Tema 5
Algunos modelos de distribuciones discretas (continuación)

1. Distribución binomial negativa

Consideremos ahora un experimento aleatorio consistente en repeticiones independientes de ensayos de Bernoulli con probabilidad de éxito constante, hasta que aparezca el éxito k-ésimo. Es decir, en lugar de fijar el número de ensayos y observar el número de éxitos en esas n realizaciones, se repiten las realizaciones hasta obtener un número determinado de éxitos y contabilizamos los fracasos. Definimos la variable aleatoria con **distribución binomial negativa** como aquella que modeliza el número de fracasos antes de que aparezca el éxito k-ésimo.

Nota.- Si en vez de contabilizar el número de fracasos antes del k-ésimo éxito se contabiliza el número de pruebas necesarias (variable aleatoria $Y = X + k$), se obtiene la distribución de Pascal que verifica
$\mathrm{P}[Y = n] = \mathrm{P}[X = n - k]$.

La variable aleatoria X puede tomar los valores $x = 0, 1, 2, \ldots$ y k debe ser un entero positivo, es decir, $k = 1, 2, \ldots$.

Calculemos la función masa de probabilidad de la variable aleatoria binomial negativa; esto es,

$$\mathrm{P}[X = x],\ x = 0, 1, 2, \ldots$$

Uno de los posibles sucesos para el cual ocurre $\{X = x\}$ sería que en las x primeras repeticiones apareciera "fracaso", en las $k - 1$ siguientes "éxito" y en la ultima "éxito"

$$\overline{E} \overset{x)}{\ldots} \overline{E} | E \overset{k-1)}{\ldots} E | E$$

y como las repeticiones de las pruebas de Bernoulli son independientes, la probabilidad del suceso anterior será

$$\mathrm{P}\left[\overline{E} \overset{x)}{\ldots} \overline{E} E \overset{k)}{\ldots} E\right] = \mathrm{P}(\overline{E}) \overset{x)}{\cdots} \mathrm{P}(\overline{E}) \cdot \mathrm{P}(E) \overset{k)}{\cdots} \mathrm{P}(E) = (1-p)^x p^k$$

Pero la obtención de los x fracasos y los $k-1$ éxitos se pueden obtener de tantas maneras como las permutaciones con repetición de $x + k - 1$ elementos en los que hay iguales x y $k - 1$, es decir

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

Por tanto, la función masa de probabilidad de esta variable aleatoria es

$$\mathrm{P}[X = x] = \binom{x+k-1}{x}(1-p)^x p^k, \qquad x = 0, 1, 2, \cdots \quad (k = 1, 2, \cdots; \quad 0 \leq p < 1)$$

y se notará como $X \sim BN(k,p)$.

Ejemplos

- Número de preguntas falladas en un examen tipo test antes de tener el décimo acierto.
- Número de unidades defectuosas antes de obtener un número concreto de correctas.

Algunas de las características de la distribución binomial negativa son:

Función de distribución:

$$F(x) = \mathrm{P}[X \leq x] = \begin{cases} 0 & x < 0 \\ \displaystyle\sum_{i=0}^{[x]} \binom{i+k-1}{i} (1-p)^i p^k & x \geq 0 \end{cases}$$

Función generatriz de momentos:

$$M(t) = \left(\frac{p}{1-(1-p)\mathrm{e}^t} \right)^k \qquad \forall t < -\ln(1-p)$$

Momentos: Existen todos dado que existe la función generatriz de momentos. Nos limitaremos a obtener los momentos de primer y segundo orden.

