilidad definida en un conjunto de enteros consecutivos contenido en $\{0,1,\dots,s\}$ con $s \leq n$. Condición necesaria y suficiente para que esta familia sea la constituida por las distribuciones de Pólya $P(N,M,s,c)$ con $c < 0$, $N=n|c|$ y $M=m|c|$, es que se verifique:

- (i) Su expresión cuando m sigue una distribución de Pólya (N',M',n,c') es la distribución de Pólya $P(N',M',s,c)$.
- (ii) Las distribuciones P_m no dependen de N' .

Demostración

Condición necesaria:

Admitamos que la familia P_m es la constituida por las distribuciones de Pólya $P(N,M,s,c)$ con $c < 0$, $N=n|c|$ y $M=m|c|$. En este caso, por el Teorema 2.1 se verifica la condición (i), mientras que la condición (ii) se verifica trivialmente.

Condición suficiente:

Consideremos en primer lugar que $c' = 0$. En este caso $P(N',M',n,c') = B(n,\theta)$ y $P(N',M',s,c) = B(s,\theta)$ con $\theta = M'/N'$, y el resultado a probar estaría demostrado con el Teorema 2.3.

Admitamos ahora que $c' \neq 0$ y que la familia P_m verifica las condiciones (i) y (ii). Vamos a probar que esta familia es la constituida por las distribuciones de Pólya $P(N,M,s,c)$ con $c < 0$, $N=n|c|$ y $M=m|c|$.

En efecto, por (i) se verifica que

$$\sum_{m=0}^n p_m(k) p(m; N', M', n, c') = p(k; N', M', s, c') \quad [2.4]$$

para todo $k:0,1,\dots,s$.

Una vez sustituidas las probabilidades puntuales $p(m; N', M', n, c')$ y $p(k; N', M', s, c')$ por sus correspondientes expresiones, [2.4] puede volverse a escribir como

$$\sum_{m=0}^n \binom{n}{m} p_m(k) \frac{M^{(m,c')}(N'-M')^{(n-m,c')}}{N'^{(n,c')}} = \binom{s}{k} \frac{M^{(k,c')}(N'-M')^{(s-k,c')}}{N'^{(s,c')}}$$

que, para cada valor k fijo, representa una ecuación lineal con $n+1$ incógnitas $p_0(k), p_1(k), \dots, p_n(k)$.

Puesto que admitimos por (ii) que las distribuciones I_m^2 no dependen de N' , podemos obtener a partir de la expresión anterior, un sistema lineal de $n+1$ ecuaciones con $n+1$ incógnitas simplemente considerando los valores $N'+ac'$ con $a:0,1,\dots,n$.

Este sistema de ecuaciones admite, al menos, la solución

$$p_m(k) = \binom{s}{k} \frac{M^{(k,c')}(N-M)^{(s-k,c')}}{N^{(s,c')}}$$

como consecuencia inmediata de la condición necesaria del teorema. El teorema quedará demostrado probando que dicho sistema tiene solución única.

La expresión del sistema de ecuaciones lineales es

$$\sum_{m=0}^n p_m(k) \frac{M^{(m,c')}(N'+ac'-M')^{(n-m,c')}}{(N'+ac')^{(n,c')}} = \binom{s}{k} \frac{M^{(k,c')}(N'+ac'-M')^{(s-k,c')}}{(N'+ac')^{(s,c')}}$$

con $a:0,1,\dots,n$; que también puede escribirse como

$$\sum_{m=0}^n p_m(k) \binom{n}{m} M^{(m,c')}(N'+ac'-M')^{(n-m,c')} = \binom{s}{k} (N'+ac')^{(n,c')} \frac{M^{(k,c')}(N'+ac'-M')^{(s-k,c')}}{(N'+ac')^{(s,c')}}$$

con $a:0,1,\dots,n$.

El determinante Δ de la matriz de los coeficientes del sistema anterior es de la forma:

$$\Delta = \begin{vmatrix} M^{(0,c)}(N-M)^{(n,c)} & M^{(1,c)}(N-M)^{(n-1,c)} & \dots & M^{(n,c)}(N-M)^{(0,c)} \\ M^{(0,c)}(N-M+c)^{(n,c)} & M^{(1,c)}(N-M+c)^{(n-1,c)} & \dots & M^{(n,c)}(N-M+c)^{(0,c)} \\ \vdots & \vdots & \vdots & \vdots \\ M^{(0,c)}(N-M+nc)^{(n,c)} & M^{(1,c)}(N-M+nc)^{(n-1,c)} & \dots & M^{(n,c)}(N-M+nc)^{(0,c)} \end{vmatrix}$$

Extrayendo factores fuera del determinante y llamando, por comodidad: $D=N-M$, se tiene que este determinante es equivalente a

$$\Delta = \binom{n}{0} \binom{n}{1} \dots \binom{n}{n} M^{(0,c)} M^{(1,c)} \dots M^{(n,c)} \delta$$

donde

$$\delta = \begin{vmatrix} D^{(n,c)} & D^{(n-1,c)} & \dots & D^{(0,c)} \\ (D+c)^{(n,c)} & (D+c)^{(n-1,c)} & \dots & (D+c)^{(0,c)} \\ \vdots & \vdots & \vdots & \vdots \\ (D+nc)^{(n,c)} & (D+nc)^{(n-1,c)} & \dots & (D+nc)^{(0,c)} \end{vmatrix}$$

La condición suficiente quedará probada demostrando que $\Delta \neq 0$. Para ello vamos a restar a cada fila del determinante δ la fila anterior, y vamos a reiterar el proceso un total de n veces, de forma que en el proceso i -ésimo restaremos a la fila j la fila $j-1$, para $j:n+1, n, \dots, i+1$.

Sin más que tener en cuenta que $A^{(0,c)}=1$, y que

$$(D+tc)^{(v,c)} - (D+(t-1)c)^{(v,c)} = (D+tc)^{(v-1,c)} \cdot vc' \quad (v:1,2,\dots,n)$$

el determinante que se obtiene tras el proceso anteriormente descrito es el siguiente:

$$\begin{vmatrix}
 D^{(n,c')} & D^{(n-1,c')} & \dots & D^{(2,c')} & D & 1 \\
 (D+c')^{(n-1,c')}nc' & (D+c')^{(n-2,c')}(n-1)c' & \dots & (D+c')2c' & c' & n \\
 (D+2c')^{(n-2,c')}n(n-1)(c')^2 & (D+2c')^{(n-3,c')}(n-1)(n-2)(c')^2 & \dots & 2(c')^2 & 0 & 0 \\
 \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
 \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
 n!(c')^n & 0 & \dots & 0 & 0 & 0
 \end{vmatrix}$$

de manera que todos los elementos de este determinante por debajo de la diagonal secundaria son nulos, mientras que los de esa diagonal son todos distintos de cero, ya que estamos admitiendo que $c' \neq 0$. Consecuentemente, el valor del determinante es distinto de cero, quedando de esta manera probado el teorema.

Corolario 2.5

Sea $\{F_M\}$, $M=0,1,\dots,N$, una familia de funciones de distribución que asignan probabilidades no nulas sólo a los enteros $\{0,1,\dots,n\}$. Condición necesaria y suficiente para que esta familia sea la formada por la familia de distribuciones hipergeométricas $H(N,M,n)$, es que se verifique:

- 1) Su composición cuando M sigue una distribución hipergeométrica $H(N',M',N)$, es la distribución hipergeométrica $H(N',M',n)$.
- 2) Las distribuciones F_M no dependen del valor M' .

El resultado de este corolario fue probado en Ramos y Ollero (1989), y se obtiene directamente del Teorema 2.4, como caso particular del mismo.

CAPÍTULO III

DESCRIPCIÓN DE LA

DISTRIBUCIÓN DE PÓLYA

COMO BINOMIAL GENERALIZADA.

APLICACIONES

CAPÍTULO III

DESCRIPCIÓN DE LA DISTRIBUCIÓN DE PÓLYA COMO BINOMIAL GENERALIZADA. APLICACIONES

La descripción de la distribución de Pólya con parámetro reemplazamiento negativo como distribución binomial de Poisson o binomial generalizada, permite obtener algunas nuevas propiedades de la distribución de Pólya y redemostrar otras ya conocidas. Nuestra aportación en este capítulo se va a centrar en utilizar esta nueva descripción para obtener resultados acerca de la distribución de la suma y diferencia de distribuciones de Pólya.

LA DISTRIBUCIÓN BINOMIAL GENERALIZADA

La distribución binomial generalizada o binomial de Poisson aparece a raíz del interés que Poisson (1837) tiene en el problema de n pruebas en las que la probabilidad de éxito varía de una a otra. Para plantear el problema, considera k tiradas independientes de n dados donde p_{ij} es la probabilidad de éxito para el dado i en la tirada j , y supone que cada resultado de cada dado es un experimento de Bernoulli de probabilidad p_{ij} , [Johnson, Kotz y Kemp (1992)].

Se define la distribución binomial generalizada o binomial de Poisson como la distribución que sigue el número total de éxitos que ocurren en un conjunto de pruebas independientes con probabilidades de éxito arbitrarias. Cada una de estas pruebas recibe el nombre de prueba o ensayo de Poisson. En el caso particular en que todas las pruebas sean equiprobables, se emplea el nombre de ensayo de Bernoulli.

Una distribución binomial generalizada queda determinada por las probabilidades de éxito de todas las pruebas que la definen, y se llama vector de parámetros al que tiene por componentes las probabilidades de éxito de cada una de las pruebas del conjunto considerado.

En lo sucesivo se van a considerar n pruebas independientes siendo la probabilidad de éxito en cada una de ellas p_i , $i: 1, \dots, n$. Así el vector de parámetros $\mathbf{p} = (p_1, p_2, \dots, p_n)$ determina la distribución binomial generalizada $B(\mathbf{p})$. Notaremos por $b_k(\mathbf{p})$ al valor de su función de masa de probabilidad en el punto k y por $B_k(\mathbf{p})$ a su función de distribución. Para las correspondientes funciones de la distribución binomial (de Bernoulli) $B(n, p)$ se empleará la notación $b_k(n, p)$ y $B_k(n, p)$.

La función de masa de probabilidad de la distribución binomial generalizada viene dada por la expresión

$$b_k(\mathbf{p}) = \sum_{\substack{i_j=0,1 \\ \sum_{j=1}^n i_j=k}} \prod_{j=1}^n p_j^{i_j} (1-p_j)^{1-i_j}$$

donde cada sumando representa la probabilidad de que los k éxitos se consigan en aquellas pruebas en que i_j sea igual a 1.

La función generatriz de probabilidad de la distribución binomial generalizada es

$$G_B(z) = \prod_{i=1}^n (q_i + p_i z)$$

Obsérvese que el lugar ocupado por cada componente en el vector de probabilidad es irrelevante ya que $G_B(z)$ es una función simétrica de dichas probabilidades.

Resulta claro que, en el caso en que $p_i = p$ para todo $i: 1, 2, \dots, n$, se obtiene la distribución binomial como caso particular de la distribución binomial generalizada.

LA DISTRIBUCIÓN DE PÓLYA COMO BINOMIAL GENERALIZADA

En el Capítulo I hicimos mención al trabajo de Ollero y Ramos (1995) en el que se describe una subfamilia del sistema de Pearson discreto como distribuciones binomiales generalizadas o binomiales de Poisson. Dicha subfamilia es notada por \mathcal{P}_H y se prueba la pertenencia de las distribuciones de Pólya con parámetro reemplazamiento negativo a \mathcal{P}_H y, en consecuencia, la descripción de estas distribuciones en términos de pruebas independientes. Es decir, cualquier distribución de Pólya $P(N, M, n, c)$ con $c < 0$ que, en principio, está asociada a un modelo de pruebas dependientes, se distribuye como el número de éxitos que tienen lugar en una secuencia de pruebas independientes.

Debemos hacer notar que, a diferencia del tratamiento que le estamos dando en esta memoria, en el citado trabajo de Ollero y Ramos no se impone sobre los parámetros de la distribución de Pólya la condición $(-c)(n-1) < \min\{M, N-M\}$. Por tanto, el rango de la distribución puede no ser el rango máximo $[0, n]$, sino que puede tener rango $[m, s]$ con $m > 0$ y $s \leq n$.

La descripción de la distribución de Pólya sin reemplazamiento ($c < 0$) como binomial generalizada se lleva a cabo a través de la identificación de las correspondientes funciones generatrices de probabilidad. Para ello se obtiene la función generatriz de probabilidad de las distribuciones pertenecientes a la subfamilia \mathcal{P}_H . En particular, para el caso de una distribución de Pólya $P(N, M, n, c)$, $c < 0$, con rango cualquiera, se obtiene la expresión de la función generatriz de probabilidad:

$$G_p(z) = f_m z^m H \left(-n+m, \frac{M}{c}+m, -\frac{N-M+c}{c} \frac{(n-1)}{+2m}; z \right),$$

donde f_m es el valor de la función de masa de probabilidad en el menor entero m con probabilidad no nula, y H es la función hipergeométrica definida por

$$H(\alpha, \beta, \gamma; z) = \sum_{k=0}^{\infty} \frac{\alpha^{(k)} \beta^{(k)}}{\gamma^{(k)}} \frac{z^k}{k!}$$

con $\gamma > 0$ y $A^{(k)} = A(A+1)\dots(A+k-1)$; $A^{(0)} = 1$.

Se prueba que $G_p(z)$ tiene s raíces reales, de las cuales $s-m$ son raíces simples y las m restantes son iguales a cero. Como consecuencia, $G_p(z)$ puede ser escrita como

$$G_p(z) = f_m \prod_{i=1}^s (z - z_i)$$

Por otra parte, la función generatriz de probabilidad de una distribución binomial generalizada con vector de probabilidades $\mathbf{p}=(p_1, p_2, \dots, p_s)$ con $p_i > 0$ para todo $i=1, 2, \dots, s$, es

$$G_B(z) = \prod_{i=1}^s (q_i + p_i z) = \prod_{i=1}^s p_i \prod_{i=1}^s \left(z + \frac{q_i}{p_i} \right)$$

Para hacer coincidir ambas funciones generatrices de probabilidad es suficiente con que

$$\frac{-q_i}{p_i} = z_i, \text{ o lo que es lo mismo, basta tomar } p_i = \frac{1}{1-z_i}, \text{ } i=1, 2, \dots, s.$$

Se observa que las $s-m$ raíces simples de $G_p(z)$ se corresponden con pruebas con probabilidad de éxito distintas entre sí, mientras que la raíz $z_i = 0$ con grado de multiplicidad m se corresponde con m pruebas determinísticas de éxito seguro en la descripción de la distribución de Pólya como binomial generalizada.

