TEMAS DE MATEMÁTICAS
(Oposiciones de Secundaria)

TEMA 65

DISTRIBUCIONES DE PROBABILIDAD DE VARIABLE DISCRETA. CARACTERÍSTICAS Y TRATAMIENTO. LAS DISTRIBUCIONES BINOMIAL Y DE POISSON. APLICACIONES.

TEMA 65

DISTRIBUCIONES DE PROBABILIDAD DE VARIABLE DISCRETA. CARACTERÍSTICAS Y TRATAMIENTO. LAS DISTRIBUCIONES BINOMIAL Y DE POISSON. APLICACIONES.

1. INTRODUCCIÓN.

DEF Diremos que X es una variable aleatoria discreta unidimensional si es una variable aleatoria que toma sólo un número finito o infinito numerable de valores del eje x.

Supongamos que X toma únicamente los valores x_1, x_2, ..., x_n, ... con probabilidades $f(x_1)$, $f(x_2)$, ..., $f(x_n)$, ... e imaginemos que A es cualquier subconjunto de los puntos x_1, x_2, ..., x_n, ... La probabilidad P(A) del suceso A (probabilidad de que x esté en A) se define como

$$P(A)=\sum_{x\in A} f(x)$$

donde el sumatorio representa la suma de f(x) para aquellos valores x_i que pertenecen a A.

Así, por ejemplo, P(X=4) indica la probabilidad de que el valor de la variable sea 4, y P(1<X<3) indica la probabilidad de que el valor de la variable aleatoria esté comprendido entre 1 y 3.

Dada una situación experimental, f(x) la tomaremos de tal manera que se ajuste al problema particular que estemos considerando.

DEF Llamaremos a f(x) Función de Densidad de una variable Discreta, si satisface las dos condiciones siguientes:

- $f(x_i)\geq 0 \quad \forall i$
- $\sum f(x_i)=1$

La función de Densidad también se conoce como Función de Cuantía.

Ejemplo. Sea un experimento aleatorio consistente en lanzar cuatro monedas simétricas y registrar el número de caras. Los valores que toma una variable aleatoria X serán 0, 1, 2, 3 y 4. Para calcular la función de densidad f(x), probabilidad de que aparezcan x caras, observamos que el número de formas en que pueden caer las cuatro monedas es 2^4. El número de formas en que pueden aparecer x caras es $\binom{4}{x}$; por tanto,

$$f(x)=\frac{\binom{4}{x}}{2^4} \qquad x=0,1,2,3,4$$

Como

$$f(x)\geq 0 \qquad 0\leq x\leq 4$$

$$\sum_{x=0}^{4} f(x) = \frac{1}{2^4}\sum_{x=0}^{4}\binom{4}{x} = 1$$

Entonces, f(x) es una función de densidad.

Es importante destacar que f(x) da las frecuencias relativas o frecuencias con que se presenta cada uno de los valores de X.

2. FUNCIONES DE CUANTÍA.

El conjunto de resultados posibles de un suceso aleatorio se divide en un cierto número de clases mutuamente excluyentes en relación con determinado atributo. A cada clase se le asocia un valor de la variable aleatoria, o variante X. La función de cuantía es una función que da la probabilidad de que ocurra un valor determinado de x.

La variante X puede describir un atributo, como era el caso anterior del lanzamiento de 4 monedas, o puede ser simplemente el resultado de una codificación. Así, al extraer bolas de una urna, pueden clasificarse según su color, y podríamos definir una variable aleatoria X estableciendo arbitrariamente una correspondencia entre los valores de X y los colores.

La función de cuantía puede ser una expresión matemática, como en el apartado anterior, o bien reducirse a una tabla de valores.

A menudo resulta necesario calcular la probabilidad cuando la variable aleatoria se encuentra en un determinado rango de valores, como por ejemplo, P(X<3) o P(1<X<5). En estos casos, y también en otras situaciones, es conveniente definir una nueva función.

DEF Llamaremos Función de Distribución Acumulativa a F(x), que se define como F(x)=P(X≤x). Es decir

$$F(x) = \sum_{x_i \le x} f(x_i)$$

PROP $P(a<x\leq b) = F(b) - F(a)$

Dem.

Inmediata a partir de la definición.

3. DISTRIBUCIONES MULTIVARIANTES.

Si el resultado de un suceso aleatorio lo podemos clasificar de más de una manera, la función de cuantía es una función de más de una variable. Así, al extraer una carta de una baraja francesa, podríamos caracterizarla según su palo y su numeración. Tendríamos la variable aleatoria X que toma los valores 1, 2, 3 y 4 (representando cada uno a un palo de la baraja) y la variable aleatoria Y que toma los valores del 1 al 13.