Media

$$\mathrm{E}[X] = \frac{k(1-p)}{p}$$

Varianza

$$\mathrm{Var}[X] = \frac{k(1-p)}{p^2}$$

Relación entre las distribuciones binomial y binomial negativa

Si X es una variable aleatoria con distribución $BN(k,p)$, el suceso $\{X = x\}$ significa la intersección de los sucesos

$A =$ {se han obtenido $k-1$ éxitos en los primeros $k+x-1$ ensayos de una serie de pruebas independientes de Bernoulli}

$B =$ {se ha obtenido un éxito en el siguiente ensayo de la serie}

El suceso A equivale a $\{Y = k-1\}$, con $Y \sim B(k+x-1,p)$, y el segundo es independiente del primero y ocurre con probabilidad p. Entonces

$$\mathrm{P}[X = x] = \mathrm{P}[Y = k-1] \cdot p$$

Caso particular: Distribución Geométrica

La distribución geométrica es un caso particular de la distribución binomial negativa para el caso de $k = 1$. Es decir, modeliza el número de fracasos antes del primer éxito en repeticiones independientes de ensayos de Bernoulli con probabilidad de éxito p. Su función masa de probabilidad es

$$\mathrm{P}[X = x] = (1-p)^x p, \qquad x = 0, 1, 2, \cdots \quad (0 < p < 1)$$

y se notará como $X \sim G(p)$.

Sus características más relevantes son

- **Función de distribución**:

$$F(x) = \mathrm{P}[X \leq x] = \begin{cases} 0 & x < 0 \\ \displaystyle\sum_{i=0}^{[x]} (1-p)^i p & x \geq 0 \end{cases}$$

  Pero como [1]

$$\sum_{i=0}^{[x]} (1-p)^i p = p \sum_{i=0}^{[x]} (1-p)^i = p \frac{1-(1-p)^{[x]+1}}{1-(1-p)} = 1 - (1-p)^{[x]+1}$$

  se puede reescribir la función de distribución como

$$F(x) = \mathrm{P}[X \leq x] = \begin{cases} 0 & x < 0 \\ 1-(1-p)^{[x]+1} & x \geq 0 \end{cases}$$

- **Función generatriz de momentos**:

$$M(t) = \frac{p}{1-(1-p)\mathrm{e}^t} \qquad \forall t < -\ln(1-p)$$

- **Media**:

$$\mathrm{E}[X] = \frac{1-p}{p}$$

[1] Recordemos que la suma de los n términos de una progresión geométrica a_n de razón r es

$$S_n = \frac{a_1 - a_n \cdot r}{1-r}$$

- **Varianza**:

$$\mathrm{Var}[X] = \frac{1-p}{p^2}$$

- **Propiedad de olvido o falta de memoria**:

  Esta propiedad se formula de la siguiente forma: Si X es una variable aleatoria con distribución geométrica de parámetro p, $G(p)$, entonces se verifica que

$$\mathrm{P}[X \geq h+k | X \geq h] = \mathrm{P}[X \geq k], \quad h, k = 0, 1, 2, \ldots$$

  "Si se ha realizado la h-ésima repetición del experimento o prueba de Bernoulli y no se ha obtenido ningún éxito, entonces la probabilidad de que se realicen por lo menos otras k repeticiones sin que se presente ningún éxito, es decir que se realicen por lo menos $h+k$ repeticiones, es la misma que si consideramos que la primera repetición es la $h+1$-ésima; es decir, esa probabilidad es la misma que la probabilidad de que realicemos al menos k repeticiones sin obtener el primer éxito, y por consiguiente se olvidan las h repeticiones realizadas inicialmente.

  Vamos a demostrarla, si $X \sim G(p)$ y $x = 0, 1, 2, \ldots$

$$\mathrm{P}[X \geq x] = 1 - \mathrm{P}[X < x] = 1 - F(x-1) = 1 - (1 - (1-p)^x) = (1-p)^x.$$

  Por tanto, si $k, h, \in \mathbb{N}$

$$\mathrm{P}[X \geq h+k | X \geq h] = \frac{\mathrm{P}[X \geq h+k, X \geq h]}{\mathrm{P}[X \geq h]} =$$

$$\frac{\mathrm{P}[X \geq h+k]}{\mathrm{P}[X \geq h]} = \frac{(1-p)^{h+k}}{(1-p)^h} = (1-p)^k = \mathrm{P}[X \geq k]$$

  Esta propiedad, además, caracteriza a la distribución geométrica como la única entero valuada (positiva) que la cumple.