En Ollero y Ramos (1995) se considera como ejemplo el caso de la distribución de Pólya $P(12, 6, 3, -2)$, que tiene rango $[0, 3]$ y cuya función de masa de probabilidad vale

$$f(0)=1/20 ; f(1)=9/20 ; f(2)=9/20 ; f(3)=1/20 .$$

Por tanto, su función generatriz de probabilidad es

$$G(z) = (1/20)(1+9z+9z^2+z^3),$$

y sus raíces son

$$z_1 = -4 + \sqrt{15} ; z_2 = -1 ; z_3 = -4 - \sqrt{15}$$

De acuerdo con el teorema anterior, los valores p_i correspondientes son

$$p_1 = (5 - \sqrt{15})/10 ; p_2 = 1/2 ; p_3 = (5 + \sqrt{15})/10 ,$$

y de aquí se tiene la equivalencia

$$P(12,6,3,-2) \equiv \mathfrak{B}(0.5 - \sqrt{0.15} ; 0.5 ; 0.5 + \sqrt{0.15})$$

ALGUNOS RESULTADOS Y PROPIEDADES DE LA DISTRIBUCIÓN DE PÓLYA SIN REEMPLAZAMIENTO DERIVADAS DE SU DESCRIPCIÓN COMO BINOMIAL GENERALIZADA

Como consecuencia de la descripción de la distribución de Pólya con parámetro reemplazamiento negativo como distribución binomial generalizada, propiedades asociadas a esta última distribución van a poder ser aplicadas para, en unos casos, obtener resultados acerca de la distribución de Pólya que, de otra forma, serían difíciles de probar y, en otros casos, pasar a deducir de una manera más simple propiedades ya conocidas de esta distribución

1.- LOCALIZACIÓN DE LA MODA Y DE LA MEDIANA

Los resultados sobre la localización de la mediana y de la moda en la distribución binomial generalizada conducen a la conclusión de que tanto la moda como la mediana de la distribución de Pólya sin reemplazamiento pueden localizarse únicamente en los enteros adyacentes a su media μ . Si la media fuera entero, entonces tanto la moda como la mediana tomarían este valor.

Los resultados concretos sobre localización de moda y mediana se recogen en los siguientes teoremas.

Teorema [Darroch (1964)]

La moda de una variable binomial de Poisson es

- i) μ , si $\delta(\mu) = 0$
- ii) $[\mu]$, si $0 < \delta(\mu) < ([\mu]+2)^{-1}$
- iii) $[\mu+1]$, si $n-[\mu] < \delta(\mu) < (n-[\mu]+1)^{-1}$
- iv) $[\mu]$, $[\mu+1]$, o ambos, si $([\mu]+2)^{-1} \leq \delta(\mu) \leq (n-[\mu])^{-1}$

siendo $\delta(\mu)$ la parte fraccionaria de μ , o sea, $\mu - [\mu]$.

Teorema [Samuels (1965)]

Sean $p_{[1]}$ y $p_{[n]}$, respectivamente, la mayor y la menor componentes del vector de probabilidades de una distribución binomial generalizada. Si $k \leq \mu \leq k+1$, entonces la moda será:

- i) k , si $\mu + p_{[1]} < k + 1$
- ii) $k+1$, si $\mu + p_{[n]} > k + 1$
- iii) $k, k+1$ o ambos, en cualquier otro caso.

Teorema [Jogdeo y Samuels (1968)]

Si $k \leq \mu < k + 1$, la mediana de la distribución binomial generalizada con media μ es:

- (i) k , si $\mu < \lambda_1$
- (ii) $k+1$, si $\mu > \lambda_2$

donde λ_1 y λ_2 son las soluciones respectivas de:

$$\frac{1}{2} = \min_{r=0, \dots, n-k-1} B_k \left(n-r, \frac{\lambda_1}{n-r} \right) ; \quad \frac{1}{2} = \min_{s=0, \dots, k} B_k \left(n-s, \frac{\lambda_2 - s}{n-s} \right)$$

2.- SOBRE LA FORMA DE LA DISTRIBUCIÓN

Samuels (1965) define por recurrencia las sucesivas diferencias a partir de la probabilidad $b_k(p)$ de obtener k éxitos en t pruebas independientes:

$$\begin{aligned} D_0(k) &= b_k(p) \\ D_r(k) &= D_{r-1}(k) - D_{r-1}(k-1) ; k:1,2,\dots \end{aligned}$$

y prueba el siguiente conjunto de desigualdades:

$$\left[\frac{D_r(k)}{\binom{t+r}{k}} \right]^2 > \frac{D_r(k-1) D_r(k+1)}{\binom{t+r}{k-1} \binom{t+r}{k+1}}$$

para $r:0,1,\dots$ y $k:1,\dots,t+r-1$.

Una versión débil de la anterior desigualdad es

$$D_r^2(k) > D_r(k-1) D_r(k+1),$$

que indica que D_{r-1} puede tener como máximo un cambio estricto de signo entre sucesivos cambios estrictos de signo de D_r . Por tanto, D_r tiene como máximo r cambios estrictos de signo. En este caso se dice que $b_k(p)$ es acampanada de orden r para todo k .

Consideremos ahora la distribución de Pólya $P(N,M,n,c)$, $c < 0$, con rango $[m,s]$ y notemos p_k al valor de su función de masa de probabilidad en el punto k . Si particularizamos la anterior desigualdad para el caso $r=0$, se tiene que

$$p_k^2 > p_{k-1} p_{k+1} ; k : m+1, m+2, \dots, s-1 .$$

Esta condición indica que la distribución de Pólya es fuertemente unimodal o, lo que es lo mismo, que p_k es log-cóncava [Keilson y Gerber (1971)].

3.- ACOTACIONES PARA LA FUNCIÓN DE DISTRIBUCIÓN

En Hoeffding (1956) se acota la función de distribución $B_k(\mathbf{p})$ de una distribución binomial generalizada con media μ en la siguiente forma

$$\begin{aligned} \text{(i)} \quad 0 \leq B_k(\mathbf{p}) \leq B_k(n; \mu/n) & \quad , \quad 0 \leq k \leq \mu-1 \\ \text{(ii)} \quad B_k(n; \mu/n) \leq B_k(\mathbf{p}) \leq 1 & \quad ; \quad \mu \leq k \leq n \end{aligned}$$

donde $B_k(n, \mu/n)$ indica el valor de la función de distribución de la distribución binomial de Bernoulli $B(n; \mu/n)$ en el punto k .

Este resultado se adaptó en Ollero y Ramos (1995) al caso de una distribución de Pólya $P(N, M, n, c)$ con parámetro reemplazamiento negativo, teniendo en cuenta el rango $[m, s]$ de la misma que indica que en su descripción como binomial generalizada habrá m pruebas determinísticas con probabilidad de éxito uno y $n-s$ pruebas determinísticas con probabilidad de éxito cero. En este caso se pueden considerar dos distribuciones aproximantes alternativas con el mismo valor esperado μ : la distribución $B(t; \mu/s)$ y la distribución binomial desplazada $m+B(s-m; (\mu-m)/(s-m))$, probándose que esta última es la que da lugar a las mejores acotaciones.

Las cotas obtenidas para la función de distribución P_k de la distribución de Pólya son

$$\begin{aligned} \text{(i)} \quad P_k = 0 & \quad \text{si} \quad 0 \leq k \leq m-1 \\ \text{(ii)} \quad 0 < P_k < B_{k-m} & \quad \text{si} \quad m \leq k \leq \mu-1 \\ \text{(iii)} \quad B_{k-m} < P_k < 1 & \quad \text{si} \quad \mu \leq k < s \\ \text{(iv)} \quad P_k = 1 & \quad \text{si} \quad s \leq k \leq n \end{aligned}$$

donde con B_k se nota la función de distribución de la distribución binomial $B(t-m, (\mu-m)/(t-m))$.

Si consideramos una distribución de Pólya $P(N, p, n, c)$ con $c < 0$ y con rango $[0, n]$, como venimos considerando a lo largo del desarrollo de esta memoria, la acotación para la función de distribución P_k vendría dada por

$$\begin{aligned} \text{(i)} \quad 0 < P_k < B_k & \quad \text{si} \quad 0 \leq k \leq \mu-1 \\ \text{(ii)} \quad B_k < P_k < 1 & \quad \text{si} \quad \mu \leq k \leq n-1 \end{aligned}$$

siendo $\mu = pn$ y donde B_k es el valor de la función de distribución de la distribución binomial $B(n,p)$ en el punto k .

Este resultado nos indica que la distribución de Pólya $P(N,p,n,c)$ y la correspondiente binomial aproximante $B(n,p)$ son comparables en todo punto excepto para valores de k pertenecientes al intervalo $(\mu-1;\mu)$ que, en este sentido, podemos denominar *intervalo de indeterminación*.

En el Capítulo V obtendremos una acotación que mejora la presentada en Ollero y Ramos (1995) ya que conseguiremos reducir el intervalo de indeterminación

4.- INTERVALOS DE CONFIANZA Y TEST DE HIPÓTESIS

Como consecuencia de las anteriores acotaciones de la función de distribución de la distribución de Pólya por la correspondiente distribución binomial, el test usual para el parámetro p de la distribución binomial puede aplicarse directamente para contrastar la correspondiente hipótesis sobre el parámetro $p=M/N$ en una distribución de Pólya $P(N,M,n,c)$ con $c < 0$. En este caso, el nivel de significación de test será como máximo el valor nominal de dicho nivel en el test binomial utilizado. De manera similar puede probarse que la potencia de dicho test cuando se tiene un valor alternativo no demasiado cercano al de la hipótesis nula es, como mínimo, igual a la potencia del correspondiente test binomial

En el mismo sentido, el intervalo de confianza para el parámetro p binomial puede utilizarse para estimar el correspondiente parámetro en la distribución de Pólya sin reemplazamiento. Cuando el nivel de confianza es suficientemente grande, el nivel de confianza real será mayor que el nominal en el intervalo para el parámetro binomial

SUMA Y DIFERENCIA DE DISTRIBUCIONES DE PÓLYA SIN REEMPLAZAMIENTO

En el siguiente epígrafe vamos a obtener unos resultados para la distribución de Pólya sin reemplazamiento (como se ha venido haciendo a lo largo del desarrollo de esta memoria, se va a trabajar con distribuciones de Pólya de rango $[0, n]$), que no aparecen recogidos en Ollero y Ramos (1995) y que se derivan de la particularización al caso de la distribución de Pólya de una caracterización de la distribución binomial generalizada dada por Ollero (1988)

Teorema [Ollero (1988)]

Sea Z una variable aleatoria obtenida como suma de dos variables X e Y , independientes y no negativas. Una condición necesaria y suficiente para que Z se distribuya según una distribución binomial generalizada $B(p_z)$ es que X e Y sigan también distribuciones binomiales generalizadas, cuyos vectores de parámetros, p_x y p_y respectivamente, verifiquen que $p_z = (p_x, p_y)$

1.- SOBRE LA SUMA DE DISTRIBUCIONES DE PÓLYA SIN REEMPLAZAMIENTO

Como consecuencia de la descripción de la distribución de Pólya con parámetro reemplazamiento negativo como binomial generalizada, y teniendo en cuenta el teorema anterior [Ollero (1988)] de caracterización de estas distribuciones, se siguen los siguientes corolarios

Corolario 3.1

La suma de dos variables aleatorias independientes X e Y que siguen distribuciones de Pólya con parámetro reemplazamiento negativo, se distribuye según una binomial generalizada.

Demostración

Basta tener en cuenta que al poder expresarse X e Y como distribuciones binomiales generalizadas con vectores de parámetros p_x y p_y , respectivamente, entonces $X+Y$ seguirá una distribución binomial generalizada con vector de parámetros p_z siendo $p_z = (p_x, p_y)$.

Corolario 3.2

Si una variable aleatoria Z sigue una distribución de Pólya con parámetro reemplazamiento negativo, y Z puede expresarse como suma de dos variables aleatorias independientes X e Y no negativas, entonces X e Y siguen distribuciones binomiales generalizadas

Demostración

Basta tener en cuenta que Z sigue una distribución binomial generalizada y aplicar la condición necesaria del teorema referido [Ollero (1988)]

■

2.- DIFERENCIA DE DISTRIBUCIONES DE PÓLYA SIN REEMPLAZAMIENTO

Consideremos X e Y variables aleatorias independientes que siguen distribuciones de Pólya con parámetros reemplazamiento negativo y con tamaño muestral n_X y n_Y , respectivamente. X e Y pueden expresarse como variables binomiales generalizadas $B(p_X)$ y $B(p_Y)$ respectivamente

Consideremos la diferencia $V = X - Y$. La variable aleatoria V tomará los valores enteros comprendidos entre $-n_Y$ y n_X . Si desplazamos V n_Y unidades a la derecha, es decir, si consideramos la variable aleatoria $V + n_Y$, tendremos que

$$V + n_Y = X - Y + n_Y = X + (n_Y - Y)$$

Puesto que la variable $n_Y - Y$ es el número de fracasos que ocurren en las n_Y pruebas, se tiene que

$$n_Y - Y \rightarrow B(q_Y)$$

En consecuencia, a partir del teorema desarrollado por Ollero (1988), se sigue que

$$V + n_Y \rightarrow B(p_V)$$

con $p_V = (p_X, q_Y)$

La conclusión puede ser expresada en el siguiente corolario

Corolario 3.3

Sean X e Y variables aleatorias independientes que siguen distribuciones de Pólya con parámetro reemplazamiento negativo y sea $V = X - Y$. Entonces existe un entero n tal que $V+n$ sigue una distribución binomial generalizada.

CAPÍTULO IV

LA FUNCIÓN DE DISTRIBUCIÓN DE LA DISTRIBUCIÓN DE PÓLYA

CAPÍTULO IV

LA FUNCIÓN DE DISTRIBUCIÓN DE LA DISTRIBUCIÓN DE PÓLYA

En este capítulo se prueban nuevas propiedades para la función de distribución de la distribución de Pólya sin establecer restricciones sobre el parámetro reemplazamiento. Así, todas las propiedades que se presentan son aplicables a distribuciones derivadas de la distribución de Pólya, como es el caso de las distribuciones binomial, hipergeométrica e hipergeométrica negativa

LA FUNCIÓN DE DISTRIBUCIÓN

Como se ha indicado, consideraremos el esquema de urna de Pólya con N bolas de las cuales M son blancas y $N-M$ negras. En cada extracción se devolverá la bola obtenida junto con c bolas más del mismo color. La operación se repite n veces. Cuando el parámetro c tome valores negativos se añadirá, como se ha especificado, la condición

$$(-c)(n-1) < \min\{M, N-M\} \quad [4.1]$$

que garantiza que la distribución tiene rango completo $[0, n]$

La función de distribución de la distribución de Pólya tiene la forma

$$F(x) = P\{X \leq x\} = \begin{cases} 0 & ; x < 0 \\ \sum_{m=0}^k \binom{n}{m} \frac{(pN)^{(m,c)}(qN)^{(n-m,c)}}{N^{(n,c)}} & ; k \leq x < k+1 \quad (k=0,1,\dots,n) \\ 1 & ; x > n \end{cases}$$

A continuación se van a desarrollar dos propiedades de la función de distribución de la distribución de Pólya a las que se hará referencia en el Capítulo V. El interés radica en estudiar el comportamiento de esta función cuando varía la composición inicial de la urna mediante la adición, si $c > 0$, o extracción, si $c < 0$, de c bolas blancas.

RESULTADOS

Sea $F(x)$ la función de distribución de una distribución de Pólya $P(N, M, n, c)$.

