ACERCA DE LA SUMA DE VARIABLES INDEPENDIENTES CON DISTRIBUCIÓN BERNOULLI

Caviezel, Pablo
Facultad de Ciencias Económicas, Universidad de Buenos Aires
pcaviezel@economicas.uba.ar
Especialidad: Estadística
Palabras clave: Binomial – Bernoulli – Convolución – Distribución

Resumen

Habitualmente se presenta la distribución binomial como distribución asociada a una variable aleatoria que representa el número de elementos que poseen determinado atributo, siendo éste dicotómico y de presencia definida. El experimento que deriva en el uso de tal distribución consiste en seleccionar en forma aleatoria un número determinado de unidades de un conjunto referencial y analizar el número de elementos seleccionados que poseen el atributo. Para la utilización del modelo binomial dos supuestos son necesarios: i) independencia en el número de elementos que compone el conjunto elegido al azar y ii) idéntica probabilidad de éxito para cada uno de los elementos que componen el conjunto referencial. Surge de esta manera la distribución binomial como suma de variables aleatorias Bernoulli de idéntico parámetro «p».

Se propone en este trabajo levantar el segundo supuesto y considerar un esquema dicotómico donde no necesariamente la probabilidad de éxito «p» – así usualmente llamada en este tipo de esquemas – resulta idéntica para todos los elementos. En términos matemáticos se avanza en la búsqueda de la función de probabilidad y de los momentos de una suma de variables aleatorias Bernoulli, no todas de igual parámetro.

Esta generalización, no obstante, no debe confundirse con la distribución multinomial que avanza en el sentido de proponer una generalización del modelo binomial donde el atributo puede tener más de dos categorías posibles de respuesta.

1. Introducción

« Estudios recientes demuestran que 85 % de las señoras inglesas toman más de una taza de té por reunión, 75 % de las señoras francesas toman más de una taza de té y sólo el 60 % de las señoras alemanas toman más de una taza de té en cualquier reunión. Suponga que seis amigas se reúnen a tomar el té: dos de ellas inglesas, tres de ellas alemanas y la restante francesa. ¿Qué probabilidad hay de que exactamente cuatro de ellas tomen más de una taza de té? ¿Y de que a lo sumo cuatro de ellas tomen más de una taza de té? Calcule asimismo esperanza y varianza de la variable aleatoria asociada al problema. »

Ninguno de los modelos que usualmente se ven en un curso de Estadística clásica resuelve el problema planteado. La propuesta del trabajo consiste en encontrar una distribución adecuada para la variable aleatoria que resuelve el planteo y caracterizarla, tanto desde el punto de vista de sus probabilidades sino también en lo que respecta a la obtención de sus momentos absolutos.

2. Las variables aleatorias de enumeración

Comenzamos definiendo como *variable aleatoria de enumeración* a aquella variable aleatoria discreta que se utiliza para enumerar o contar elementos. El dominio de esta variable aleatoria está compuesto por un número finito o infinito de números naturales pero siempre empezando su recorrido por el valor cero que implica ausencia de elementos.

Distinguimos entonces:

Variable aleatoria de enumeración con dominio finito: $X = \{0;1;2;3;....;n-1;n\}$

Variable aleatoria de enumeración con dominio infinito: $X = \{0;1;2;3;....\}$

En este trabajo trataremos con variables aleatorias de enumeración con dominio finito y por lo tanto todas las propiedades y generalizaciones que se demuestren se harán sobre este esquema.

2.1. Su función generatriz de probabilidades

Una variable aleatoria X de enumeración con dominio finito, tendrá como función generatriz de probabilidades (en adelante FGP) a la función continua $G_x(t)$ sí y sólo sí:

$$G_x(t) = P(X=0) + tP(X=1) + t^2P(X=2) + t^3P(X=3) + ... + t^rP(X=r) + ... + t^nP(X=n)$$

Es decir; $G_X(t) = \sum_{r=0}^{n} t^r P(X = r) = E(t^X)$

Es importante aclarar que el coeficiente que acompaña a la potencia r-ésima de la variable continua t en la FGP corresponde a la probabilidad de que la variable aleatoria de enumeración con dominio finito tome exactamente el valor r; es decir: $P(X = r)$

Tres propiedades útiles que utilizaremos en este trabajo son las siguientes:

Propiedad 1

Existe una relación unívoca y biyectiva entre una distribución de probabilidad y su FGP en el sentido que para cada distribución existe una única FGP y, viceversa, cada FGP describe unívocamente a una sola distribución de probabilidad.

Propiedad 2

Las derivadas sucesivas de la FGP valuadas en $t = 1$ proporcionan el valor de los momentos factoriales; es decir:

$G_X^{(k)}(1) = E\left[X(X-1)(X-2)...(X-k+1)\right]$ donde k representa el orden de derivación.