La distribución de probabilidad geométrica se aplica frecuentemente en el estudio de la distribución de la duración de tiempos de espera. Así pues, si las repeticiones del experimento se realizan a intervalos regulares de tiempo, entonces la variable aleatoria con distribución geométrica nos dará el número de intervalos de tiempo transcurridos hasta que aparezca el primer éxito.

EJERCICIOS

1.- *Un examen de Estadística consta de 20 preguntas tipo test y se conoce de experiencias anteriores que un alumno tiene probabilidad 0.7 de contestar bien cada pregunta. Obtener:*

a) La probabilidad de que la primera pregunta que contesta bien sea la cuarta.
b) Sabiendo que para aprobar el examen es necesario contestar bien a 10 preguntas, ¿cuál es la probabilidad de que apruebe al contestar la pregunta duodécima?

a) $X=$ Número de fallos antes del primer éxito, $\quad X \sim G(0.7)$

$$\mathrm{P}[X=3]=(0.3)^3 0.7=0.0189$$

b) $X=$ Número de fallos antes del décimo éxito, $\quad X \sim BN(10, 0.7)$

$$\mathrm{P}[X=2]=\binom{2+10-1}{2}(0.3)^2(0.7)^{10}=\binom{11}{2}(0.3)^2(0.7)^{10}.$$

Esta probabilidad se puede obtener mediante la relación con la binomial $Y=$ Número de éxitos en las 11 primeras pruebas, $\quad Y \sim B(11, 0.7)$

$$P[X=2]=P[Y=9]\cdot 0.7$$

2.- *Una máquina dedicada a la fabricación de piezas de alta precisión produce las piezas de una en una, siendo independiente la fabricación de cada pieza. La probabilidad de fabricar una pieza defectuosa es 0.15. Obtener*
a) La probabilidad de que la primera pieza defectuosa durante ese día sea la número 40.
b) Sabiendo que en la fabricación de cada pieza se tardan 20 segundos ¿cuál será el tiempo medio que hay que esperar hasta que sea producida la primera pieza defectuosa?
c) Probabilidad de que las ocho primeras piezas fabricadas sean todas buenas.

a) $X=$ Número de piezas buenas antes de la primera defectuosa, $\quad X \sim G(0.15)$

$$P[X=39]=(0.85)^{39}\cdot 0.15$$

b) $EX=\frac{1-p}{p}=\frac{0.85}{0.15}=5.666.$ que es el número medio de piezas buenas antes de la primera defectuosa. Por tanto, el tiempo pedido es $20\times(5.666+1)=133.33$ segundos.

c) $P[X\geq 8]=1-F(7)=1-P[X\leq 7]=0{,}2724905$
Si consideramos $Y=$ Número de piezas buenas en las 8 primeras, $\quad Y \sim B(8, 0.85)$

$$P[Y=8].$$

2. Distribución hipergeométrica

Se denomina **población** a cualquier colección de individuos, objetos o elementos arbitrarios y una **muestra de tamaño n** de esa población es cualquier subconjunto con n elementos.

El procedimiento de selección de muestras de una población se denomina **muestreo** y éste puede realizarse de distintas formas, dando lugar a distintos tipos de muestras de las que aquí vamos a distinguir las dos siguientes (suponiendo una población finita):

MUESTRA ALEATORIA CON REEMPLAZAMIENTO (CON DEVOLUCIÓN O SIMPLE).

Se selecciona al azar un elemento de la población (suponiendo que todos tienen la misma probabilidad de ser seleccionado), se devuelve a la población (después de anotar las características de interés) y se selecciona otro también de forma aleatoria; se continua el proceso hasta completar el tamaño muestral deseado.

Notemos que una muestra aleatoria con reemplazamiento puede contener elementos repetidos y, en ciertas ocasiones, esto no es conveniente (incluso puede no ser posible, por ejemplo, al estudiar la duración de vida de cierto tipo de bombillas).