Por $F'(x)$ notaremos a la función de distribución correspondiente a una distribución de Pólya $P(N, M+c, n, c)$. En el caso de $c < 0$ admitiremos que los parámetros de esta distribución satisfacen la condición [4.1] y, por tanto, que tiene, asimismo, rango $[0, n]$

Teorema 4.1

Para todo $x: 0, 1, \dots, n-1$, se verifica:

$$F(x) - F'(x) = \binom{n}{x} (n-x)c \frac{(M+c)^{(x,c)} (N-M)^{(n-x-1,c)}}{N^{(n,c)}} \quad [4.2]$$

Demostración

Se va a probar por inducción

Para $x=0$:

$$\begin{aligned} F(0) - F'(0) &= \binom{n}{0} \frac{1}{N^{(n,c)}} [M^{(0,c)} (N-M)^{(n,c)} - (M+c)^{(0,c-1)} (N-M-c)^{(n,c)}] = \\ &= \frac{1}{N^{(n,c)}} [(N-M)^{(n,c)} - (N-M-c)^{(n,c)}] \end{aligned}$$

Teniendo en cuenta que

$$(N-M)^{(n,c)} = (N-M)^{(n-1,c)} \cdot (N-M - (n-1)c),$$

$$(N-M-c)^{(n,c)} = (N-M)^{(n-1,c)} \cdot (N-M-c) ;$$

la expresión anterior se puede escribir como

$$F(0) - F'(0) = \frac{(N-M)^{(n-1,c)}}{N^{(n,c)}} [N-M+(n-1)c-N+M+c] = \frac{(N-M)^{(n-1,c)}}{N^{(n,c)}} nc$$

que coincide con [4.2] en el caso particular en que $x=0$

Supuesto válido hasta $x-1$, se va a probar para x . En efecto,

$$\begin{aligned} F(x) - F'(x) &= \\ &= F(x-1) - F'(x-1) + \binom{n}{x} \frac{1}{N^{(n,c)}} [M^{(x,c)}(N-M)^{(n-x,c)} - (M+c)^{(x,c)}(N-M-c)^{(n-x,c)}] \end{aligned}$$

y teniendo en cuenta que

$$M^{(x,c)} = (M+c)^{(x,c)} \frac{M}{(M+xc)}$$

$$(N-M)^{(n-x,c)} = (N-M-c)^{(n-x,c)} \frac{(N-M+(n-x-1)c)}{(N-M-c)}$$

la expresión anterior se puede escribir como sigue

$$\begin{aligned} F(x) - F'(x) &= \\ &= F(x-1) - F'(x-1) + \binom{n}{x} \frac{(M+c)^{(x,c)}(N-M)^{(n-x,c)}}{N^{(n,c)}} \left[\frac{M}{M+xc} - \frac{N-M-c}{N-M+(n-x-1)c} \right] = \\ &= F(x-1) - F'(x-1) + \binom{n}{x} \frac{(M+c)^{(x,c)}(N-M)^{(n-x,c)}}{N^{(n,c)}} \left[\frac{Mnc - Nxc + xc^2}{(M+xc)(N-M+(n-x-1)c)} \right] \end{aligned}$$

sin más que operar la expresión entre corchetes. Por otra parte, de acuerdo con [4.2],

$$F(x-1) - F'(x-1) = \binom{n}{x-1} (n-x+1)c \frac{(M+c)^{(x-1,c)}(N-M)^{(n-x,c)}}{N^{(n,c)}}$$

Teniendo en cuenta que

$$\begin{aligned} (N-M)^{(n-x,c)} &= (N-M)^{(n-x-1,c)} (N-M+(n-x-1)c) \\ (M+c)^{(x,c)} &= (M+c)^{(x-1,c)} (M+xc) \end{aligned} \quad [4.3]$$

se puede escribir

$$F(x-1) - F'(x-1) = \binom{n}{x} xc \frac{(M+c)^{(x,c)}(N-M)^{(n-x-1,c)} (N-M+(n-x-1)c)}{N^{(n,c)} (M+xc)}$$

En consecuencia, y sin más que tener en cuenta de nuevo [4.3],

$$F(x-1) - F'(x-1) = \binom{n}{x} \frac{(M+c)^{(x,c)}(N-M)^{(n-x-1,c)}c}{N^{(n,c)}} \left[\frac{Mn-Nx+xc+x(N-M+(n-x-1)c)}{M+xc} \right]$$

Y el corchete puede simplificarse ya que

$$\begin{aligned} \frac{Mn-Nx+xc+xN-Mx+xcn-x^2c-xc}{M+xc} &= \frac{Mn-Mx+xcn-x^2c}{M+xc} = \\ &= \frac{M(n-x)+xc(n-x)}{M+xc} = \frac{(M+xc)(n-x)}{M+xc} = (n-x) \end{aligned}$$

con lo que queda probada la proposición. ■

Nota: La expresión [4.2] no es válida para $x=n$ ya que $(N-M)^{(n-x-1,c)}$ no estaría definido para dicho valor. Por otra parte, puesto que estamos admitiendo rango $[0,n]$ para ambas distribuciones, tenemos que

$$F(n) - F'(n) = 1 - 1 = 0.$$

El siguiente corolario se obtiene como particularización del teorema anterior cuando $c=1$. En este caso, el rango $[0,n]$ está garantizado por ser c positivo.

Corolario 4.2

La función de distribución de la distribución hipergeométrica negativa verifica

$$F(x) - F'(x) = \frac{\binom{M+x}{x} \binom{N-M+n-x-2}{n-x-1}}{\binom{N+n-1}{n}} ; \quad x=0,1,\dots,n-1$$

Demostración

Aplicando directamente el Teorema 4.1 para $c = 1$, se tiene que

$$F(x) - F'(x) = \binom{n}{x} (n-x) \frac{(M+1)^{(x,1)} (N-M)^{(n-x-1,1)}}{N^{(n,1)}}$$

y como

$$(M+1)^{(x,1)} = x! \binom{M+x}{x}$$

$$N^{(n,1)} = n! \binom{N+n-1}{n}$$

$$(N-M)^{(n-x-1,1)} = (n-x-1)! \binom{N-M+n-x-2}{n-x-1}$$

se llega a la expresión dada sin más que tener en cuenta que

$$\binom{n}{x} (n-x) \frac{x!(n-x-1)!}{n!} = 1$$

■

A continuación se va a considerar $F''(x)$, denominándose así a la función de distribución correspondiente a una distribución de Pólya $P(N, M-c, n, c)$

Siempre que la distribución $P(N, M-c, n, c)$ tenga rango completo $[0, n]$, se verifica el siguiente corolario.

Corolario 4.3

$$F''(x) - F'(x) = \binom{n}{x} (n-x)c \frac{(M)^{(x,c)}(N-M+c)^{(n-x-1,c)}}{N^{(n,c)}}, \quad x = 1, 2, \dots, n-1$$

Demostración

Este resultado se deriva directamente del obtenido en el Teorema 4.1

■

Los dos siguientes corolarios hacen referencia a la función de distribución de distribuciones hipergeométricas. Si por $F(x)$ notamos ahora a la función de distribución de la hipergeométrica $H(N, M, n)$, $F'(x)$ y $F''(x)$ serán, respectivamente, las funciones de distribución de las hipergeométricas $H(N, M-1, n)$ y $H(N, M+1, n)$

Corolario 4.4

La función de distribución de la distribución hipergeométrica verifica:

$$F(x) - F''(x) = \frac{\binom{M}{x} \binom{N-M-1}{n-x-1}}{\binom{N}{n}}, \quad x = 0, 1, \dots, n-1$$

Demostración

De la particularización del Corolario 4.3 al caso en que $c = -1$, se tiene que

$$F''(x) - F(x) = -(F(x) - F''(x)) = -\binom{n}{x}(n-x)(-1)\frac{(M)^{(x,-1)}(N-M-1)^{(n-x-1,-1)}}{N^{(n,-1)}}$$

y como

$$M^{(x,-1)} = \binom{M}{x}x!$$

$$(N-M-1)^{(n-x-1,-1)} = \binom{N-M-1}{n-x-1}(n-x-1)!$$

$$N^{(n,-1)} = \binom{N}{n}n!$$

la diferencia anterior puede escribirse como sigue:

$$F(x) - F''(x) = \binom{n}{x}(n-x)\frac{\binom{M}{x}x!\binom{N-M-1}{n-x-1}(n-x-1)!}{\binom{N}{n}n!}$$

de donde se obtiene el resultado buscado. ■

Este resultado fue obtenido por Uhlmann (1966) en un estudio acerca de la distribución hipergeométrica siguiendo una metodología distinta.

El corolario siguiente es una particularización del Teorema 4.1, cuando $c = -1$.

Corolario 4.5

La función de distribución de la distribución hipergeométrica verifica:

$$F'(x) - F(x) = \frac{\binom{M-1}{x} \binom{N-M}{n-x-1}}{\binom{N}{n}} ; x:0,1,\dots,n-1$$

Demostración

De la particularización del Teorema 4.1 al caso en que $c = -1$, se tiene que

$$F'(x) - F(x) = -(F(x) - F'(x)) = -\binom{n}{x}(n-x)(-1) \frac{(M-1)^{(x,-1)}(N-M)^{(n-x-1,-1)}}{N^{(n,-1)}}$$

y como

$$(M-1)^{(x,-1)} = \binom{M-1}{x} x!$$

$$(N-M)^{(n-x-1,-1)} = \binom{N-M}{n-x-1} (n-x-1)!$$

$$N^{(n,-1)} = \binom{N}{n} n!$$

la diferencia anterior puede escribirse como sigue:

$$F'(x) - F(x) = \binom{n}{x}(n-x) \frac{\binom{M-1}{x} x! \binom{N-M}{n-x-1} (n-x-1)!}{\binom{N}{n} n!}$$

de donde se obtiene el resultado buscado.

Teorema 4.6

Para valores de $x:0,1,\dots,n-1$, se verifica que la segunda diferencia de la función de distribución de la distribución de Pólya, tiene la siguiente expresión:

$$F''(x) - 2F'(x) + F(x) = \binom{n}{x} \frac{(n-x)c(M+c)^{(x,c)}(N-M)^{(n-x-1,c)}}{N^{(n,c)}} \frac{Nc(p(n-1)-x)}{(M+xc)(N-M)}$$

Demostración

La expresión de la segunda diferencia también puede escribirse:

$$F''(x) - 2F'(x) + F(x) = (F''(x)-F'(x)) - (F'(x)-F(x))$$

Aplicando los resultados del Teorema 4.1 y del Corolario 4.3, se tiene que

$$\binom{n}{x} \frac{(n-x)c}{N^{(n,c)}} (M^{(x,c)}(N-M+c)^{(n-x-1,c)} - (M+c)^{(x,c)}(N-M)^{(n-x-1,c)})$$

que es igual a

$$\binom{n}{x} (n-x)c \frac{(M+c)^{(x,c)}(N-M)^{(n-x-1,c)}}{N^{(n,c)}} \left(\frac{M}{M+xc} \frac{N-M+(n-x-1)c}{N-M} - 1 \right) \tag{4.4}$$

ya que

$$M^{(x,c)} = (M+c)^{(x,c)} \frac{M}{(M+xc)}$$

$$(N-M)^{(n-x,c)} = (N-M-c)^{(n-x,c)} \frac{(N-M+(n-x-1)c)}{(N-M-c)}$$

Operando ahora el paréntesis de la expresión [4.4], resulta que

$$\frac{M(N-M) + M(n-x-1)c - M(N-M) - xc(N-M)}{(M+xc)(N-M)} =$$

$$= \frac{M(n-1)c - Nxc}{(M+xc)(N-M)} = \frac{Nc(p(n-1)-x)}{(M+xc)(N-M)}$$

quedando, de esta forma, probado el teorema. ■

Observamos que esta segunda diferencia se anula para $x = p(n-1)$. (Basta sustituir en la expresión obtenida para comprobarlo).

Corolario 4.7

La función de distribución de la distribución hipergeométrica verifica la siguiente expresión para la segunda diferencia:

$$F''(x) - 2F'(x) + F(x) = \frac{\binom{M-1}{x} \binom{N-M}{n-x-1}}{\binom{N}{n}} - \frac{\binom{M}{x} \binom{N-M-1}{n-x-1}}{\binom{N}{n}} ; x:0,1,\dots,n-1$$

Demostración

Se trata de una particularización del Teorema 4.6 al caso en que $c=-1$. También puede deducirse inmediatamente a partir de los Corolarios 4.4 y 4.5. ■

Uhlmann (1996) obtuvo también la expresión correspondiente a la segunda diferencia de la distribución hipergeométrica siguiendo una metodología distinta.

En el siguiente teorema se van a comparar las funciones de distribución de dos distribuciones de Pólya con distintos valores en el parámetro c . No se harán restricciones sobre el signo que tome este parámetro. Se va a notar por $F_{N,M,n,c}(x)$ a la función de distribución de una distribución de Pólya de parámetros $P(N,M,n,c)$, y por $F_{N,M,n,c'}(x)$ se notará a la función de distribución de una distribución de Pólya de parámetros $P(N,M,n,c')$. En el caso en que $c' < 0$ se mantendrá la condición: $(-c')(n-1) < \min\{M, N-M\}$, con lo que el rango será $[0, n]$ en ambos casos.

Teorema 4.8

Si $c > c'$ y $n \geq 2$, se verifica que:

- $F_{N,M,n,c}(0) > F_{N,M,n,c'}(0)$
- $F_{N,M,n,c}(n-1) < F_{N,M,n,c'}(n-1)$

Demostración

a) Puesto que

$$F_{N,M,n,c}(0) = P[X \leq 0] = P[X=0],$$

basta probar que

$$\frac{(N-M)^{(n,c)}}{N^{(n,c)}} > \frac{(N-M)^{(n,c')}}{N^{(n,c)'}}$$

lo que se va a hacer probando que la desigualdad se cumple para cada término del producto que define la expresión $A^{(n,c)}$, ya que por los requerimientos exigidos a los parámetros, está garantizado que todos los factores son positivos.

Debemos probar, por tanto, que

$$\frac{N-M+tc}{N+tc} > \frac{N-M+tc'}{N+tc'} \quad , \quad t=1,2,\dots,n-1$$

(para $t=0$ se da la igualdad, con lo que la demostración no se ve afectada)

En efecto,

$c > c' \Rightarrow c - c' > 0$, y por tanto, para $t: 1, 2, \dots, n-1$ se tiene que

$$\begin{aligned} Nt(c-c') &> (N-M)t(c-c') = Ntc + (N-M)tc' > Ntc' + (N-M)tc = \\ &= N(N-M) + (N-M)tc' + Ntc + t^2cc' > \\ &> N(N-M) + (N-M)tc + Ntc' + t^2cc' = \end{aligned}$$

$$\frac{N-M+tc}{N+tc} > \frac{N-M+tc'}{N+tc'}$$

b) Puesto que

$$F_{N,M,n,c}(n-1) = P\{X \leq n-1\} = 1 - P\{X=n\},$$

de manera análoga al apartado anterior se prueba que

$$\frac{M^{(n,c)}}{N^{(n,c)}} > \frac{M^{(n,c')}}{N^{(n,c)'}}$$

y de esta forma se tiene que

$$F_{N,M,n,c}(n-1) = 1 - \frac{M^{(n,c)}}{N^{(n,c)}} < 1 - \frac{M^{(n,c')}}{N^{(n,c)'}} = F_{N,M,n,c'}(n-1)$$

Corolario 4.9

En los casos particulares $c=1$, $c'=0$ y $c''=-1$, (distribuciones hipergeométrica negativa, binomial e hipergeométrica, respectivamente), se verifica para $n \geq 2$ que

- a) $F_{N,M,n,a}(0) > F_{N,M,n,b}(0) > F_{N,M,n,c}(0)$
- b) $F_{N,M,n,a}(n-1) < F_{N,M,n,b}(n-1) < F_{N,M,n,c}(n-1)$

Demostración

Es inmediata a partir de la aplicación del Teorema 4.8.