La variable aleatoria (X,Y) es bidimensional y la probabilidad de extraer una carta se representa por f(X,Y). Si cada carta tiene la misma probabilidad de ser extraida, la función de cuantía será

$$f(X,Y) = \frac{1}{52} \qquad 1 \le X \le 4 \quad 1 \le Y \le 13$$

Este tipo de funciones pueden representarse sobre un plano. Las probabilidades vienen representadas por segmentos verticales en los puntos (x,y) del plano horizontal en el cual están definidas. En el ejemplo, los segmentos serán de igual altura.

En general, la variable aleatoria k-dimensional (x_1, x_2, ..., x_n) se denomina función de cuantía de la variable aleatoria k-dimensional. Sea A un subconjunto del conjunto de valores de la variable aleatoria, entonces

$$P\big((x_1, x_2, \ldots, x_k) \in A\big) = \sum_A f(x_1, x_2, \ldots, x_k)$$

donde el sumatorio supone la suma que toma la función de cuantía para los valores de la variable que pertenecen al conjunto A.

DEF Sea $x_{i_1}, x_{i_2}, \ldots, x_{i_t}$ cualquier subconjunto de las variables aleatorias discretas x_1, x_2, ..., x_n. La distribución marginal de la variable aleatoria t-dimensional $(x_{i_1}, x_{i_2}, \ldots, x_{i_t})$ es

$$f_{i_1, i_2, \ldots, i_t}(x_{i_1}, x_{i_t}, \ldots, x_{i_t}) = \sum f(x_1, x_2, \ldots, x_n)$$

donde la suma se extiende a todos los valores, a excepción de $x_{i_1}, x_{i_2}, \ldots, x_{i_t}$.

DEF Sean $x_{i_1}, x_{i_2}, \ldots, x_{i_t}$ y $x_{j_1}, x_{j_2}, \ldots, x_{j_s}$ dos subconjuntos disjuntos de las variables aleatorias discretas x_1, x_2, ..., x_n. La distribución condicional de la variable aleatoria t-dimensional ($x_{i_1}, x_{i_2}, \ldots, x_{i_t}$) dado el valor ($x_{j_1}, x_{j_2}, \ldots, x_{j_s}$) es

$$g(x_{i_1}, x_{i_2}, \ldots, x_{i_t} \mid x_{j_1}, x_{j_2}, \ldots, x_{j_s}) = \frac{f_{i_1, i_2, \ldots, i_t, j_1, j_2, \ldots, j_s}(x_{i_1}, x_{i_2}, \ldots, x_{i_t}, x_{j_1}, x_{j_2}, \ldots, x_{j_s})}{f_{j_1, j_2, \ldots, j_s}(x_{j_1}, x_{j_2}, \ldots, x_{j_s})}$$

para todos los valores de x para los cuales no se anula el denominador.

DEF Las variables aleatorias discretas x_1, x_2, ..., x_k son Independientes si

$$f(x_1, x_2, \ldots, x_k) = f(x_1) \cdot f(x_2) \cdot \ldots \cdot f(x_k)$$

para todos los valores en que está definida la variable aleatoria.

Ejemplo Supongamos que extraemos 12 cartas sin reemplazamiento de una baraja francesa. Sea X_1 el número de ases, X_2 el número de doses, X_3 el número de treses y X_4 el número de cuatros. La distribución de estas variables viene expresada por una función de cuatro variables, que es:

$$f(x_1, x_2, x_3, x_4) = \frac{\binom{4}{x_1} \cdot \binom{4}{x_2} \cdot \binom{4}{x_3} \cdot \binom{4}{x_4} \cdot \binom{36}{12 - x_1 - x_2 - x_3 - x_4}}{\binom{52}{12}}$$

siendo el recorrido de cada una de las variables aleatorias de $1 \leq X_i \leq 4$ y verificando la restricción de

$$\sum X_i \leq 12$$

Existe una gran cantidad de distribuciones marginales y condicionales asociadas con esta distribución. Veamos dos ejemplos de distribuciones marginales y otro de distribución condicional.