MUESTRA ALEATORIA SIN REEMPLAZAMIENTO (SIN DEVOLUCIÓN).

Se selecciona al azar un elemento de la población (suponiendo que todos tienen la misma probabilidad de ser seleccionado); a continuación, sin devolverlo a la población, se selecciona otro suponiendo que los restantes tienen la misma probabilidad de ser elegidos y así sucesivamente.

Notemos que ahora todos los elementos de la muestra serán distintos.

El muestreo sin reemplazamiento es útil en muchas situaciones prácticas. Por ejemplo, consideremos una población con N individuos que deben elegir entre dos candidatos A y B a cierto puesto. Con objeto de realizar un sondeo de opinión antes de la elección se efectúa un muestreo de n individuos para ver su preferencia. En una situación de este tipo parece lógico hacer un muestreo sin reemplazamiento, que proporcionará más información sobre la intención de voto al tener todos los individuos diferentes.

Observemos que si en la población hay N_1 votantes de A y la muestra se elige con reemplazamiento, cada elemento de la muestra tiene probabilidad N_1/N de ser votante de A. Por tanto, la selección de la muestra equivale a la repetición de n pruebas de Bernoulli independientes con probabilidad constante de éxito (ser votante de A) y la variable X: Número de votantes de A en la muestra $\sim B(n, \frac{N_1}{N})$

Por el contrario, si el muestreo se realiza sin reemplazamiento, la probabilidad de éxito varía después de cada selección y las pruebas de Bernoulli no son independientes:

- La probabilidad de que el primer individuo seleccionado vote al candidato A es

$$\frac{N_1}{N}$$

- La probabilidad de que el segundo individuo seleccionado vote al candidato A depende de lo que haya ocurrido en la primera prueba

$$P(2A/1A) = \frac{N_1 - 1}{N - 1}, \qquad P(2A/1B) = \frac{N_1}{N - 1}.$$

Por tanto, la variable X: Número de votantes de A en la muestra, no tiene ahora una distribución binomial, sino que su distribución es la denominada HIPERGEOMÉTRICA.

Antes de introducir formalmente esta distribución notemos que, a efectos de calcular probabilidades, el muestreo sin reemplazamiento equivale a la selección simultánea de n elementos de la población, suponiendo que todos los subconjuntos de n elementos tienen la misma probabilidad de ser elegidos.

En efecto, si la selección se realiza simultáneamente no cabe distinguir entre dos muestras que tengan los mismos elementos (no se tiene en cuenta el orden) y hay en total $\binom{N}{n}$ muestras distintas y la probabilidad de elegir una cualquiera de ellas (esto es, que n individuos determinados formen parte de la muestra) es $\frac{1}{\binom{N}{n}}$.

Por otra parte, si la selección se realiza elemento a elemento, dos muestras con los mismos elementos en distinto orden serán distintas y, en total, hay $N(N-1)\cdots(N-n+1)$ muestras distintas; la probabilidad de que una muestra esté formada por n individuos concretos es

$$\frac{n!}{N(N-1)\cdots(N-n+1)} = \frac{1}{\binom{N}{n}}$$

que es igual a la anterior.

Definición

Supongamos una población de N individuos divididos en dos categorías de N_1 y $N_2(= N-N_1)$ individuos cada uno. Se elige una muestra de n individuos de la población (sin reemplazamiento o simultáneamente). La variable aleatoria X que contabiliza el número de individuos de la primera categoría en la muestra se dice que tiene distribución **hipergeométrica** de parámetros N, N_1 y n y se nota

$$X \sim H(N, N_1, n), \qquad n, N_1, N \in \mathbb{N} - \{0\},\ n, N_1 \leq N$$

Función masa de probabilidad

Evidentemente, un valor de la variable X debe ser un número natural (debe incluirse el cero) verificando

$$\text{máx}(0, n-(N-N_1)) \leq x \leq \text{mín}(n, N_1)$$

pues como hay N_1 elementos en la primera subpoblación y $N-N_1$ en la segunda subpoblación se tiene que cumplir

$$\begin{array}{l} x \leq n,\ x \leq N_1 \quad \Rightarrow \quad x \leq \text{mín}(n, N_1) \\ x \geq 0,\ n-x \leq N-N_1 \quad \Rightarrow \quad x \geq \text{máx}(0, n-(N-N_1)) \end{array}$$