Teorema 4.10

La distribución de Pólya es simétrica para $p = q = 1/2$.

Demostración

Basta probar que para todo $k: 0, 1, \dots, n$, se verifica que

$$\sum_{i=0}^k \binom{n}{i} \frac{(pN)^{i,c} (N-pN)^{(n-i,c)}}{N^{(n,c)}} = \sum_{j=n-k}^n \binom{n}{j} \frac{(pN)^{j,c} (N-pN)^{(n-j,c)}}{N^{(n,c)}}$$

En efecto, si $p = 1/2$ la igualdad anterior puede ser escrita como

$$\sum_{i=0}^k \binom{n}{i} \frac{(pN)^{i,c} (pN)^{(n-i,c)}}{N^{(n,c)}} = \sum_{j=n-k}^n \binom{n}{j} \frac{(pN)^{j,c} (pN)^{(n-j,c)}}{N^{(n,c)}}$$

igualdad que se verifica trivialmente sin más que tener en cuenta que el sumando que se obtiene para i ($i: 0, 1, \dots, k$) en el primer sumatorio es igual al sumando que se obtiene para $j = n-i$ ($i: 0, 1, \dots, k$) en el segundo sumatorio.

■

Corolario 4.11

Para $p = 1/2$ y n impar, se verifica que

$$\sum_{i=0}^m \binom{n}{i} (pN)^{i,c} (N-pN)^{(n-i,c)} = \sum_{i=m+1}^n \binom{n}{i} (pN)^{i,c} (N-pN)^{(n-i,c)} = \frac{1}{2} N^{(n,c)}$$

Demostración

Se deduce de forma inmediata del Teorema 4.9.

■

CAPÍTULO V

UNA ACOTACIÓN PARA LA FUNCIÓN DE DISTRIBUCIÓN DE LA DISTRIBUCIÓN DE PÓLYA

CAPÍTULO V

UNA ACOTACIÓN PARA LA FUNCIÓN DE DISTRIBUCIÓN DE LA DISTRIBUCIÓN DE PÓLYA

En este capítulo se compara la función de distribución de la distribución de Pólya con parámetro reemplazamiento negativo con la función de distribución de la binomial aproximante (igual valor esperado y tamaño muestral), obteniéndose un resultado que permite la comparación de ambas funciones de distribución en todo punto salvo como máximo en un intervalo de amplitud menor que la unidad.

INTRODUCCIÓN Y DEFINICIONES PREVIAS

En este capítulo vamos a considerar exclusivamente distribuciones de Pólya $P(N, M, n, c)$ con parámetro reemplazamiento $c < 0$ y rango completo $[0, n]$, condición esta última que queda garantizada con la exigencia

$$(-c)(n-1) < \min\{M; N-M\}.$$

El objetivo de este capítulo es llegar a poder comparar la función de distribución de la distribución $P(N, p, n, c)$ [$p = M/N$] con la de la distribución binomial $B(n, p)$. El resultado fundamental, Teorema 5.19, permite la comparación de ambas funciones de distribución en todo \mathbb{R} salvo en un intervalo de amplitud menor que la unidad. El interés de este resultado radica en que mejora el obtenido por Ollero y Ramos (1995) comentado en el Capítulo III de esta memoria, ya que el intervalo de indeterminación de la comparación obtenido por estos autores contiene estrictamente al que nosotros obtenemos.

Para alcanzar nuestro objetivo vamos a utilizar la metodología sugerida en Uhlmann (1966) en un estudio acerca de la distribución hipergeométrica. En este sentido vamos a utilizar dos funciones auxiliares $G_{N, x, n, c}(p)$ y $H_{N, x, n, c}(p)$ relacionadas con la función de distribución $F_{N, p, n, c}(x)$ de la distribución de Pólya $P(N, p, n, c)$ que definimos a continuación

Recordemos en primer lugar que

$$F_{N,M,n,c}(x) = P[X \leq x] = \begin{cases} 0 & ; x < 0 \\ \sum_{m=0}^k \binom{n}{m} \frac{(\rho N)^{(m,c)}(qN)^{(n-m,c)}}{N^{(n,c)}} & ; k \leq x < k+1 \quad (k:0,1,\dots,n) \\ 1 & ; x > n \end{cases}$$

Definamos ahora

$$G_{N,x,n,c}(\rho) = \begin{cases} 0 & ; 1 + \frac{c(n-x-1)}{N} < \rho \leq 1 \\ \sum_{m=0}^x \binom{n}{m} \frac{(\rho N)^{(m,c)}(qN)^{(n-m,c)}}{N^{(n,c)}} & ; \frac{-cx}{N} \leq \rho \leq 1 + \frac{c(n-x-1)}{N} \\ 1 & ; 0 \leq \rho < \frac{-cx}{N} \end{cases}$$

Consideremos, por otra parte, la función auxiliar

$$H_{N,x,n,c}(\rho) = \sum_{m=0}^x \binom{n}{m} \frac{(\rho N)^{(m,c)}(qN)^{(n-m,c)}}{N^{(n,c)}} \quad ; \quad 0 \leq \rho \leq 1 \quad [5.1]$$

que coincide con $G_{N,x,n,c}(\rho)$ en el intervalo: $\left[\frac{-cx}{N} ; 1 + \frac{c(n-x-1)}{N} \right]$.

Si consideramos fijos los valores N , n y c , se tiene que, para todo $x:0,1,\dots,n$, y para todo valor de ρ compatible con el modelo de urna asociado a la distribución de Pólya con parámetro reemplazamiento negativo,

$$G_{N,x,n,c}(\rho) = H_{N,x,n,c}(\rho) = F_{N,\rho,n,c}(x)$$

RESULTADOS PREVIOS

Teorema 5.1

Para valores $x:0,1,\dots,n-1$, se verifica que

$$H_{N,x,n,c}(p) - H_{N,x,n,c}\left(p + \frac{c}{N}\right) = \binom{n}{x} (n-x)c \frac{(pN+c)^{(x,c)}(qN)^{(n-x-1,c)}}{N^{(n,c)}} \quad [5.2]$$

Demostración

La demostración es similar a la empleada en el Teorema 4.1. Se procederá a probar por inducción.

Para $x=0$:

$$\begin{aligned} H_{N,0,n,c}(p) - H_{N,0,n,c}\left(p + \frac{c}{N}\right) &= \\ &= \frac{1}{N^{(n,c)}} [(pN)^{(0,c)}(N-pN)^{(n,c)} - (pN+c)^{(0,c)}(N-pN-c)^{(n,c)}] = \\ &= \frac{(N-pN)^{(n,c)}}{N^{(n,c)}} \frac{nc}{N-pN+(n-1)c} = nc \frac{(N-pN)^{(n-1,c)}}{N^{(n,c)}} \end{aligned}$$

Consideremos ahora el resultado válido para $x-1$ y probémoslo para x ($x=1,2,\dots,n-1$). En efecto, de acuerdo con la expresión [5.1],

$$\begin{aligned} H_{N,x,n,c}(p) - H_{N,x,n,c}\left(p + \frac{c}{N}\right) &= H_{N,x-1,n,c}(p) - H_{N,x-1,n,c}\left(p + \frac{c}{N}\right) + \\ &+ \binom{n}{x} \frac{(pN)^{(x,c)}(qN)^{(n-x,c)}}{N^{(n,c)}} - \binom{n}{x} \frac{(pN+c)^{(x,c)}(qN-c)^{(n-x,c)}}{N^{(n,c)}} \end{aligned}$$

y, supuesto que la expresión [5.2] se verifica para $x-1$, tendremos que

$$\begin{aligned} H_{N,x,n,c}(p) - H_{N,x-1,n,c}\left(p + \frac{c}{N}\right) &= \\ &= \binom{n}{x-1} (n-x+1)c \frac{(pN+c)^{(x-1,c)}(N-pN)^{(n-x,c)}}{N^{(n,c)}} + \binom{n}{x} \frac{(pN)^{(x,c)}(N-pN)^{(n-x,c)}}{N^{(n,c)}} \\ &\quad - \binom{n}{x} \frac{(pN+c)^{(x,c)}(N-pN-c)^{(n-x,c)}}{N^{(n,c)}} \end{aligned}$$

Esta expresión puede ser escrita como la suma de tres términos B, C y D respectivamente, en la forma B+C-D donde

$$C - D = \binom{n}{x} \frac{1}{N^{(n,c)}} \left((pN)^{(x,c)}(N-pN)^{(n-x,c)} - (pN+c)^{(x,c)}(N-pN-c)^{(n-x,c)} \right)$$

Como

$$\begin{aligned} (pN)^{(x,c)} &= (pN+c)^{(x,c)} \frac{pN}{pN+xc} \\ (N-pN)^{(n-x,c)} &= (N-pN-c)^{(n-x,c)} \frac{N-pN+(n-x-1)c}{N-pN-c} \end{aligned}$$

el término C-D queda:

$$C - D = \binom{n}{x} \frac{(pN+c)^{(x,c)}(N-pN)^{(n-x,c)}}{N^{(n,c)}} \left[\frac{pN}{pN+xc} - \frac{N-pN-c}{N-pN+(n-x-1)c} \right]$$

Operando el término entre corchetes, y después de simplificar, queda:

$$C - D = \binom{n}{x} \frac{(pN+c)^{(x-1,c)}(N-pN)^{(n-x-1,c)}}{N^{(n,c)}} (pN(n-x)c - xc(N-pN-c))$$

Antes de sumar el término B, tendremos en cuenta que

$$\binom{n}{x-1} (n-x+1) = \binom{n}{x} x,$$

por lo que el resultado final de B+C-D es el siguiente

$$\begin{aligned} B+C-D &= \binom{n}{x} x c \frac{(pN+c)^{(x-1,c)} (qN)^{(n-x,c)}}{N^{(n,c)}} + \binom{n}{x} c \frac{(pN+c)^{(x-1,c)} (qN)^{(n-x-1,c)}}{N^{(n,c)}} [pN(n-x) - x(qN-c)] \\ &= \binom{n}{x} c \frac{(pN+c)^{(x-1,c)}}{N^{(n,c)}} (x(qN)^{(n-x,c)} + (qN)^{(n-x-1,c)} [pN(n-x) - x(qN-c)]) \end{aligned}$$

Esta expresión puede ser simplificada ya que

$$\begin{aligned} x(qN)^{(n-x,c)} + (qN)^{(n-x-1,c)} [pN(n-x) - x(qN-c)] &= \\ = x(qN)^{(n-x-1,c)} (qN + (n-x-1)c) + (qN)^{(n-x-1,c)} [pN(n-x) - x(qN-c)] &= \\ = (qN)^{(n-x-1,c)} [x(qN + (n-x-1)c) + pN(n-x) - x(qN-c)] &= (qN)^{(n-x-1,c)} (n-x)(pN + xc). \end{aligned}$$

En consecuencia,

$$\begin{aligned} H_{N,x,n,c}(p) - H_{N,x,n,c}\left(p + \frac{c}{N}\right) &= B+C-D = \\ &= \binom{n}{x} c \frac{(pN+c)^{(x-1,c)} (qN)^{(n-x-1,c)} (n-x)(pN+xc)}{N^{(n,c)}} = \binom{n}{x} (n-x) c \frac{(pN+c)^{(x,c)} (qN)^{(n-x-1,c)}}{N^{(n,c)}} \end{aligned}$$

quedando probado el teorema. ■

El resultado que se ha presentado en este Teorema 5.1 tiene un homólogo en el Teorema 4.1 donde se obtuvo la expresión de la primera diferencia para el caso de la función de distribución de una distribución de Pólya $P(N, M, n, c)$ pero no se impuso la condición de negatividad en el parámetro c .

Corolario 5.2

Para $x: 0, 1, \dots, n-1$, se verifica que:

$$H_{N,x,n,c}(p) - H_{N,x,n,c}\left(p - \frac{c}{N}\right) = -\binom{n}{x} (n-x)c \frac{(pN)^{(x,c)}(qN+c)^{(n-x-1,c)}}{N^{(n,c)}}$$

Demostración

Este resultado se obtiene teniendo en cuenta que

$$H_{N,x,n,c}(p) - H_{N,x,n,c}\left(p - \frac{c}{N}\right) = -\left(H_{N,x,n,c}\left(p - \frac{c}{N}\right) - H_{N,x,n,c}(p)\right)$$

y aplicando la expresión [5.2] sin más que considerar que

$$\left(p - \frac{c}{N}\right)N + c = pN - c + c = pN$$

$$\left(q + \frac{c}{N}\right)N = qN + c$$

■

Corolario 5.3

En el caso particular de la distribución hipergeométrica ($c = -1$), se verifica para $x: 0, 1, \dots, n-1$ que

$$H_{N,x,n,-1}(p) - H_{N,x,n,-1}\left(p + \frac{1}{N}\right) = \frac{\binom{pN}{x} \binom{qN-1}{n-x-1}}{\binom{N}{n}}$$

Demostación

A este resultado, obtenido por Uhlmann (1966), se llega aquí como una particularización del Corolario 5.2 para el caso en que $c = -1$. En efecto,

$$H_{N,x,n,-1}(p) - H_{N,x,n,-1}\left(p + \frac{1}{N}\right) = \binom{n}{x} (n-x) \frac{(pN)^{(x-1)} (qN-1)^{(n-x-1,-1)}}{N^{(n-1)}}$$

Puesto que

$$(pN)^{(x-1)} = x! \binom{pN}{x}$$

$$N^{(n-1)} = n! \binom{N}{n}$$

$$(qN-1)^{(n-x-1,-1)} = (n-x-1)! \binom{qN-1}{n-x-1}$$

se tiene que

$$H_{N,x,n,-1}(p) - H_{N,x,n,-1}\left(p + \frac{1}{N}\right) = \binom{n}{x} (n-x) \frac{\binom{pN}{x} \binom{qN-1}{n-x-1} x!(n-x-1)!}{\binom{N}{n} n!} = \frac{\binom{pN}{x} \binom{qN-1}{n-x-1}}{\binom{N}{n}}$$

como queríamos demostrar. ■

Teorema 5.4

$H_{N,x,n,c}(p)$ es decreciente en el intervalo $\left(\frac{-cx}{N} ; 1 + \frac{c(n-x-1)}{N} \right)$.

Demostración:

$H_{N,x,n,c}(p)$ es una función polinómica de grado n que verifica:

$$H_{N,x,n,c}(p) = 1 \text{ cuando } p = \frac{-ic}{N} ; i: 0, 1, \dots, x \quad [5.3]$$

$$H_{N,x,n,c}(p) = 0 \text{ cuando } p = 1 + \frac{c(n-x-i)}{N} ; i: 0, 1, \dots, n-x \quad [5.4]$$

Por otra parte, $\frac{d}{dp}H_{N,x,n,c}(p)$ es un polinomio en p de grado $n-1$ que tiene $n-1$ raíces distintas

$p'_1, p'_2, \dots, p'_{n-1}$.