$$f_{23}(x_3, x_4) = \frac{\binom{4}{x_2} \cdot \binom{4}{x_3} \cdot \binom{44}{12 - x_2 - x_3}}{\binom{52}{12}} \qquad 0 \leq x_i \leq 4$$

$$f_4(x_4) = \frac{\binom{4}{x_4} \cdot \binom{48}{12 - x_4}}{\binom{52}{12}} \qquad 0 \leq x_4 \leq 4$$

$$f(x_2, x_4 \mid x_1, x_3) = \frac{\binom{4}{x_2} \cdot \binom{4}{x_4} \cdot \binom{36}{12 - x_1 - x_2 - x_3 - x_4}}{\binom{44}{12 - x_1 - x_3}} \qquad \begin{matrix} 0 \leq x_i \leq 4 \\ x_2 + x_4 \leq 12 - x_1 - x_3 \end{matrix}$$

4. DISTRIBUCIÓN DE BERNOUILLI.

Llamaremos Prueba de Bernouilli a toda realización de un experimento aleatorio en el que sólo sean posibles dos resultados mutuamente exclusivos (por ejemplo, éxito y fracaso). Por tanto, podemos definir un espacio muestral como

$$\Omega = \{0,1\}$$

donde 0 representa “fracaso” y 1 representa “éxito”. Definamos una variable aleatoria X que pueda tomar sólo los valores anteriores, con probabilidades asociadas

$$P(\text{éxito}) = p \qquad P(\text{fracaso}) = q = 1 - p$$

Es claro que la función de distribución de probabilidades (también llamada de cuantía) de X es:

$$f(x) = P(X=x) = p^x(1-p)^{1-x} \quad \text{para } x=0,1$$

Esta distribución de probabilidades recibe el nombre de Distribución de Bernouilli.

La media y la varianza de una variable aleatoria que posee una distribución de probabilidades de Bernouilli son:

$$E(X) = 0\cdot f(0) + 1\cdot f(1) =00\cdot q + 1\cdot p = p$$

$$Var(X) = E(X^2) - [E(X)]^2 = 0\cdot f(0) + 1\cdot f(1) - p^2 = p - p^2 = p(1-p) = pq$$

5. DISTRIBUCIÓN BINOMIAL.

Supongamos n pruebas independientes entre sí de Bernouilli tales que la probabilidad de éxito, p, se mantiene constante a lo largo de todas ellas. Definimos la variable aleatoria X como el número de veces que sucedió el suceso éxito. Bajo estas condiciones podemos definir

DEF Diremos que X es una variable aleatoria Binomial de parámetros n y p si su función de probabilidad es

$$P(X=x) = \binom{n}{x} p^x q^{n-x} \qquad \text{con } x = 0,1,...,n$$

Veamos como se ha originado esta función de probabilidad. Supongamos que el resultado de la i-ésima prueba forma una variable aleatoria unidimensional que designaremos por X_i, siendo su valor $X_i=0$ si el resultado es un fracaso y $X_i=1$ si es un éxito. Sea $f(x_1,x_2,...,x_n)$ la probabilidad de que la variable aleatoria $X_i=x_i$ con i:1,2,...,n.

Puesto que las n variables aleatorias, todas ellas iguales, son independientes, podemos aplicar la definición de independencia para obtener

$$f(x_1, x_2,..., x_n) = f(x_1)\cdot f(x_2)\cdot\ldots\cdot f(x_n) = p^{x_1}q^{1-x_1}\, p^{x_2}q^{1-x_2}\cdot\ldots\cdot p^{x_n}q^{1-x_n} = p^{\sum x_i} q^{n-\sum x_i}$$

donde 0 y 1 son los valores que toman cada una de las variables x_i.

Evidentemente, para cualquier ordenación de k éxitos y n–k fracasos, la probabilidad es

$$p^k q^{n-k}$$

puesto que el sumatorio será k al haber esa cantidad de 1's. Si ahora tenemos en cuenta el número total de formas de ordenar k 1's en n lugares, obtenemos el número combinatorio

$$\binom{n}{k}$$

siendo entonces la probabilidad de k éxitos el número

$$\binom{n}{k} p^k q^{n-k}$$

La función de distribución de variable aleatoria binomial presenta otras dos variables, p y n, de carácter distinto. Su variación corresponde a distribuciones binomiales diferentes, ya que para una distribución binomial dada, p y n tienen valores fijos. Las variables de este tipo reciben el nombre de parámetros de la distribución.

La función f(x;n,p) representa una familia de distribuciones con dos parámetros, obteniéndose un miembro de esta familia al especificar los valores de p y n.

DEF El parámetro n recibe el nombre de Parámetro Discreto. El parámetro p es el Parámetro Contínuo.

El nombre de n es debido a que sólo puede tomar valores naturales. En cambio, p puede tomar valores dentro del intervalo cerrado [0,1]

En general, la distribución binomial tendrá un valor máximo, el cual pasamos a determinar.