La función masa de probabilidad de esta variable aleatoria es

$$\mathrm{P}[X=x] = \frac{\binom{N_1}{x}\binom{N-N_1}{n-x}}{\binom{N}{n}} \qquad \text{máx}(0, n-(N-N_1)) \leq x \leq \text{mín}(n, N_1)$$

Ejemplos y/o aplicaciones

La distribución hipergeométrica se aplica en el control estadístico de calidad de una fabricación en serie. Así pues, si el lote bajo control contiene N_1 elementos buenos y $N_2 = N-N_1$

elementos defectuosos, cuando tomamos una muestra de tamaño n sin reemplazamiento estaremos interesados en saber el número de elementos buenos que han aparecido en la muestra de tamaño n para así determinar la calidad del proceso de fabricación.

En sondeos de opinión también tiene aplicación la distribución hipergeométrica. Podemos realizar una encuesta para intentar conocer si los individuos de una población tienen o no intención de votar en las próximas elecciones de tal manera que el número de individuos, de una muestra sin reemplazamiento, que tienen intención de votar sigue una distribución hipergeométrica.

Algunas de sus características son:

- **Función de distribución**:

$$F(x) = \begin{cases} 0 & x < \text{máx}(0, n-(N-N_1)) \\ \sum_{i=0}^{[x]} \frac{\binom{N_1}{i}\binom{N-N_1}{n-i}}{\binom{N}{n}} & \text{máx}(0, n-(N-N_1)) \le x \le \text{mín}(n, N_1) \\ 1 & x > \text{mín}(n, N_1) \end{cases}$$

- **Función generatriz de momentos**:

  Existe $\forall t \in \mathbb{R}$ porque la variable toma un número finito de valores.

$$M_X(t) = \sum_{x=\text{máx}(0,n-(N-N_1))}^{\text{mín}(n,N_1)} \mathrm{e}^{tx} \frac{\binom{N_1}{x}\binom{N-N_1}{n-x}}{\binom{N}{n}}$$

  No se conoce una forma funcional específica.

- **Media**:

$$\mathrm{E}[X] = n\frac{N_1}{N}$$

- **Varianza**:

$$\mathrm{Var}[X] = \frac{n(N-n)N_1(N-N_1)}{N^2(N-1)}$$

Relación entre la distribución hipergeométrica y la binomial

Sea $X \sim H(N, N_1, n)$ y sea $p = \frac{N_1}{N}$ (probabilidad de que al elegir un individuo al azar sea de la primera categoría). La función masa de probabilidad, la media y la varianza se pueden escribir como

$$\mathrm{P}\{X = x\} = \frac{\binom{Np}{x}\binom{N(1-p)}{n-x}}{\binom{N}{n}}, \quad \text{máx}(0, n - N(1-p)) \leq \text{mín}(n, Np)$$

$$\mathrm{E}[X] = np$$

$$\mathrm{Var}[X] = np(1-p)\frac{N-n}{N-1}$$

Recordemos que la diferencia entre esta distribución y la binomial radica en que la probabilidad de éxito en las n repeticiones de Bernoulli realizadas para seleccionar la muestra no es constante.

Ahora bien, si el tamaño de la población N es muy grande en comparación al tamaño de muestra, las probabilidades de éxito si se realiza reemplazamiento o no son muy parecidas ya que al extraer un individuo (o un número pequeño de individuos) de una población muy grande, las proporciones no varían prácticamente.

De hecho, se observa que $\mathrm{E}[X] = np$ coincide con la de la binomial y la varianza se diferencia en $\frac{N-n}{N-1}$, al que se denomina “factor de corrección para poblaciones finitas", que tiende a 1 si $N \to \infty$.