Teniendo en cuenta las expresiones [5.3] y [5.4] y considerando el carácter de función polinómica de $H_{N,x,n,c}(p)$ se tiene, necesariamente, que

$$0 < p'_1 < \frac{-c}{N} < p'_2 < \frac{-2c}{N} < \dots < \frac{(-x+1)c}{N} < p'_x < \frac{-xc}{N}$$

$$1 + \frac{c(n-x-1)}{N} < p'_{x+1} < 1 + \frac{c(n-x+2)}{N} < \dots < 1 + \frac{c}{N} < p'_{n-1} < 1$$

es decir, las $n-1$ raíces se encuentran fuera del intervalo $\left(\frac{-cx}{N} ; 1 + \frac{c(n-x-1)}{N} \right)$, por lo que en

dicho intervalo $H_{N,x,n,c}(p)$ es estrictamente decreciente.

■

Corolario 5.5

Si $0 < x < n-1$, existe un punto $p_0 \in \left(\frac{-cx}{N} ; 1 + \frac{c(n-x-1)}{N} \right)$ tal que

$$\frac{d^2}{dp^2} H_{N,x,n,c}(p) \begin{cases} < 0 & \text{si } p'_x \leq p < p_0 \\ > 0 & \text{si } p_0 < p \leq p'_{x+1} \end{cases}$$

Demostación

Es inmediata a la vista de la distribución de las $n-1$ raíces $p'_1, p'_2, \dots, p'_{n-1}$ de la función polinómica $\frac{d}{dp} H_{N,x,n,c}(p)$, que $H_{N,x,n,c}(p)$ tiene un máximo relativo en p'_x y un mínimo relativo en p'_{x+1} . Consecuentemente, en el intervalo $[p'_x; p'_{x+1}]$ $H_{N,x,n,c}(p)$ comienza siendo cóncava y finaliza siendo convexa, existiendo un punto de inflexión $p_0 \in \left(\frac{-cx}{N} ; 1 + \frac{c(n-x-1)}{N} \right)$

Corolario 5.6

Para los valores $x=0$ y $x=n-1$ se tiene, respectivamente, que:

a) $H_{N,0,n,c}(p)$ es convexa en el intervalo $\left(0 ; 1 + \frac{c(n-1)}{N} \right)$

b) $H_{N,n-1,n,c}(p)$ es cóncava en el intervalo $\left(\frac{-c(n-1)}{N} ; 1 \right)$

Demostación:

Basta tener en cuenta el carácter polinómico de la función $H_{N,x,n,c}(p)$ y la distribución de las raíces p'_i ($i:1,2,\dots,n-1$) particularizada para los casos $x=0$ y $x=n-1$.

Teorema 5.7

Para valores de $x: 1, 2, \dots, n-2$, se verifica que la segunda diferencia toma la forma:

$$\begin{aligned}
 & H_{N,x,n,c}\left(p - \frac{c}{N}\right) - 2H_{N,x,n,c}(p) + H_{N,x,n,c}\left(p + \frac{c}{N}\right) = \\
 & = \binom{n}{x} \frac{(n-x)c}{N^{(n,c)}} (pN+c)^{(x,c)} (N-pN)^{(n-x-1,c)} \frac{Nc(p(n-1)-x)}{(pN+xc)(N-pN)}
 \end{aligned}$$

Demostración

La expresión:

$$H_{N,x,n,c}\left(p - \frac{c}{N}\right) - 2H_{N,x,n,c}(p) + H_{N,x,n,c}\left(p + \frac{c}{N}\right)$$

también puede escribirse como sigue:

$$\left(H_{N,x,n,c}\left(p - \frac{c}{N}\right) - H_{N,x,n,c}(p) \right) - \left(H_{N,x,n,c}(p) - H_{N,x,n,c}\left(p + \frac{c}{N}\right) \right)$$

y aplicando a cada una de estas diferencias el resultado presentado en el Teorema 5.1, se tiene que es igual a

$$\begin{aligned}
 & \binom{n}{x} \frac{(n-x)c}{N^{(n,c)}} \left((pN)^{(x,c)} (N-pN+c)^{(n-x-1,c)} - (pN+c)^{(x,c)} (N-pN)^{(n-x-1,c)} \right) = \\
 & = \binom{n}{x} \frac{(n-x)c}{N^{(n,c)}} (pN+c)^{(x,c)} (N-pN)^{(n-x-1,c)} \left(\frac{pN}{pN+xc} \frac{N-pN+(n-x-1)c}{N-pN} - 1 \right)
 \end{aligned}$$

ya que

$$\begin{aligned}
 (pN)^{(x,c)} &= (pN)(pN+c)(pN+2c)\dots(pN+(x-1)c) \\
 (pN+c)^{(x,c)} &= (pN+c)(pN+2c)(pN+3c)\dots(pN+xc)
 \end{aligned}$$

$$(N-pN)^{(n-x-1,c)} = (N-pN)(N-pN+c)(N-pN+2c)\dots(N-pN+(n-x-2)c)$$

$$(N-pN+c)^{(n-x-1,c)} = (N-pN+c)(N-pN+2c)(N-pN+3c)\dots(N-pN+(n-x-1)c).$$

Finalmente, puesto que

$$\frac{pN}{pN+xc} \frac{N-pN+(n-x-1)c}{N-pN} - 1 =$$

$$= \frac{pN(N-pN) + pN(n-x-1)c - pN(N-pN) - xc(N-pN)}{(pN+xc)(N-pN)} =$$

$$= \frac{Nc(p(n-x-1) - x(1-p))}{(pN+xc)(N-pN)} = \frac{Nc(p(n-1) - x)}{(pN+xc)(N-pN)}$$

tenemos que

$$H_{N,x,n,c}\left(p - \frac{c}{N}\right) - 2H_{N,x,n,c}(p) + H_{N,x,n,c}\left(p + \frac{c}{N}\right) =$$

$$= \binom{n}{x} \frac{(n-x)c}{N^{(n,c)}} (pN+c)^{(x,c)} (N-pN)^{(n-x-1,c)} \frac{Nc(p(n-1)-x)}{(pN+xc)(N-pN)}$$

como queríamos demostrar. ■

Observamos que esta segunda diferencia se anula para $p = \frac{x}{(n-1)}$ (basta sustituir en la expresión obtenida para comprobarlo).

El resultado que se ha presentado en este Teorema 5.7 tiene un equivalente en el Teorema 4.6, donde se obtuvo la expresión de la segunda diferencia para el caso de la función de distribución de una distribución de Pólya $P(N, M, n, c)$ pero no se impuso la condición de negatividad en el parámetro c .

La misma consideración se puede hacer respecto del Corolario 5.8 y del Corolario 4.7.

Corolario 5.8

La segunda diferencia de la función correspondiente al valor $c = -1$ (distribución hipergeométrica) verifica la siguiente expresión:

$$\begin{aligned}
 &H_{N,x,n,-1}\left(p - \frac{1}{N}\right) - 2H_{N,x,n,-1}(p) + H_{N,x,n,-1}\left(p + \frac{1}{N}\right) = \\
 &= \frac{\binom{pN-1}{x} \binom{qN}{n-x-1}}{\binom{N}{n}} - \frac{\binom{pN}{x} \binom{qN-1}{n-x-1}}{\binom{N}{n}}.
 \end{aligned}$$

Demostración

Se trata de una particularización del Teorema 5.7 al caso en que $c = -1$. Uhlmann (1966) calculó la expresión correspondiente a dicha segunda diferencia en el caso de la distribución hipergeométrica, siendo este resultado equivalente al suyo.



Corolario 5.9

Para $x: 1, 2, \dots, n-1$, $H_{N,x,n,c}(p)$ es cóncava en el intervalo $\left(\frac{(-x+1)c}{N}; \frac{x}{n-1}\right)$ y convexa en el intervalo $\left(\frac{x}{n-1}; 1 + \frac{c(n-x-2)}{N}\right)$.

Demostración

Por el Teorema 5.7 sabemos que para $x: 1, 2, \dots, n-2$,

$$\begin{aligned}
 &H_{N,x,n,c}\left(p - \frac{c}{N}\right) - 2H_{N,x,n,c}(p) + H_{N,x,n,c}\left(p + \frac{c}{N}\right) = \\
 &= \binom{n}{x} \frac{(n-x)c}{N^{(n,c)}} (pN+c)^{(x,c)} (N-pN)^{(n-x-1,c)} \frac{Nc(p(n-1)-x)}{(pN+xc)(N-pN)}.
 \end{aligned}$$

De aquí se sigue que la segunda diferencia se anula sólo para $p = \frac{x}{(n-1)}$. Para este valor tenemos que:

$$H_{N,x,n,c}\left(\frac{x}{n-1} - \frac{c}{N}\right) - 2H_{N,x,n,c}\left(\frac{x}{n-1}\right) + H_{N,x,n,c}\left(\frac{x}{n-1} + \frac{c}{N}\right) = 0$$

lo cual indica que los puntos de abscisa $\frac{x}{n-1} + \frac{c}{N}$; $\frac{x}{n-1}$; $\frac{x}{n-1} - \frac{c}{N}$ están alineados

En consecuencia, la localización del punto de inflexión p_0 señalada en el Corolario 5.5 queda precisada: $p_0 = \frac{x}{(n-1)}$.

Por tanto, y de acuerdo con el Corolario 5.5, $H_{N,x,n,c}(p)$ será cóncava para todo p tal que $\frac{(-x+1)c}{N} \leq p \leq \frac{x}{n-1}$ y convexa para todo p tal que $\frac{x}{n-1} \leq p \leq 1 - \frac{c(n-x-2)}{N}$

Lema 5.10

Para todo ε real tal que $0 < \varepsilon < 1$ y para todo $x = 1, 2, \dots, n-2$, se verifica:

$$(i) H_{N,x,n,c}\left(p + \frac{\varepsilon c}{N}\right) - H_{N,x,n,c}(p) > \varepsilon \left\{ H_{N,x,n,c}\left(p + \frac{c}{N}\right) - H_{N,x,n,c}(p) \right\}$$

para todo p tal que $\frac{-xc}{N} \leq p \leq \frac{x}{n-1}$.

$$(ii) H_{N,x,n,c}\left(p + \frac{\varepsilon c}{N}\right) - H_{N,x,n,c}(p) < \varepsilon \left\{ H_{N,x,n,c}\left(p + \frac{c}{N}\right) - H_{N,x,n,c}(p) \right\}$$

para todo p tal que $\frac{x}{n-1} - \frac{c}{N} \leq p \leq 1 - \frac{c(n-x-2)}{N}$.

Demostración

(i) Por el Corolario 5.9 sabemos que $H_{N,x,n,c}(p)$ es cóncava para todo p tal que

$$\frac{-xc}{N} + \frac{c}{N} \leq p \leq \frac{x}{n-1}$$

Consideremos un punto p tal que $\frac{-xc}{N} \leq p \leq \frac{x}{n-1}$. Es obvio, siendo $c < 0$, que $p + \frac{c}{N}$ sigue perteneciendo al intervalo de concavidad de $H_{N,x,n,c}(p)$. En consecuencia,

$$\begin{aligned} H_{N,x,n,c} \left(\varepsilon \left(p + \frac{c}{N} \right) + (1-\varepsilon)p \right) &= H_{N,x,n,c} \left(p + \frac{\varepsilon c}{N} \right) > \varepsilon H_{N,x,n,c} \left(p + \frac{c}{N} \right) + (1-\varepsilon)H_{N,x,n,c}(p) = \\ &= \varepsilon \left\{ H_{N,x,n,c} \left(p + \frac{c}{N} \right) - H_{N,x,n,c}(p) \right\} + H_{N,x,n,c}(p). \end{aligned}$$

De aquí,

$$\varepsilon \left\{ H_{N,x,n,c} \left(p + \frac{c}{N} \right) - H_{N,x,n,c}(p) \right\} < H_{N,x,n,c} \left(p + \frac{\varepsilon c}{N} \right) - H_{N,x,n,c}(p).$$

(ii) Por el Corolario 5.9 sabemos que $H_{N,x,n,c}(p)$ es convexa para todo p tal que

$$\frac{x}{n-1} \leq p \leq 1 + \frac{c(n-x-2)}{N}$$

Consideremos un punto p tal que $\frac{x}{n-1} - \frac{c}{N} \leq p \leq 1 + \frac{c(n-x-2)}{N}$. Por ser $c < 0$ resulta claro que

$p + \frac{c}{N}$ sigue perteneciendo al intervalo de convexidad de $H_{N,x,n,c}(p)$.

En consecuencia, razonando en términos similares a lo efectuado en (i), quedaría probado (ii). ■

Lema 5.11

Para $0 < x < n-1$, las funciones las funciones $H_{N,x,n,c}(\rho)$ y $H_{N-c,x,n,c}(\rho)$ tienen un único punto de corte en el intervalo $\left(\frac{-cx}{N} ; 1 + \frac{c(n-x-1)}{N} \right)$.

Demostración

Aplicando los resultados conocidos, expresiones [5.3] y [5.4] y Teorema 5.4, de la función $H_{N,x,n,c}(\rho)$ a la función $H_{N-c,x,n,c}(\rho)$, tendremos que

$$H_{N-c,x,n,c}(\rho) = 1 \text{ cuando } \rho = \frac{-ic}{N-c} ; i=0,1, \dots, x \quad [5.5]$$

$$H_{N-c,x,n,c}(\rho) = 0 \text{ cuando } \rho = 1 + \frac{c(n-x-i)}{N-c} ; i=0,1, \dots, n-x \quad [5.6]$$

siendo decreciente en el intervalo $\left(\frac{-cx}{N-c} ; 1 + \frac{c(n-x-1)}{N-c} \right)$ con un único punto de inflexión ρ^*_0 en el intervalo tal que en ρ^*_0 la función pasa de ser cóncava a convexa

Por una parte, al ser $c < 0$ y $0 < x < n-1$, se tiene que

$$\frac{-cx}{N-c} < \frac{-cx}{N} < 1 + \frac{c(n-x-1)}{N} < 1 + \frac{c(n-x-1)}{N-c} \quad [5.7]$$

Por otra parte, de las expresiones [5.3], [5.4] y [5.5], [5.6], se sigue que

$$\frac{-ic}{N-c} < \frac{-ic}{N} < \frac{-(i+1)c}{N-c} ; i=1,2, \dots, x-1$$

$$1 + \frac{c(n-x-i)}{N} < 1 + \frac{c(n-x-i)}{N-c} < 1 + \frac{c(n-x-i-1)}{N-c} ; i=1,2, \dots, n-x-1$$

como puede comprobarse fácilmente.