Sea m la parte entera del número (n+1)·p y sea e la parte fraccionaria.

$$m = [(n+1)\cdot p] \qquad y \qquad e = (n+1)\cdot p - [(n+1)\cdot p]$$

PROP El mayor valor de f(x) se obtiene al tomar x = m, recibiendo m el nombre de valor modal o, simplemente, moda de la variable aleatoria X. Si e=0, también se alcanza en x = m–1.

<u>Dem.</u>

Supongamos que el número e, definido anteriormente, es distinto de cero. Tomemos la razón

$$\frac{f(x+1)}{f(x)}$$

Vamos a comprobar que esta razón es inferior a 1 cuando x es mayor o igual a m, y superior a 1 cuando x es inferior a m.

$$\frac{f(x+1)}{f(x)} = \frac{p}{q} \cdot \frac{n-x}{x+1}$$

y si $x \geq m$ tenemos

$$\frac{p}{q} \cdot \frac{n-x}{x+1} \leq \frac{p}{q} \cdot \frac{n-m}{m+1}$$

Sustituyendo m por (n+1)p–e, el segundo miembro puede escribirse como

$$\frac{p}{q}\cdot\frac{n-m}{m+1}=\frac{(n+1)-\left[\dfrac{1-e}{q}\right]}{(n+1)+\left[\dfrac{1-e}{p}\right]}$$

que es menor que la unidad

Si $x < m-1$ tenemos

$$\frac{p}{q}\cdot\frac{n-x}{x+1}>\frac{p}{q}\cdot\frac{n-(m-1)}{m}>\frac{p}{q}\cdot\frac{(n+1)q+e}{(n+1)p-e}>\frac{n+1+\dfrac{e}{q}}{n+1-\dfrac{e}{p}}$$

que es mayor que 1.

Ahora vamos a estudiar el caso que nos hemos dejado pendiente: $x = m-1$, para el cual se verifica:

$$\frac{f(x+1)}{f(x)}=\frac{p}{q}\cdot\frac{n-m+1}{m}=\frac{(n+1)+\dfrac{e}{q}}{(n+1)-\dfrac{e}{p}}$$

que es también superior a 1 si $e\neq 0$. Si $e=0$, la razón es igual a la unidad, verificándose que $f(x+1) = f(x)$, teniendo la distribución dos valores de x en los que alcanza el máximo, $x = m$ y $x = m-1$.

PROP La Media y la Varianza de una variable aleatoria que posee una distribución de probabilidades Binomial es:

1) $E(X) = np$
2) $Var(X) = npq$

<u>Dem.</u>

Para la demostración, vamos a considerar que la variable aleatoria X con distribución binomial se puede dividir en n variables aleatorias independientes, cada una de ellas asociada a una distribución de Bernouilli de probabilidad p. Es decir

$$X = X_1 + X_2 + ... + X_n$$

donde las X_i son variables aleatorias independientes que siguen una distribución de Bernouilli, representando si sucede o no el suceso A en la prueba i-ésima, con i:1,2,...,n. Tomando esperanzas a ambos lados de la igualdad:

$$E(X) = E(X_1 + X_2 + ... + X_n) = E(X_1) + E(X_2) + ... + E(X_n) = p + p + ... + p = np$$

y tomando varianzas

$$Var(X) = Var(X_1 + X_2 + ... + X_n) = Var(X_1) + Var(X_2) + ... + Var(X_n) =$$

$$= pq + pq + ... + pq = npq$$

6. DISTRIBUCIÓN DE POISSON.

DEF Diremos que la variable aleatoria X se distribuye según una distribución de Poisson si la función de distribución es de la forma

$$f(x) = \frac{e^{-\lambda}\lambda^x}{x!} \text{ con } x=0,1,2,3,...$$

siendo λ cualquier número positivo.

Si tenemos en cuenta que la exponencial e^{λ} tiene un desarrollo en serie de potencias

$$e^{\lambda} = 1 + \lambda + \frac{\lambda^2}{2!} + \ldots + \frac{\lambda^x}{x!} + \ldots$$

deducimos que

$$\sum_{x=0}^{\infty} f(x) = 1$$

Esta distribución se suele aplicar en aquellas situaciones en que un gran número de objetos se encuentran distribuidos sobre un gran recinto de considerable extensión. Consideremos un ejemplo que concrete esta situación.

Supongamos un volumen V de un fluido que contiene un gran número N de pequeños organismos. Podemos suponer que estos organismos carecen de instintos sociales y que la probabilidad de que aparezcan en cualquier parte del fluido es la misma para un determinado volumen.