Además, como probamos a continuación, si $N \to \infty$ las probabilidades de la distribución hipergeométrica $H(N, N_1, n)$ convergen a las de una $B(n, p)$.

Teorema.- Sea X una variable aleatoria con distribución hipergeométrica $H(N, N_1, n)$ y $p = \frac{N_1}{N}$ constante. Entonces

$$\lim_{N \longrightarrow \infty} \frac{\binom{Np}{x}\binom{N(1-p)}{n-x}}{\binom{N}{n}} = \binom{n}{x} p^x (1-p)^{n-x}$$

Es decir, cuando N sea suficientemente grande podemos aproximar probabilidades hipergeométricas por binomiales. En la práctica se considera una buena aproximación cuando $N > 50$ y $n \leq 0{,}1N$, o bien, para $n < 0{,}5N$.

EJERCICIOS

1.- Sea una baraja de 40 cartas. De ella se toma una muestra de 5 cartas sin reemplazamiento. Obtener la probabilidad de obtener al menos dos ases.

Tenemos 40 cartas, de las cuales 4 son ases y se realiza un muestreo de tamaño 5 sin reemplazamiento.

X: Número de ases en la muestra $\sim H(40, 4, 5)$
y se pide

$$\mathrm{P}[X \geq 2] = 1 - \mathrm{P}[X = 0] - \mathrm{P}[X = 1] = 1 - \frac{\binom{4}{0}\binom{36}{5}}{\binom{40}{5}} - \frac{\binom{4}{1}\binom{36}{4}}{\binom{40}{5}} = 0{,}06899$$

2.- Una determinada empresa quiere aumentar su plantilla de vendedores en 20 personas y se presentan 40 personas al proceso de selección. Determinar la probabilidad de que, después de realizar todas las pruebas de selección, entre las 20 personas seleccionadas estén los 10 mejores de las 40 personas que se presentaron.

Consideramos la variable aleatoria X: Número de los mejores vendedores entre los 20 seleccionados.

$$X \sim H(40, 10, 20)$$

Nos piden

$$\mathrm{P}[X = 10] = \frac{\binom{10}{10}\binom{30}{10}}{\binom{40}{20}} = 0{,}000217$$

ANEXO

Demostración de que la función masa de probabilidad de una distribución hipergeométrica suma uno

En efecto, las probabilidades son no negativas y además [2]

$$\sum_{x=\text{máx}(0,n-(N-N_1))}^{\text{mín}(n,N_1)} \text{P}[X=x] = \sum_{x=\text{máx}(0,n-(N-N_1))}^{\text{mín}(n,N_1)} \frac{\binom{N_1}{x}\binom{N-N_1}{n-x}}{\binom{N}{n}} = \frac{\binom{N_1+N-N_1}{n}}{\binom{N}{n}} = \frac{\binom{N}{n}}{\binom{N}{n}} = 1$$

Cálculo de la media de una distribución hipergeométrica

$$EX = \sum_{x=\text{máx}(0,n-(N-N_1)}^{\text{mín}(n,N_1)} x\frac{\binom{N_1}{x}\binom{N-N_1}{n-x}}{\binom{N}{n}} = \frac{1}{\binom{N}{n}} \sum_{x=\text{máx}(1,n-(N-N_1)}^{\text{mín}(n,N_1)} N_1\binom{N_1-1}{x-1}\binom{N-N_1}{n-x} =$$

haciendo $x-1=k$

$$= \frac{N_1}{\binom{N}{n}} \sum_{k=\text{máx}(0,(n-1)-(N-N_1)}^{\text{mín}(n-1,N_1-1)} \binom{N_1-1}{k}\binom{N-N_1}{(n-1)-k} = \frac{N_1}{\binom{N}{n}}\binom{N-1}{n-1} = \frac{nN_1}{N}$$