En consecuencia, y teniendo en cuenta la expresión obtenida [5.7], ambas funciones $H_{N,x,n,c}(p)$ y $H_{N-c,x,n,c}(p)$ son decrecientes en el intervalo $\left(\frac{-cx}{N} ; 1 + \frac{c(n-x-1)}{N}\right)$, verificándose:

$$H_{N,x,n,c}\left(\frac{-cx}{N}\right) = 1 > H_{N-c,x,n,c}\left(\frac{-cx}{N}\right)$$

$$H_{N,x,n,c}\left(1 + \frac{c(n-x-1)}{N}\right) = 0 < H_{N-c,x,n,c}\left(1 + \frac{c(n-x-1)}{N}\right)$$

Por todo ello, atendiendo a la continuidad de la función H , $H_{N,x,n,c}(p)$ y $H_{N-c,x,n,c}(p)$ tendrán un único punto de corte en el intervalo $\left(\frac{-cx}{N} ; 1 + \frac{c(n-x-1)}{N}\right)$ como queríamos probar

Teorema 5.12

Si $x=0$, entonces,

$$G_{N,0,n,c}(p) - G_{N-c,0,n,c}(p) = \begin{cases} 0 & ; p=0 \\ <0 & ; 0 < p < 1 + \frac{c(n-1)}{(N-c)} \\ 0 & ; p > 1 + \frac{c(n-1)}{(N-c)} \end{cases}$$

y si $x=n-1$, entonces,

$$G_{N,n-1,n,c}(p) - G_{N-c,n-1,n,c}(p) = \begin{cases} 0 & ; p \leq \frac{-c(n-1)}{(N-c)} \\ >0 & ; \frac{-c(n-1)}{(N-c)} < p < 1 \\ 0 & ; p=1 \end{cases}$$

Demostración

Por la definición dada de las funciones $G_{N,x,n,c}(p)$ y $H_{N,x,n,c}(p)$, podemos escribir:

$$G_{N,x,n,c}(p) = \begin{cases} 0 & ; \quad ; -\frac{c(n-x-1)}{N} < p \leq 1 \\ H_{N,x,n,c}(p) & ; \quad ; -\frac{cx}{N} \leq p \leq 1 + \frac{c(n-x-1)}{N} \\ 1 & ; \quad ; 0 \leq p < -\frac{cx}{N} \end{cases}$$

En particular, para $x=0$,

$$G_{N,0,n,c}(p) = \begin{cases} 1 & ; \quad ; p = 0 \\ H_{N,0,n,c}(p) & ; \quad ; 0 < p \leq 1 + \frac{c(n-1)}{N} \\ 0 & ; \quad ; p > 1 + \frac{c(n-1)}{N} \end{cases}$$

$$G_{N-c,0,n,c}(p) = \begin{cases} 1 & ; \quad ; p = 0 \\ H_{N-c,0,n,c}(p) & ; \quad ; 0 < p \leq 1 + \frac{c(n-1)}{N-c} \\ 0 & ; \quad ; p > 1 + \frac{c(n-1)}{N-c} \end{cases}$$

Por el Corolario 5.6 sabemos que $H_{N,0,n,c}(p)$ es decreciente y convexa en el intervalo $\left(0; 1 + \frac{c(n-1)}{N}\right)$ y $H_{N-c,0,n,c}(p)$ es decreciente y convexa en el intervalo $\left(0; 1 + \frac{c(n-1)}{N-c}\right)$. Por otra parte, por ser $c < 0$,

$$1 + \frac{c(n-1)}{N} < 1 + \frac{c(n-1)}{N-c}$$

Entonces, se verifica necesariamente que, para valores de p pertenecientes a $\left(0; 1 + \frac{c(n-1)}{N-c}\right)$,

$$G_{N,0,n,c}(p) - G_{N-c,0,n,c}(p) < 0$$

completándose el resultado de forma trivial.

La demostración del resultado para el caso $x=n-1$ se haría de forma análoga siguiendo un planteamiento dual.

Lema 5.13

Si $0 \leq x < n \leq N$, para todo $0 \leq p \leq 1$, se verifica que

$$\begin{aligned} H_{N,x,n,c} \left(p - \frac{cp}{N} \right) - H_{N-c,x,n,c}(p) &= \\ &= \frac{c}{N-c} \sum_{i=0}^x \frac{\binom{n}{i} (p(N-c))^{(i,c)} (q(N-c)+c)^{(n-i,c)} ((c-N)i + pn(N-c))}{N^{(n,c)}} \end{aligned}$$

Demostración

A partir de la definición de $H_{N,x,n,c}(p)$ podemos escribir

$$\begin{aligned} H_{N,x,n,c} \left(p - \frac{cp}{N} \right) - H_{N-c,x,n,c}(p) &= \\ &= \sum_{i=0}^x \binom{n}{i} \frac{(p(N-c))^{(i,c)} (qN+pc)^{(n-i,c)}}{N^{(n,c)}} - \sum_{i=0}^x \binom{n}{i} \frac{(p(N-c))^{(i,c)} (q(N-c))^{(n-i,c)}}{(N-c)^{(n,c)}} \end{aligned}$$

sin más que tener en cuenta que

$$\left(p - \frac{cp}{N}\right)N = Np - cp = p(N-c)$$

$$\left(1 - p + \frac{cp}{N}\right)N = N - Np + cp = Nq + pc$$

Por otra parte, puesto que

$$\frac{N^{(n,c)}}{(N-c)^{(n,c)}} = \frac{N(N+c)\dots(N+(n-1)c)}{(N-c)N\dots(N+(n-1)c-c)} = \frac{N+(n-1)c}{N-c}$$

entonces,

$$\begin{aligned} & H_{N,x,n,c} \left(p - \frac{cp}{N}\right) - H_{N-c,x,n,c}(p) = \\ & = \sum_{i=0}^x \binom{n}{i} \frac{(p(N-c))^{(i,c)}}{N^{(n,c)}} \left((qN+pc)^{(n-i,c)} - \frac{N+(n-1)c}{N-c} (q(N-c))^{(n-i,c)} \right) \end{aligned} \quad [5.8]$$

Realizamos a continuación algunas operaciones en el segundo miembro de la expresión [5.8]. En primer lugar tenemos que

$$(qN+pc) = (qN+(1-q)c) = qN - qc + c = q(N-c) + c.$$

De aquí,

$$\frac{(q(N-c))^{(n-i,c)}}{(qN+pc)^{(n-i,c)}} = \frac{(q(N-c))^{(n-i,c)}}{(q(N-c)+c)^{(n-i,c)}} =$$

$$= \frac{(q(N-c))(q(N-c)+c)\dots(q(N-c)+(n-i-1)c)}{(q(N-c)+c)(q(N-c)+2c)\dots(q(N-c)+(n-i)c)} = \frac{q(N-c)}{(q(N-c)+(n-i)c)}$$

Teniendo en cuenta lo anterior, el segundo miembro de la expresión [5.8] puede ser escrito como

$$\sum_{i=0}^x \binom{n}{i} \frac{(p(N-c))^{(i,c)}(qN+pc)^{(n-i,c)}}{N^{(n,c)}} \left(1 - \frac{(N-(n-1)c)}{(N-c)} \frac{(q(N-c))}{(q(N-c)+(n-i)c)} \right) \quad [5.9]$$

Desarrollando el último paréntesis de la expresión [5.9] tenemos que

$$1 - \frac{(N-(n-1)c)}{(N-c)} \frac{(q(N-c))}{(q(N-c)+(n-i)c)} = \frac{(N-c)(q(N-c)+(n-i)c) - (N-(n-1)c)(q(N-c))}{(N-c)(q(N-c)+(n-i)c)}$$

cuyo numerador puede reducirse ya que

$$\begin{aligned} & (N-c)(q(N-c)) + (N-c)(n-i)c - Nq(N-c) - (n-1)cq(N-c) = \\ & = -cq(N-c) + (N-c)(n-i)c - (n-1)cq(N-c) = \\ & = -cq(N-c) + N(n-i)c - (n-i)c^2 - (n-1)cq(N-c) = \\ & = c[-q(N-c) + N(n-i) - (n-i)c - (n-1)q(N-c)] = \\ & = c[N(n-i) - (n-i)c - nq(N-c)] = \\ & = c[Nn - Ni - nc + ic - n(1-p)(N-c)] = \\ & = c[Nn - Ni - nc + ic - Nn + nc + npN - npc] = \\ & = c[-Ni + ic + pn(N-c)] = c[(c-N)i + pn(N-c)] \end{aligned}$$

Sustituyendo en [5.9] la expresión obtenida, aquella puede volver a escribirse como

$$\sum_{i=0}^x \binom{n}{i} \frac{(p(N-c))^{(i,c)}(qN+pc)^{(n-i,c)}}{N^{(n,c)}} \frac{c}{(N-c)} \frac{((c-N)i + pn(N-c))}{(q(N-c)+(n-i)c)} \quad [5.10]$$

Finalmente, esta última expresión puede ser simplificada teniendo en cuenta que

$$\frac{(qN+pc)^{(n-i,c)}}{(q(N-c)+(n-i)c)} = \frac{(q(N-c)+c)^{(n-i,c)}}{q(N-c)+(n-i)c} = (q(N-c)+c)^{(n-i-1,c)}$$

por lo que la expresión [5.10] puede ser escrita de nuevo como

$$\frac{c}{(N-c)} \sum_{i=0}^x \binom{n}{i} \frac{(p(N-c))^{(i,c)}(q(N-c)+c)^{(n-i-1,c)}((c-N)i+pn(N-c))}{N^{(n,c)}}$$

quedando de esta manera probado el Lema. ■

Teorema 5.14

Si $0 \leq x < n \leq N$, para todo $0 \leq p \leq 1$, se verifica que

$$\begin{aligned} H_{N,x,n,c}(p) - H_{N-c,x,n,c}(p) &= \\ &= H_{N,x,n,c}(p) - H_{N,x,n,c}\left(p - \frac{cp}{N}\right) + \binom{n}{x+1} \frac{(x+1)c(p(N-c))^{(x-1,c)}(q(N-c)+c)^{(n-x-1,c)}}{(N-c)N^{(n,c)}} \end{aligned}$$

Demostración

Teniendo en cuenta el Lema 5.13, será suficiente con probar que:

$$\frac{c}{(N-c)} \sum_{i=0}^x \binom{n}{i} \frac{(p(N-c))^{(i,c)}(q(N-c)+c)^{(n-i-1,c)}((c-N)i+pn(N-c))}{N^{(n,c)}} = \tag{5.11}$$

$$= \binom{n}{x+1} \frac{(x+1)c(p(N-c))^{(x-1,c)}(q(N-c)+c)^{(n-x-1,c)}}{(N-c)N^{(n,c)}} \tag{5.12}$$

Lo probaremos por inducción.

En efecto, para $x=0$ basta sustituir este valor en [5.11] y en [5.12] para comprobar la igualdad:

$$\frac{c}{(N-c)} \binom{n}{0} \frac{(p(N-c))^{(0,c)}(q(N-c)+c)^{(n-1,c)}pn(N-c)}{N^{(n,c)}} = \frac{nc}{(N-c)} \frac{(p(N-c))^{(1,c)}(q(N-c)+c)^{(n-1,c)}}{N^{(n,c)}}$$

Supuesto cierto para $x-1$, demostrémoslo para x . En efecto, descomponiendo la expresión [5.11] se tiene que

$$\begin{aligned} & \frac{c}{(N-c)} \sum_{i=0}^x \binom{n}{i} \frac{(p(N-c))^{(i,c)}(q(N-c)+c)^{(n-i-1,c)}[(c-N)i+pn(N-c)]}{N^{(n,c)}} = \\ & = \frac{c}{(N-c)} \binom{n}{x} \frac{(p(N-c))^{(x,c)}(q(N-c)+c)^{(n-x,c)}}{N^{(n,c)}} + \\ & + \frac{c}{(N-c)} \binom{n}{x} \frac{(p(N-c))^{(x,c)}(q(N-c)+c)^{(n-x-1,c)}[(c-N)x+pn(N-c)]}{N^{(n,c)}} = \\ & = \frac{c}{(N-c)} \binom{n}{x} \frac{(p(N-c))^{(x,c)}(q(N-c)+c)^{(n-x-1,c)}[x(q(N-c)+cn-x)+(c-N)x+pn(N-c)]}{N^{(n,c)}} \quad [5.13] \end{aligned}$$

obteniéndose la última igualdad sin más que tener en cuenta que

$$(q(N-c)+c)^{(n-x,c)} = (q(N-c)+c)^{(n-x-1,c)} (qN-c+nc-xc)$$

en [5.13], con lo que el factor multiplicador se reduce a

$$x(q(N-c)+c(n-x))+(c-N)x+pn(N-c) =$$

$$\begin{aligned}
 &= x(N-c) - xp(N-c) + xc(n-x) - x(N-c) + pn(N-c) = \\
 &= xc(n-x) + p(N-c)(n-x) = (n-x)(xc + p(N-c))
 \end{aligned}$$

y la expresión [5.13] puede escribirse ahora como

$$\begin{aligned}
 &\frac{c}{(N-c)} \binom{n}{x} \frac{(p(N-c))^{(x,c)}(q(N-c)+c)^{(n-x-1,c)}(n-x)(p(N-c)+xc)}{N^{(n,c)}} = \\
 &= \frac{c}{(N-c)} \binom{n}{x} (n-x) \frac{(p(N-c))^{(x-1,c)}(q(N-c)+c)^{(n-x-1,c)}}{N^{(n,c)}}
 \end{aligned}$$

Teniendo en cuenta, por último, que

$$\binom{n}{x} (n-x) = \binom{n}{x+1} (x+1) .$$

la expresión [5.13] puede ser finalmente escrita como

$$\frac{c}{N-c} \binom{n}{x+1} (x+1) \frac{(p(N-c))^{(x+1,c)}(q(N-c)+c)^{(n-x-1,c)}}{N^{(n,c)}} .$$

quedando, de esta manera, probado el teorema. ■

Corolario 5.15

Si $0 \leq x < n \leq N$, para todo $0 \leq p \leq 1$, se verifica que

$$\begin{aligned}
 &H_{N,x,n-1}(p) - H_{N-1,x,n-1}(p) = \\
 &= H_{N,x,n-1}(p) - H_{N,x,n-1}\left(p + \frac{p}{N}\right) - \frac{p \binom{p(N+1)-1}{x} \binom{q(N+1)-1}{n-x-1}}{\binom{N}{n}}
 \end{aligned}$$

Demostración

Particularizando el resultado del Teorema 5.14 para el caso $c = -1$ (distribución hipergeométrica) se tiene que

$$\begin{aligned} & H_{N,x,n,-1}(\rho) - H_{N+1,x,n,-1}(\rho) = \\ & = H_{N,x,n,-1}(\rho) - H_{N,x,n,-1}\left(\rho + \frac{\rho}{N}\right) + \binom{n}{x+1} (x+1)(-1) \frac{[\rho(N+1)]^{(x+1,-1)} [q(N+1)-1]^{(n-x-1,-1)}}{(N+1)N^{(n,-1)}} \end{aligned}$$

Teniendo en cuenta que

$$A^{(b,-1)} = \frac{A!}{(A-b)!} = b! \binom{A}{b}$$

podemos escribir

$$[\rho(N+1)]^{(x+1,-1)} = \rho(N+1)[\rho(N+1)-1]^{(x,-1)} = \rho(N+1)x! \binom{\rho(N+1)-1}{x}$$

$$[q(N+1)-1]^{(n-x-1,-1)} = (n-x-1)! \binom{q(N+1)-1}{n-x-1}$$

$$N^{(n,-1)} = n! \binom{N}{n}$$

Y, por tanto,

$$\binom{n}{x+1} (x+1)(-1) \frac{[\rho(N+1)]^{(x+1,-1)} [q(N+1)-1]^{(n-x-1,-1)}}{(N+1)N^{(n,-1)}} =$$

$$\begin{aligned}
 &= - \binom{n}{x+1} \frac{(x+1)x!(n-x-1)!}{n!} p^{(N+1)} \frac{\binom{(p(N+1)-1)}{x} \binom{(q(N+1)-1)}{n-x-1}}{(N+1) \binom{N}{n}} = \\
 &= - \frac{p \binom{(p(N+1)-1)}{x} \binom{(q(N+1)-1)}{n-x-1}}{\binom{N}{n}}
 \end{aligned}$$

quedando probado, de esta forma, el corolario. ■

Este resultado coincide con el obtenido por Uhlmann (1966) en su estudio de la distribución hipergeométrica.