Examinemos una gota de volumen D al microscopio. ¿Cuál es la probabilidad de que se hallen x organismos en la gota?. Se supone que V es mucho mayor de D. Puesto que suponemos que los organismos están distribuidos con probabilidad uniforme por todo el fluido, se deduce que la probabilidad de encontrar uno cualquiera de ellos en D es $\frac{D}{V}$. Y al suponer que no tienen instintos sociales, la presencia de uno de ellos en D no influye en la de cualquiera de los otros. Por tanto, la probabilidad de que haya x organismos en D es

$$\binom{N}{x} \cdot \left(\frac{D}{V}\right)^x \cdot \left(\frac{V-D}{V}\right)^{N-x}$$

Supongamos ahora también que los organismos son tan pequeños que podemos prescindir del espacio que ocupan; los N reunidos no ocuparían parte apreciable del volumen D. La función de Poisson es una aproximación de la expresión anterior, que es simplemente una función binomial en la que $p = \frac{D}{V}$ es un valor muy pequeño.

La distribución de Poisson se obtiene haciendo que V y N tiendan a infinito, de tal modo que la densidad de organismos $d = \dfrac{N}{V}$ permanezca constante. si escribimos de nuevo el producto anterior de la siguiente forma

$$\frac{N(N-1)(N-2)\cdots(N-x+1)}{x!\cdot N^x}\cdot\left(\frac{ND}{V}\right)^x\cdot\left(1-\frac{ND}{NV}\right)^{N-x} =$$

$$= \frac{\left(1-\frac{1}{N}\right)\left(1-\frac{2}{N}\right)\cdots\left(1-\frac{x-1}{N}\right)\cdot(Dd)^x\cdot\left(1-\frac{Dd}{N}\right)^{N-x}}{x!}$$

Podemos ver que el límite de la expresión anterior cuando N tiende a infinito es

$$\frac{e^{-Dd}(Dd)^x}{x!}$$

y si sustituimos Dd por λ obtenemos la función de distribución de Poisson.. Esto nos demuestra que λ es el valor medio de x, ya que D, volumen de la porción examinada, multiplicado por la densidad d, da el promedio esperado en el volumen D.

Hemos visto este ejemplo con cierto detalle porque a menudo se aplica de forma equivocada a datos que no verifican todas las condiciones requeridas. Por ejemplo, no se puede aplicar en el estudio de distribuciones de larvas de insectos en una gran extensión de cultivo, porque los insectos depositan sus huevos en grupos, de modo que en caso de encontrar uno en una pequeña área dada, lo probable es que se encuentren también otros.

PROP La Media y la Varianza de una distribución de Poisson son iguales al parámetro λ.

Dem.

$$E(X) = \sum_{x=0}^{\infty} x\frac{l^x e^{-l}}{x!} = \sum_{x=1}^{\infty} x\frac{l^x e^{-l}}{x!} = l\sum_{x=1}^{\infty}\frac{l^{x-1}e^{-l}}{(x-1)!}$$

Haciendo el cambio de y = x–1 y operando, obtenemos

$$E(X) = le^{-l}\sum_{y=1}^{\infty}\frac{l^y}{y!} = le^{-l}e^{l} = l$$

En cuanto a la varianza, nos vamos a apoyar en su obtención en el cálculo de la esperanza.

$$E[X(X-1)] = \sum_{x=0}^{\infty} x(x-1)\frac{l^x e^{-l}}{x!} = \sum_{x=2}^{\infty} x(x-1)\frac{l^x e^{-l}}{x!} = l^2\sum_{x=2}^{\infty}\frac{l^{x-2}e^{-l}}{(x-2)!}$$

Haciendo el cambio de variable y = x–2 y operando, obtenemos

$$E[X(X-1)] = \lambda^2 \sum_{y=0}^{\infty} \frac{\lambda^y e^{-\lambda}}{y!} = \lambda^2$$

con lo cual obtenemos el valor de la varianza

$$Var(X) = E(X^2) - [E(X)]^2 = E(X^2) - \mu^2 = E[X(X-1)] + \mu - \mu^2 = \mu = \lambda$$

7. AJUSTE DE UNA DISTRIBUCIÓN BINOMIAL A UNA POISSON.

La distribución de Poisson también se obtiene como límite de la binomial cuando el tamaño de muestra n crece, $n\to\infty$, y la probabilidad de éxito tiende a cero, $p\to 0$, de manera que el número medio de "éxitos", $np=\lambda$, permanezca constante. Las probabilidades binomiales las podemos expresar de la forma siguiente:

$$P(X=x) = \binom{n}{x} \cdot \left(\frac{\lambda}{n}\right)^x \cdot \left(1-\frac{\lambda}{n}\right)^{n-x}$$