Cálculo de la varianza de una distribución hipergeométrica

Calculemos en primer lugar el momento no centrado de orden dos; es decir

$$E[X^2] = E[X(X-1)] + E[X]$$

$$E[X(X-1)] = \sum_{x=\text{máx}(2,n-(N-N_1)}^{\text{mín}(n,N_1)} x(x-1)\frac{\binom{N_1}{x}\binom{N-N_1}{n-x}}{\binom{N}{n}} =$$

[2] Usando que dados dos números cualesquiera a y b y un entero positivo n, se verifica

$$\sum_{x=0}^{n} \binom{a}{x}\binom{b}{n-x} = \binom{a+b}{n}$$

aunque realmente, dado que $\binom{a}{x}=0$ si $x>a$ al haber un término nulo en $a(a-1)\cdots(a-x+1)$ se debe tomar $x \leq a$, y dado que $\binom{b}{n-x}=0$ si $n-x>b$ se debe tomar $x \geq n-b$. Por tanto la suma parte de $\text{máx}(0,n-b)$ y llega a $\text{mín}(n,a)$ y realmente se tiene

$$\sum_{x=\text{máx}(0,n-b)}^{\text{mín}(n,a)} \binom{a}{x}\binom{b}{n-x} = \binom{a+b}{n}$$

$$= \frac{N_1(N_1-1)}{\binom{N}{n}} \sum_{x=\text{máx}(2,n-(N-N_1)}^{\text{mín}(n,N_1)} \binom{N_1-2}{x-2}\binom{N-N_1}{n-x} =$$

haciendo $x-2=k$

$$= \frac{N_1(N_1-1)}{\binom{N}{n}} \sum_{k=\text{máx}(0,(n-2)-(N-N_1)}^{\text{mín}(n-2,N_1-2)} \binom{N_1-2}{k}\binom{N-N_1}{(n-2)-k} =$$

$$= \frac{N_1(N_1-1)}{\binom{N}{n}}\binom{N-2}{n-2} = \frac{n(n-1)N_1(N_1-1)}{N(N-1)}$$

$$E[X^2] = \frac{n(n-1)N_1(N_1-1)}{N(N-1)} + \frac{nN_1}{N}$$

de donde se deduce la expresión de la varianza.

$$Var[X] = \frac{n(n-1)N_1(N_1-1)}{N(N-1)} + \frac{nN_1}{N} - \frac{n^2N_1^2}{N^2} = \frac{n(N-n)N_1(N-N_1)}{N^2(N-1)}$$

Demostración del Teorema que relaciona la distribución hipergeométrica con la binomial

Veamos la demostración del resultado expuesto

$$\text{P}[X=x] = \frac{\binom{Np}{x}\binom{N(1-p)}{n-x}}{\binom{N}{n}} = \frac{(Np)!}{x!(Np-x)!}\frac{(N(1-p))!}{(n-x)!(N(1-p)-n+x)!}\frac{n!(N-n)!}{N!} =$$

$$= \frac{n!}{x!(n-x)!}\frac{(Np)!(N(1-p))!(N-n)!}{(Np-x)!(N(1-p)-n+x)!N!} =$$

$$= \binom{n}{x}\frac{Np(Np-1)\cdots(Np-x+1)N(1-p)(N(1-p)-1)\cdots(N(1-p)-n+x+1)}{N(N-1)(N-2)\cdots(N-n+1)}$$

(Esta expresión se deduce directamente de la expresión de la función masa de probabilidad al considerar una extracción sucesiva y sin reemplazamiento).

Tomando límites cuando $N \longrightarrow \infty$, se tiene [3]

$$\lim_{N\to\infty} \text{P}[X=x] = \binom{n}{x}p^x(1-p)^{n-x}$$

que es la función masa de probabilidad de una $B(n,p)$, como queríamos demostrar.

[3] El numerador es un polinomio en N de grado n, cuyo término de mayor grado es $(Np)^x(N(1-p))^{n-x} = N^np^x(1-p)^{n-x}$ y el denominador es también un polinomio en N de grado n, cuyo término de mayor grado es N^n, luego el límite de ese cociente de polinomios del mismo grado es $p^x(1-p)^{n-x}$