Teorema 5.16

Si $0 < x < n-1 < N$, se verifica:

$$G_{N,x,n,c}(p) - G_{N-c,x,n,c}(p) = \begin{cases} 0 & ; \quad 0 \leq p \leq \frac{x(-c)}{N-c} \\ > 0 & ; \quad \frac{x(-c)}{N-c} < p \leq \frac{xN}{(n-1)(N-c)} \\ < 0 & ; \quad \frac{x}{n-1} + \frac{(-c)(n-x-1)}{(n-1)(N-c)} \leq p < 1 + \frac{c(n-x-1)}{N-c} \\ 0 & ; \quad 1 + \frac{c(n-x-1)}{N-c} \leq p \leq 1 \end{cases}$$

Demostación

Se probará para cada uno de los cuatro intervalos.

a) Consideremos $0 \leq p \leq \frac{x(-c)}{N-c}$.

En este caso, al ser $c < 0$, $\frac{x(-c)}{N-c} < \frac{x(-c)}{N}$, por lo que, teniendo en cuenta la expresión

de $G(p)$, se tiene que

$$G_{N,x,n,c}(p) = G_{N-c,x,n,c}(p) = 1$$

y, por lo tanto, la diferencia es cero.

b) Consideremos $1 + \frac{c(n-x-1)}{N-c} \leq p \leq 1$.

En este caso, $1 + \frac{c(n-x-1)}{N-c} > 1 + \frac{c(n-x-1)}{N}$, por lo que:

$$G_{N,x,n,c}(p) = G_{N-c,x,n,c}(p) = 0$$

c) Sea $\frac{x(-c)}{N-c} < p \leq \frac{xN}{(n-1)(N-c)}$.

Por el Lema 5.10 sabemos que para todo $0 < \varepsilon < 1$, para todo $0 < x < n-1$ y para todo p tal que $\frac{-xc}{N} \leq p \leq \frac{x}{n-1}$, se verifica que

$$H_{N,x,n,c} \left(p + \frac{\varepsilon c}{N} \right) - H_{N,x,n,c}(p) > \varepsilon \left\{ H_{N,x,n,c} \left(p + \frac{c}{N} \right) - H_{N,x,n,c}(p) \right\} \quad [5.14]$$

Aplicando el Teorema 5.1

$$H_{N,x,n,c} \left(p + \frac{c}{N} \right) - H_{N,x,n,c}(p) = - \binom{n}{x} (n-x) c \frac{(pN+c)^{(x,c)} (qN)^{(n-x-1,c)}}{N^{(n,c)}} \quad [5.15]$$

En consecuencia, de [5.14] y [5.15] se tiene que

$$H_{N,x,n,c}\left(p + \frac{\varepsilon c}{N}\right) - H_{N,x,n,c}(p) > -\varepsilon \binom{n}{x} (n-x) c \frac{(pN+c)^{(x,c)}(qN)^{(n-x-1,c)}}{N^{(n,c)}} \quad [5.16]$$

Haciendo el cambio: $p = \frac{N-c}{N}p'$ y $\varepsilon = p'$, se tiene:

- $p + \frac{\varepsilon c}{N} = \frac{N-c}{N}p' + \frac{p'c}{N} = p'$
- $pN + c = \frac{N-c}{N}p'N + c = p'(N-c) + c$
- $qN = (1-p)N = N - pN = N - \frac{N-c}{N}p'N = N - Np' + cp' =$
 $= q'N + c(1-q') = q'(N-c) + c$

y podemos escribir de nuevo, volviendo, por comodidad, a notar p en lugar de p' , (5.16) como sigue

$$H_{N,x,n,c}(p) - H_{N,x,n,c}\left(p - \frac{cp}{N}\right) > -p \binom{n}{x} (n-x) \frac{c(p(N-c)+c)^{(x,c)}(q(N-c)+c)^{(n-x-1,c)}}{N^{(n,c)}}$$

expresión que será cierta, teniendo en cuenta que $\frac{-xc}{N} \leq p \leq \frac{x}{n-1}$ y el cambio efectuado, para todo p tal que

$$\frac{-xc}{N-c} \leq \frac{N-c}{N}p \leq \frac{x}{n-1}$$

o, de modo equivalente, para todo p tal que

$$\frac{-xG}{N-c} \leq p \leq \frac{xN}{(n-1)(N-c)}$$

Por otra parte, aplicando el Teorema 5.14, tenemos que:

$$\begin{aligned} H_{N,x,n,c}(p) - H_{N-c,x,n,c}(p) &= \\ H_{N,x,n,c}(p) - H_{N,x,n,c}\left(p - \frac{cG}{N}\right) &+ \frac{\binom{n}{x+1} (x+1)c [p(N-c)]^{(x-1,c)} [q(N-c)+c]^{(n-x-1,c)}}{(N-c)N^{(n,c)}} > \\ > \frac{-p \binom{n}{x} (n-x)c [p(N-c)+c]^{(x,c)} [q(N-c)+c]^{(n-x-1,c)}}{N^{(n,c)}} + \\ &+ \frac{\binom{n}{x+1} (x+1)c [p(N-c)]^{(x-1,c)} [q(N-c)+c]^{(n-x-1,c)}}{(N-c)N^{(n,c)}} \end{aligned}$$

Teniendo en cuenta que

$$\binom{n}{x} (n-x) = \binom{n}{x+1} (x+1)$$

$$[p(N-c)]^{(x+1,c)} = p(N-c)[p(N-c)+c]^{(x,c)}$$

el segundo miembro de la desigualdad anterior es igual a cero, por lo que

$$H_{N,x,n,c}(\rho) - H_{N-c,x,n,c}(\rho) > 0$$

para todo ρ tal que $\frac{-xc}{N-c} \leq \rho \leq \frac{xN}{(n-1)(N-c)}$

Finalmente, basta tener en cuenta las definiciones de $G_{N,x,n,c}(\rho)$ y $H_{N,x,n,c}(\rho)$ para que se deduzca de forma inmediata que $G_{N,x,n,c}(\rho) - G_{N-c,x,n,c}(\rho) > 0$ para todo ρ tal que

$$\frac{-xc}{N-c} \leq \rho \leq \frac{xN}{(n-1)(N-c)}$$

como queríamos demostrar.

d) Sea $\frac{x}{n-1} - \frac{c(n-x-1)}{(n-1)(N-c)} \leq \rho < 1 + \frac{c(n-x-1)}{N-c}$

Por el Lema 5.10 sabemos que para todo $0 < \varepsilon < 1$; $x: 1, 2, \dots, n-2$; y para todo ρ tal que

$$\frac{x}{n-1} - \frac{c}{N} \leq \rho \leq 1 + \frac{c(n-x-2)}{N} \tag{5.7}$$

se verifica que

$$H_{N,x,n,c} \left(\rho + \frac{\varepsilon c}{N} \right) - H_{N,x,n,c}(\rho) < \varepsilon \left\{ H_{N,x,n,c} \left(\rho + \frac{c}{N} \right) - H_{N,x,n,c}(\rho) \right\}$$

Aplicando el Teorema 5.1, se tiene que

$$H_{N,x,n,c} \left(\rho + \frac{\varepsilon c}{N} \right) - H_{N,x,n,c}(\rho) < -\varepsilon \binom{n}{x} (n-x)c \frac{(\rho N + c)^{(x,c)} (qN)^{(n-x-1,c)}}{N^{(n,c)}} \tag{5.18}$$

Llamando $\rho = \frac{N-c}{N} \rho'$ y $\varepsilon = \rho'$, al igual que hicimos con la expresión [5.16] del

apartado anterior, y volviendo, por comodidad, a notar p en lugar de p' , podemos describir la expresión [5.18] como sigue:

$$H_{N,x,n,c}(p) - H_{N,x,n,c}\left(p - \frac{cp}{N}\right) < -p \binom{n}{x} (n-x) \frac{c[p(N-c)+c]^{(x,c)}[q(N-c)+c]^{(n-x-1,c)}}{N^{(n,c)}}$$

Teniendo en cuenta [5.17] y el cambio efectuado, esta última expresión será cierta para todo p tal que

$$\frac{x}{n-1} - \frac{c}{N} \leq \frac{N-c}{N} p \leq 1 + \frac{c(n-x-2)}{N}$$

o, de forma equivalente,

$$\frac{x}{n-1} - \frac{c(n-x-1)}{(n-1)(N-c)} \leq p \leq 1 + \frac{c(n-x-1)}{N-c}$$

Por otra parte, aplicando el Teorema 5.14, tenemos que

$$\begin{aligned} H_{N,x,n,c}(p) - H_{N-c,x,n,c}(p) &= \\ &= H_{N,x,n,c}(p) - H_{N,x,n,c}\left(p - \frac{cp}{N}\right) + \frac{\binom{n}{x+1} (x+1)c[p(N-c)]^{(x+1,c)}[q(N-c)+c]^{(n-x-1,c)}}{(N-c)N^{(n,c)}} < \\ &< \frac{-p \binom{n}{x} (n-x)c[p(N-c)+c]^{(x,c)}[q(N-c)+c]^{(n-x-1,c)}}{N^{(n,c)}} + \\ &+ \frac{\binom{n}{x+1} (x+1)c[p(N-c)]^{(x+1,c)}[q(N-c)+c]^{(n-x-1,c)}}{(N-c)N^{(n,c)}} \end{aligned}$$

Puesto que el segundo miembro de la anterior desigualdad es igual a cero como ya probamos en el desarrollo del apartado c), queda finalmente que

$$H_{N,x,n,c}(p) - H_{N-c,x,n,c}(p) < 0$$

para todo p tal que : $\frac{x}{n-1} - \frac{c(n-x-1)}{(n-1)(N-c)} \leq p < 1 + \frac{c(n-x-1)}{N-c}$

Teniendo, por último, en cuenta las definiciones de $G_{N,x,n,c}(p)$ y $H_{N,x,n,c}(p)$ se deduce inmediatamente que $G_{N,x,n,c}(p) - G_{N-c,x,n,c}(p) < 0$ para todo p tal que

$$\frac{x}{n-1} - \frac{c(n-x-1)}{(n-1)(N-c)} \leq p < 1 + \frac{c(n-x-1)}{N-c}$$

como queríamos demostrar. ■

COMPARACIÓN ENTRE LAS FUNCIONES DE DISTRIBUCIÓN DE LAS DISTRIBUCIONES DE PÓLYA Y BINOMIAL DE IGUAL VALOR ESPERADO Y TAMAÑO MUESTRAL

Consideremos la distribución binomial B(n,p). Para valores x:0,1,...,n, definimos:

$$G_{x,n}(p) = \sum_{m=0}^x \binom{n}{m} p^m (1-p)^{n-m} ; 0 \leq p \leq 1$$

El siguiente teorema permite establecer la comparación entre esta función y la función $G_{N,x,n,c}(p)$ para todo p, excepto para valores de p pertenecientes al siguiente intervalo:

$$\left(\frac{xN}{(n-1)(N-c)} ; \frac{x}{n-1} - \frac{c(n-x-1)}{(n-1)(N-c)} \right)$$

Teorema 5.17

Para $0 < x < n-1 < N$, se verifica que

$$G_{N,x,n,c}(p) - G_{x,n}(p) = \begin{cases} 0 & \text{si } p=0 \\ >0 & \text{si } 0 < p \leq \frac{xN}{(n-1)(N-c)} \\ <0 & \text{si } \frac{x}{n-1} - \frac{c(n-x-1)}{(n-1)(N-c)} \leq p < 1 \\ 0 & \text{si } p=1 \end{cases}$$

Demostración

Teniendo en cuenta que $\lim_{N \rightarrow \infty} G_{N,x,n,c}(p) = G_{x,n}(p)$ podemos escribir:

$$G_{N,x,n,c}(p) - G_{x,n}(p) = \sum_{i=0}^{\infty} [G_{N-ic,x,n,c}(p) - G_{N-(i+1)c,x,n,c}(p)] \tag{5.19}$$

A partir del Teorema 5.16 se obtiene que

$$G_{N-ic,x,n,c}(p) - G_{N-(i+1)c,x,n,c}(p) = \begin{cases} 0 & ; 0 \leq p \leq \frac{x(-c)}{N-(i+1)c} \\ >0 & ; \frac{x(-c)}{N-(i+1)c} < p \leq \frac{x(N-ic)}{(n-1)(N-(i+1)c)} \\ <0 & ; \frac{x}{n-1} - \frac{c(n-x-1)}{(n-1)(N-(i+1)c)} \leq p < 1 + \frac{c(n-x-1)}{N-(i+1)c} \\ 0 & ; 1 + \frac{c(n-x-1)}{N-(i+1)c} \leq p \leq 1 \end{cases}$$

(i) Si $p=0$ ó $p=1$, entonces,

$$G_{N-ic,x,n,c}(p) - G_{N-(i+1)c,x,n,c}(p) = 0$$

y, por tanto, $G_{N,x,n,c}(p) - G_{x,n}(p) = 0$.

(ii) Puesto que: $\frac{xN}{(n-1)(N-c)} < \frac{x(N-ic)}{(n-1)(N-(i+1)c)}$ para todo $i:1,2,\dots$, se tiene que si

$0 < \rho \leq \frac{xN}{(n-1)(N-c)}$, entonces $G_{N,x,n,c}(\rho) - G_{x,n}(\rho) > 0$, ya que, por pequeño que sea ρ ,

existirá un i suficientemente grande como para que sea $\rho > \frac{x(-c)}{(N-(i+1)c)}$.

A partir de ese valor de i , los términos de la serie [5.19] serán estrictamente positivos, mientras que los primeros términos serán nulos.

(iii) Puesto que $\frac{x}{n-1} - \frac{c(n-x-1)}{(n-1)(N-c)} > \frac{x}{n-1} - \frac{c(n-x-1)}{(n-1)(N-(i+1)c)}$, entonces $G_{N,x,n,c}(\rho) - G_{x,n}(\rho) < 0$

ya que, por grande que sea ρ , existirá un i suficientemente grande como para que sea

$$\rho < 1 + \frac{c(n-x-1)}{(N-(i+1)c)}$$

A partir de ese valor de i , los términos de la serie [5.19] serán estrictamente negativos, mientras que los primeros términos serán nulos.

■

En las definiciones de las funciones $G_{N,x,n,c}(\rho)$ y $G_{x,n}(\rho)$ subyacen las funciones de distribución de las distribuciones de Pólya $P(N,p,n,c)$ y binomial $B(n,p)$ respectivamente, consideradas estas últimas como función de ρ , para un valor x fijo de las correspondientes variables aleatorias, con la salvedad de que en $G_{N,x,n,c}(\rho)$ y en $G_{x,n}(\rho)$ no se hace ninguna restricción sobre los valores que puede tomar ρ .