Tomando límites:

$$\lim_{n\to\infty} P(X=x) = \lim_{n\to\infty}\left[\binom{n}{x} \cdot \left(\frac{\lambda}{n}\right)^x \cdot \left(1-\frac{\lambda}{n}\right)^{n-x}\right] = \frac{\lambda^x}{x!} \lim_{n\to\infty}\left[\frac{n(n-1)\dots(n-x+1)}{\left(1-\frac{\lambda}{n}\right)^x n^x}\left(1-\frac{\lambda}{n}\right)^n\right] =$$

$$= \frac{\lambda^x}{x!} \lim_{n\to\infty}\left[\frac{n(n-1)\dots(n-x+1)}{\left(1-\frac{\lambda}{n}\right)^x n^x}\right] \cdot \lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^n = \frac{\lambda^x}{x!} \cdot 1 \cdot e^{-\lambda}$$

En consecuencia

$$P(X=x) \to \frac{\lambda^x e^{-\lambda}}{x!} \quad \text{para } x = 0, 1, 2, ...$$

llegando así a la distribución de Poisson. La aproximación se puede considerar que es bastante buena cuando se verifican las siguientes condiciones:

1) $n > 25$
2) $p < 0'1$
3) $np = \lambda \leq 5$

Debido a esto, se conoce a la distribución de Poisson como Ley de los Sucesos Raros ya que se presenta en aquellos fenómenos en que estudiemos la ocurrencias de sucesos con probabilidades pequeñas.

8. OTRAS DISTRIBUCIONES DISCRETAS.

8.1. Distribución Multinomial.

Supongamos que un experimento aleatorio puede conducir a k resultados excluyentes y exhaustivos E_1, E_2, ..., E_k, con probabilidades respectivas p_1, p_2, ..., p_k, entonces la distribución de probabilidad de las variables aleatorias X_1, X_2, ..., X_k, que representan los números de ocurrencias para E_1, E_2, ..., E_k, en n ensayos independientes es:

$$f(x_1, x_2, \dots, x_k) = \binom{n}{x_1, x_2, \dots, x_k} p_1^{x_1} p_2^{x_2} \dots p_k^{x_k} = \frac{n!}{x_1! x_2! \dots x_k!} p_1^{x_1} p_2^{x_2} \dots p_k^{x_k}$$

donde

$$\sum_{i=1}^{k} x_i = n \qquad 0 \le p_1 \le 1 \qquad \sum_{i=1}^{k} p_i = 1$$

Si la variable aleatoria X tiene distribución Multinomial, entonces sus Medias, Varianzas y Covarianzas están dadas por las siguientes expresiones:

1) $E(X_i) = \mu_i = np_i$
2) $Var(X_i) = \sigma_i^2 = np_i(1-p_i)$
3) $Cov(X_i,X_j) = \sigma_{ij} = -np_ip_j$ con $i \neq j$

8.2. Distribución Hipergeométrica.

Supongamos que tenemos una población con N unidades estadísticas, de las cuales N_A son de la clase A y el resto, $N-N_A$, de la clase contraria $\overline{A}$. Sea un experimento consistente en tomar, sin reemplazamiento, n unidades de población.

DEF Bajo las condiciones anteriores, llamamos variable aleatoria Hipergeométrica a

X = "número de elementos de la clase A que hay en la muestra"

A la variable aleatoria así definida le corresponde una función de probabilidad dada por

$$P(X = x) = \frac{\binom{N_A}{x}\binom{N - N_A}{n - x}}{\binom{N}{n}}$$

Los parámetros de la distribución son

N Tamaño de la población
N_A Nº de unidades de la clase A
n Tamaño de la muestra sin reemplazamiento

PROP Si la variable aleatoria X tiene una distribución Hipergeométrica, entonces su media y su varianza están dadas por

$$E(X) = m = \frac{nN_A}{N} = np \qquad\qquad Var(X) = s^2 = \frac{N-n}{N-1}np(1-p)$$

donde $p = \dfrac{N_A}{N}$.

<u>Dem.</u>

$$E(X) = \sum_{x=0}^{n} x \frac{\binom{N_A}{x} \cdot \binom{N-N_A}{n-x}}{\binom{N}{n}} = \frac{N_A}{\binom{N}{n}} \sum_{x=1}^{n} \frac{(N_A - 1)!}{(x-1)!(N_A - x)!}\binom{N-N_A}{n-x} =$$

$$= \frac{N_A}{\binom{N}{n}} \sum_{x=1}^{n} \frac{(N_A - 1)!}{(x-1)!(N_A - x)!}\binom{N-N_A}{n-x} = \frac{N_A}{\binom{N}{n}} \sum_{x=1}^{n} \binom{N_A - 1}{x-1}\binom{N-N_A}{n-x}$$

y haciendo el cambio de variable de y = x–1 convertimos la expresión anterior en

$$E(X) = \frac{N_A \sum_{y=0}^{n-1} \binom{N_A - 1}{y} \cdot \binom{N-N_A}{n-y-1}}{\binom{N}{n}}$$

y teniendo en cuenta que $\binom{N}{n} = \dfrac{N}{n}\binom{N-1}{n-1}$ podemos dejar la expresión anterior como

$$E(X) = \frac{N_A}{N} n \sum_{y=0}^{n-1} \frac{\binom{N_A - 1}{y} \cdot \binom{N-N_A}{n-y-1}}{\binom{N-1}{n-1}} = \frac{N_A}{N} n = np$$

ya que el sumatorio es el de todas las probabilidades en el caso de una muestra de tamaño n–1, de un lote con N–1 objetos de los que N_A–1 son de la clase A.

Para el cálculo de la varianza vamos a tener en cuenta lo siguiente:

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

donde

$$E[X(X-1)] = \frac{N_A(N_A - 1)n(n-1)}{N(N-1)}$$

y al operar obtenemos

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

La diferencia más importante entre esta distribución y la binomial es que en esta última las probabilidades permanecen constantes a lo largo de las pruebas (muestreo con reemplazamiento), mientras que en la hipergeométrica varían de prueba a prueba en función de los resultados de las pruebas anteriores (muestreo sin reemplazamiento).

Cuando se verifican las condiciones siguientes:

1) $N \to \infty$
2) $\frac{N_A}{N} = p = cte$
3) n = cte.

entonces podemos aproximar la distribución hipergeométrica por una binomial, ya que puede probarse que en esas condiciones se cumple que:

$$\frac{\binom{N_A}{x} \cdot \binom{N-N_A}{n-x}}{\binom{N}{n}} \to \binom{n}{x} p^x q^{n-x}$$

8.3. Distribución Hipergeométrica Multivariante.

Consideremos una población con N unidades estadísticas repartidas en k clases disjuntas y exhaustivas E_1, E_2, ..., E_k con N_1, N_2, ..., N_k unidades en cada clase respectivamente, y verificándose que $N_1+N_2+ ...+N_k=N$. Si tomamos n unidades estadísticas sin reemplazamiento y contabilizamos cuántas unidades de cada clase nos aparecen, entonces esas k variables discretas constituyen una variable hipergeométrica multivariante.

X_i = Nº de elementos de la clase A_i que hay en la muestra

Estas variables X_i son linealmente dependientes, ya que su suma es evidentemente el tamaño n de la muestra. Su función de probabilidad viene dada por

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_k = x_k) = \frac{\binom{N_1}{x_1}\binom{N_2}{x_2}\cdots\binom{N_k}{x_k}}{\binom{N}{n}}$$

y sus medias y varianzas valen

$$E(X_i) = m_i = np_i$$

$$Var(X_i) = \sigma_i^2 = \frac{N-n}{N-1} np_i(1-p_i)$$

$$Cov(X_i, X_j) = \sigma_{ij} = -\frac{N-n}{N-1} np_i p_j \qquad i \neq j$$

donde $p_i = \dfrac{N_i}{N}$.

Vemos que las covarianzas son negativas, al igual que sucedía con la distribución multinomial. De manera totalmente análoga a la expuesta en el punto anterior, podemos aproximar esta distribución por una binomial, bajo las siguientes condiciones:

1) $N \rightarrow \infty$
2) $\dfrac{N_i}{N} = p_i = cte.$
3) n = cte.

y la aproximación de la distribución hipergeométrica multivariante por una multinomial se debe al hecho de que en esas condiciones se verifica

$$\frac{\binom{N_1}{x_1}\binom{N_2}{x_2}\cdots\binom{N_k}{x_k}}{\binom{N}{n}} \rightarrow \frac{n!}{x_1!x_2!\ldots x_k!} p_1^{x_1} p_2^{x_2} \ldots p_k^{x_k}$$

8.4. Distribución Geométrica o de Pascal.

Supongamos que realizamos un experimento y estamos interesados en la ocurrencia o no de un determinado suceso A. Supongamos también, al igual que hacíamos en la distribución binomial, que realizamos repeticiones del experimento de forma que estas repeticiones son independientes, y que en cada una de las repeticiones las probabilidades de A, p, y de su complementario $\overline{A}$, 1–p, permanecen constantes. En esta situación, repetimos el experimento hasta que ocurre A por primera vez. Podemos definir la variable aleatoria

X = número de repeticiones del experimento necesarias para que tenga lugar la primera ocurrencia de A

la cual, diremos, tiene una distribución Geométrica o de Pascal de parámetro p.