En consecuencia, si notamos por $F_{N,p,n,c}(x)$ y $F_{p,n}(x)$ a las funciones de distribución de las distribuciones de Pólya y binomial respectivamente, es claro que, para todo $x:0,1,\dots,n-1$ y para todo valor ρ fijo,

$$F_{N,p,n,c}(x) = G_{N,x,n,c}(p)$$

$$F_{p,n}(x) = G_{p,n}(x)$$

Vamos a poder, por tanto, reinterpretar el resultado obtenido en el Teorema 5.17 y comparar las funciones de distribución de las distribuciones de Pólya y binomial de igual valor esperado.

Veamos, en primer lugar, el siguiente lema previo.

Lema 5.18

Dadas las distribuciones de Pólya $P(N,p,n,c)$ ($c < 0$) y binomial $B(n,p)$, se verifica:

$$(i) F_{N,p,n,c}(0) < F_{p,n}(0)$$

$$(ii) F_{N,p,n,c}(n-1) > F_{p,n}(n-1)$$

Demostración

$$(i) F_{N,p,n,c}(0) = \binom{n}{0} \frac{M^{(0,c)}(N-M)^{(n,c)}}{N^{(n,c)}} = \frac{(N-M)^{(n,c)}}{N^{(n,c)}} ; \left(\frac{M}{N} = p \right)$$

$$F_{p,n}(0) = \binom{n}{0} p^0(1-p)^n = \left(\frac{N-M}{N} \right)^n$$

Basta tener en cuenta que $\frac{N-M+ic}{N+ic} < \frac{N-M}{N}$ para todo $i: 1, 2, \dots, n-1$ ($n > 1$) para que quede probado que $F_{N,p,n,c}(0) < F_{p,n}(0)$.

$$(ii) F_{N,p,n,c}(n-1) = 1 - F_{N,p,n,c}(n) = 1 - \frac{M^{(n,c)}}{N^{(n,c)}}$$

$$F_{p,n}(n-1) = 1 - F_{p,n}(n) = 1 - \left(\frac{M}{N} \right)^n$$

Se comprueba, de forma análoga a lo efectuado en (i), que (para $n > 1$): $\frac{M^{(n,c)}}{N^{(n,c)}} < \left(\frac{M}{N}\right)^n$,

por lo que $F_{N,p,n,c}(n-1) > F_{p,n}(n-1)$.

Teorema 5.19

Dadas las distribuciones de Pólya $P(N,p,n,c)$ ($c < 0$) y binomial $B(n,p)$, se verifica la siguiente relación entre las correspondientes funciones de distribución $F_{N,p,n,c}(x)$ y $F_{p,n}(x)$ para todo $x: 0, 1, \dots, n-1$ ($n > 1$):

$$F_{N,p,n,c}(x) - F_{p,n}(x) \begin{cases} < 0 & \text{si } x \leq \frac{p(n-1)(N-c) + c(n-1)}{N} \\ > 0 & \text{si } x \geq \frac{p(n-1)(N-c)}{N} \end{cases}$$

Demostración

Teniendo en cuenta que para todo $x: 0, 1, \dots, n-1$, y para cada valor de p fijo compatible con el resto de parámetros de la distribución de Pólya,

$$\begin{aligned} F_{N,p,n,c}(x) &= G_{N,x,n,c}(p) \\ F_{p,n}(x) &= G_{x,n}(p), \end{aligned}$$

basta con reinterpretar el resultado dado en el Teorema 5.17, expresando la diferencia en función de x , para p constante. Así, se tiene que para $0 < x < n-1$,

$$F_{N,p,n,c}(x) - F_{p,n}(x) \begin{cases} < 0 & \text{si } x \leq \frac{p(n-1)(N-c) + c(n-1)}{N} \\ > 0 & \text{si } x \geq \frac{p(n-1)(N-c)}{N} \end{cases}$$

La demostración queda completada teniendo en cuenta (Lema 5.18) que $F_{N,p,n,c}(0) < F_{p,n}(0)$ y $F_{N,p,n,c}(n-1) > F_{p,n}(n-1)$.

CONCLUSIONES

A la vista del Teorema 5.19 se concluye que como máximo existe un único valor de x entero para el que la comparación de las funciones de distribución de las distribuciones de Pólya con parámetro reemplazamiento negativo y binomial de igual tamaño muestral y valor esperado, queda sin determinar. Concretamente esto ocurre si x pertenece al intervalo I siendo

$$I = \left(\frac{p(n-1)(N-c)+c(n-1)}{N} ; \frac{p(n-1)(N-c)}{N} \right)$$

Puede comprobarse sin dificultad que la amplitud de este intervalo es menor que la unidad. En efecto, la amplitud del intervalo es $\frac{(-c)(n-1)}{N}$ que es menor que la unidad ya que, por hipótesis estamos aceptando que $(-c)(n-1) < \min \{M; N-M\} < N$.

Este resultado, Teorema 5.19, supone mejorar la acotación para la función de distribución de la distribución de Pólya obtenida por Ollero y Ramos (1995) a partir de la descripción de esta distribución como binomial generalizada, resultado que recogimos en el Capítulo III y que con la notación que estamos utilizando aquí sería:

$$\begin{aligned} (i) \quad & 0 < F_{N,p,n,c}(x) < F_{p,n}(x) \quad \text{si } 0 \leq x \leq \mu - 1 \\ (ii) \quad & F_{p,n}(x) < F_{N,p,n,c}(x) < 1 \quad \text{si } \mu \leq x < n \end{aligned}$$

Las funciones de distribución de la distribución de Pólya y de la distribución binomial aproximante son aquí comparables salvo en el intervalo de amplitud unidad $(\mu - 1; \mu)$.

Veamos seguidamente que el intervalo de indeterminación I en nuestro resultado está contenido estrictamente en $(\mu - 1; \mu)$. En efecto,

$$\begin{aligned} a) \quad & p(n-1)(N-c)+c(n-1) = Nnp - cnp - Np + cp + cn - c = Nnp - Np + cn(1-p) + c(p-1) = \\ & = Nnp - Np + c(p-1)(1-n) = Nnp - Np - (1-p)(-c)(n-1) > Nnp - Np - (1-p)N = \\ & = Nnp - N = N(np-1) = N(\mu-1) \end{aligned}$$

Luego,

$$\frac{p(n-1)(N-c)+c(n-1)}{N} > \mu - 1.$$

b)

$$\begin{aligned} p(n-1)(N-c) &= Nnp - cnp - Np + cp = Nnp - p(cn+N-c) = \\ &= Nnp - p(N+c(n-1)) < Nnp = N\mu \end{aligned}$$

ya que estamos aceptando que $(-c)(n-1) < \min \{M, N-M\} < N$ y, por tanto, $N+c(n-1)$ es una cantidad positiva.

En conclusión:

$$\frac{p(n-1)(N-c)}{N} < \mu.$$

BIBLIOGRAFÍA

BIBLIOGRAFÍA

ABRAMOWITZ, M. y STEGUN, I.A. (1972): *Handbook of Mathematical Functions*. Dover Publications, Inc., New York.

BAGCHI, A. y PAL, A.K. (1985): "Asymptotic normality in the generalized Pólya-Eggenberger urn model, with an application to computer data structures". *SIAM J. Alg. Disc. Meth.*, Vol. 6, n° 3, pp. 394-405.

BERG, S. (1989): "Urn Models". *Encyclopedia of Statistical Sciences*, 9, pp. 424-436, Wiley.

BOSCH, A.J. (1963): "The Pólya distribution". *Statistica Neerlandica*, 17, n° 3, pp. 201-213.

COLEMAN, J. (1964): *Introduction to Mathematical Sociology*, (Cap. 11) The Free Press, New York.

DARROCH, J.N. (1964): "On the Distribution of the Number of Successes in Independent Trials", *Ann. Math. Statist.*, 31, pp. 1317-1321

DAVIES, O.L. (1933): "On asymptotic formulae for the hypergeometric series I". *Biometrika*, 25, pp. 295-322.

DAVIES, O.L. (1934): "On asymptotic formulae for the hypergeometric series II". *Biometrika*, 26, pp. 59-107.

EGGENBERGER, F. y POLYA, G. (1923): "Über die Statistik verketteter Vorgänge". *Zeitschrift für Angewandte Mathematik und Mechanik*, Vol. 3, pp. 279-289.

EGGENBERGER, F. y POLYA, G. (1928): "Sur l'interprétation de certaines courbes de fréquence". *Comptes Rendus, Académie des Sciences, Paris*, 187, pp. 870-872.

EHRENFEST, P. y EHRENFEST, T. (1907): "Über zwei bekannte Einwände gegen das Boltzmannsche Theorem". *Physikalische Zeitschrift*, Vol. 8, pp. 311-314.

FELLER, W. (1993): *Introducción a la teoría de probabilidades y sus aplicaciones*, Vol. 1, pp. 131-132, Ed. Limusa, Méjico.

FREEDMAN, D.A. (1965): "Bernard Friedman's urn" *Annals of Mathematical Statistics*, 36, pp. 956-970.

FRIEDMAN, B. (1949): "A simple urn model". *Comm. Pure Applied Math.*, 2, pp. 59-70.

GOUET, R. (1989): "A martingale approach to strong convergence in a generalized Pólya-Eggenberger urn model". *Statistics & Probability Letters*, 8, pp. 225-228.

HALD, A. (1960): "The Compound Hypergeometric Distribution and a System of Single Sampling Inspection Plans Based on Prior Distributions and Costs", *Technometrics*, 2, pp. 275-340.

HEITELE, D. (1975): *Educational Studies in Mathematics*, Vol. 6. Reidel, Dordrecht, Netherlands.

HOEFFDING, W. (1956): "On the Distribution of the Number of Successes in Independent Trials". *Ann. Math. Statist.*, 27, p. 713-721.

JANARDAN, K.G. (1984): "On characterizing the Markov-Pólya distribution". *Sankhya*, Vol. 46, Series A, Pt. 3, pp. 444-453.

JANARDAN, K.G. y SHAEFFER, D.J. (1977): "A Generalization of Markov-Pólya Distribution. Its Extensions and Applications". *Biometrische Zeitschrift*, Vol. 19, n. 2, pp. 87-106.

JOGDEO, K. y SAMUELS, S.M. (1968): "Monotone Convergence of Binomial Probabilities and a Generalization of Ramanujan's Equation". *Ann. Math. Statist.*, 39, pp. 1191-1195.

JOHNSON, N.L. y KOTZ, S. (1969): *Distributions in Statistics. Discrete Distributions*. John Wiley & Sons. New York.

JOHNSON, N.L. y KOTZ, S. (1977): *Urn Models and Their Application*. Wiley, New York.

JOHNSON, N.L., KOTZ, S. y KEMP, A.W. (1992): *Univariate Discrete Distributions*. Wiley. New York.

JORDAN, C. (1927): "Sur un cas généralisé de la probabilité des épreuves répétées" *Comptes Rendus, Académie des Sciences, Paris*, 184, pp. 315-317.

KEILSON, J. y GERBER, H. (1971): "Some results for Discrete Unimodality" *Journal of the American Statistical Association*, 66, pp. 386-389

KEMP, C.D. y KEMP, A.W. (1956): "Generalized Hypergeometric Distributions" *Journal of the Royal Statistical Society, Series B*, 18, pp. 202-211

KEMP, A.W. (1968): *Studies in Univariate Discrete Distribution Theory Based on the Generalized Hypergeometric Function and Associated Differential Equations*. Ph.D. thesis, Belfast: The Queen's University of Belfast.

KEMP, A.W. (1968): "A wide class of discrete distributions and the associated differential equations". *Sankhya, Series A*, 30, pp. 401-410.

KRYUKOV, V.F. (1973): "Theoretical Justification of Pearson Distribution Curve". *Soviet Hidrology Selected papers*, n 1, pp. 71-76.

MARKOV, A.A. (1917): "On limiting formulas for computing probabilities" (en ruso). *Imperatovskiy Akademy Nauk*, 6, pp. 171-186.

MARSHALL, A.W. y OLKIN, I. (1990): "Bivariate Distributions Generated from Pólya-Eggenberger Urn Models". *Journal of Multivariate Analysis*, 35, pp. 48-65.

NAOR, P. (1956): "On machine interference". *Journal of the Royal Statistical Society, Series B*, 18, pp. 280-287.

NAOR, P. (1957): "Normal approximation to machine interference with many repairmen". *Journal of the Royal Statistical Society, Series B*, 1, pp. 334-341.

NOACK, A. (1950): "A class of random variables with discrete distributions". *Annals of Mathematical Statistics*, 21, pp. 127-132.

OLLERO, J. (1988): "Una propiedad caracterizadora de la distribución Binomial de Poisson". *XVI Reunión Nacional de Estadística, Investigación Operativa e Informática*. Universidad de Alicante.

OLLERO, J. y RAMOS, H.M. (1995): "Description of a Subfamily of the Discrete Pearson System as Generalized-Binomial Distributions" *J. Ital. Statist. Soc.*, 2, pp. 235-249

PATEL, J.K., KAPADIA, C.H. y OWEN, D.B. (1976): *Handbook of Statistical Distributions*. Marcel Dekker, Inc. New York and Basel.

PATIL, G.P. y JOSHI, S.W. (1968): *A dictionary and bibliography of discrete distributions*. Oliver & Boyd. Edinburgh.

PEARSON, K. (1895): "Memoir on skew variation in homogeneous material" *Phil. Trans. Roy. Soc. A*, 186, pp. 343-414.

POISSON, S.D. (1837): *Recherches sur la Probabilité des Jugements en Matière Criminelle et en Matière Civile, Précédées des Règles Générales du Calcul des Probabilités*. Bachelier.

POLYA, G. (1954): *Patterns of Plausible Inference*. Princeton University Press, Princeton, N.J.

RAMOS, H.M. (1989): *Aportaciones al estudio de la distribución hipergeométrica: caracterización y propiedades de monotonía*. Tesis Doctoral. Universidad de Granada.

- RAMOS, H.M. y OLLERO, J. (1989): "Teoremas de caracterización de la distribución hipergeométrica". *Cuadernos de Estadística Matemática. Serie A*, nº 11, pp. 85-101.
- RENYI, A. (1970): *The Teaching of Probability and Statistics*. L. Rade, ed. pp. 273-282.
- SAMUELS, S.M. (1965): "On the Number of Successes in independent Trials". *Ann. Math. Statist.*, 36, pp. 1272-1278.
- SKELLAM. (1948): "A probability distribution derived from the binomial distribution by regarding the probability of success as variable between the sets of trials". *Journal of the Royal Statistical Society*, vol. 10, p. 257.
- SKIBINSKY, M. (1970): "A characterization of hypergeometric distributions". *Journal of the American Statistical Association*, 65, pp. 926-929.
- STANCU, D.D. (1968): "On the Markov-Pólya Probability Distribution", *Bull. Math. Soc. Sci. Math. de la R.S. de Roumanie*, Tomo 12 (60) n 4, pp. 203-208.
- STEYN, H.S. (1957): "On discrete multivariate probability function". *Kon. Nederl. Wetensch. Proc., Series A*, 54, pp. 23-30.
- UHLMANN, W. (1966): "Vergleich der hypergeometrischen mit der Binomial-Verteilung". *Metrika*, 10, pp. 145-158.
- WEI, LEE-JEN. (1976): *A class of designs for sequential clinical trials*. Department of Mathematics, University of South Carolina.
- WOODBURY, M.A. (1949): "On a probability distribution". *Annals of Mathematical Statistics*, 20, pp. 311-313.