Tengamos en cuenta que aquí nos separamos de las hipótesis mantenidas para la distribución binomial. Allí el número de repeticiones era predeterminado, mientras que aquí es una variable aleatoria. Tenemos que X=k siempre que en las k–1 repeticiones anteriores no haya sucedido A. Entonces, la función de probabilidad es

$$f(x) = pq^{k-1} \qquad para\ k = 1,2,3,\ldots$$

PROP Si la variable aleatoria X tiene una distribución geométrica, entonces su media y su varianza vienen determinadas por

$$E(X) = \frac{1}{p} \qquad\qquad Var(X) = \frac{q}{p^2}$$

Dem.

$$E(X) = \sum_k kpq^{k-1} = p\sum_k \frac{d(q^k)}{dq} = p\frac{d\left(\sum_k q^k\right)}{dq} = p\frac{d}{dq}\left(\frac{q}{1-q}\right) = \frac{p}{(1-q)^2} = \frac{p}{p^2} = \frac{1}{p}$$

y de forma análoga, con la derivada segunda respecto de q podemos razonar para obtener la varianza a través de la esperanza de X(X–1)

$$Var(X) = E[X(X-1)] + E(X) - E(X)^2 = \frac{q}{p^2}$$

8.5. Distribución Binomial Negativa.

Esta distribución es una generalización de la anterior. Para obtenerla, y bajo las mismas condiciones que para la distribución geométrica, supongamos que un experimento se repite hasta que un suceso particular A se repite r veces. Sea la variable aleatoria

X = número de repeticiones necesarias para que A suceda exactamente r veces.

Esta distribución recibe el nombre de Distribución Binomial Negativa de Parámetros r y p.

Es evidente que para r=1 obtenemos la distribución geométrica de parámetro p. El suceso X=K ocurre si y sólo si el suceso A ocurre en la k-ésima repetición, verificándose que A ya sucedió r-1 veces en los k-1 experimentos anteriores. Por ello, la probabilidad de este suceso es

$$P(X = k) = \binom{k-1}{r-1} p^r q^{k-r} \quad \text{para k = r, r+1, ...}$$

PROP La media y la varianza de una distribución binomial negativa son

$$E(X) = \frac{r}{p} \qquad Var(X) = \frac{rq}{p^2}$$

Dem.

Para comprobarlo, definamos las siguientes variables aleatorias.

Z_1 = Nº de repeticiones necesarias hasta la primera ocurrencia de A

Z_2 = Nº de repeticiones necesarias desde la primera ocurrencia de A hasta incluir la segunda ocurrencia de A.

....

Z_r = Nº de repeticiones necesarias desde la ocurrencia r–1 de A hasta incluir la r-ésima ocurrencia de A.

Todas las Z_i son variables aleatorias independientes, teniendo cada una de ellas una distribución geométrica de parámetro p. Además, la variable aleatoria X es suma de todas ellas. Por tanto

$$E(X) = E(Z_1 + Z_2 + \ldots + Z_r) = E(Z_1) + E(Z_2) + \ldots + E(Z_r) = \frac{1}{p} + \frac{1}{p} + \ldots + \frac{1}{p} = \frac{r}{p}$$

de manera análoga, tomando varianzas

$$Var(X) = Var(Z_1 + Z_2 + \ldots + Z_r) = Var(Z_1) + Var(Z_2) + \ldots + Var(Z_r) = \frac{rq}{p^2}$$

BIBLIOGRAFÍA RECOMENDADA.

Estadística Teórica. Aut. J.M.Doblado y M.C. Nieto. Edit. UNED

Introducción a la Estadística Teórica. Aut.: G Arnáiz. Edit.: Lex Nova

Estadística Teórica y Aplicada. Aut.: A. Nortes. Edit.: S. Rodríguez.

Introducción a la Probabilidad y la Medida (I). Aut.: P Zoroa y N. Zoroa. Edit.: Maior DM.

Introducción a la Teoría de la Estadística. Aut.: Mood, Graybill. Edit.: Aguilar.