Interesa puntualmente las siguientes relaciones:

$G'_X(1) = E(X)$

$G''_X(1) = E\left[X(X-1)\right] = E(X^2 - X) = E(X^2) - E(X)$

De donde se desprende una forma adicional para obtener la varianza de la variable aleatoria:

$Var(X) = E(X^2) - E^2(X) = G''_X(1) + G'_X(1) - \left[G'_X(1)\right]^2$

Propiedad 3

La FGP de una suma de variables aleatorias independientes es igual al producto de las funciones generatrices de probabilidades individuales de las variables aleatorias.

Es decir:

Si $R = X_1 + X_2 + ... + X_n$ y además todas las variables aleatorias son independientes, entonces se cumplirá que $G_R(t) = G_{X_1}(t) \cdot G_{X_2}(t) \cdot G_{X_3}(t) \cdot ... \cdot G_{X_n}(t)$

En efecto, asumiendo independencia:

$G_R(t) = E(t^R) = E(t^{X_1 + X_2 + ... + X_n}) = E(t^{X_1} t^{X_2} ... t^{X_n}) = E(t^{X_1}) E(t^{X_2}) ... E(t^{X_n}) =$
$G_{X_1}(t) \cdot G_{X_2}(t) \cdot ... \cdot G_{X_n}(t)$

3. La variable aleatoria Bernoulli

Asumamos un experimento con únicamente dos resultados posibles (éxito y fracaso) y designemos con la letra «p» a la probabilidad de que el experimento resulte exitoso. La variable aleatoria X que asigna como valor numérico "0" al fracaso y "1" al éxito tiene distribución Bernoulli y las siguientes características:

Dominio: $X = \{0;1\}$

Función de probabilidad: $P(X = r) = p^r (1-p)^{1-r}$ con $0 < p < 1$

r	P(X=r)
0	1-p
1	p

$E(X) = p$

$E(X^2) = p^2$

$Var(X) = p^2 - p = p(1-p)$

$G_x(t) = P(X = 0) + P(X = 1)t = (1-p) + pt = 1 + (t-1)p$

Es decir:

☛ $G_x(t) = 1 + (t-1)p$

4. Suma de «n» variables aleatorias independientes con distribución Bernoulli

Presentamos la variable aleatoria R como suma de variables aleatorias independientes X_i, cada una de ellas con distribución Bernoulli.

$R = X_1 + X_2 + X_3 + ... + X_n$

Dominio: $R = \{0;1;2;...n\}$

$E(R) = E(X_1 + X_2 + ... + X_n) = E(X_1) + E(X_2) + ... + E(X_n)$

$E(R^2) = E(X_1{}^2 + X_2{}^2 + ... + X_n{}^2) = E(X_1{}^2) + E(X_2{}^2) + ... + E(X_n{}^2)$

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

$G_R(t) = G_{X_1}(t) \cdot G_{X_2}(t) \cdot G_{X_3}(t) \cdot ... \cdot G_{X_n}(t)$

Para simplificar su estudio, vamos a considerar un primer caso donde todas las variables aleatorias Bernoulli tienen el mismo parámetro «p» y luego vamos a considerar que no todos los parámetros toman el mismo valor.

4.1. Suma de «n» variables aleatorias independientes con distribución Bernoulli con idéntico «p»

En este caso, nuestra variable aleatoria R se dice que tiene distribución binomial con parámetros «n» y «p»

Y, reemplazando en el caso general resulta: $R = X_1 + X_2 + X_3 + ... + X_n$

$$E(R) = p + p + ... + p = np$$

$$E(R^2) = p^2 + p^2 + ... + p^2 = np^2$$

$$Var(R) = np^2 - np = np(1-p)$$

$$G_R(t) = \left[(1-p) + pt\right]^n$$

Aplicando en este caso el desarrollo del Binomio de Newton resulta:

$$G_R(t) = \left[(1-p) + pt\right]^n = \sum_{r=0}^{n} \binom{n}{r} (1-p)^{n-r} (pt)^r = \sum_{k=0}^{n} \left[\binom{n}{r} (1-p)^{n-r} p^r \right] t^r$$

de donde se desprende, por definición de la FGP que:

$P(R = r) = \binom{n}{r} p^r (1-p)^{n-r}$ que es la función de probabilidad de la Distribución Binomial

4.2. Suma de «n» variables aleatorias independientes con distribución Bernoulli

Este punto, eje central del trabajo en cuestión, propone que nuestra variable aleatoria R está formada por la suma de «n» variables aleatorias Bernoulli de parámetros igual a $p_1; p_2; ... p_n$ respectivamente, no necesariamente iguales. Y, reemplazando en el caso general resulta: $R = X_1 + X_2 + X_3 + ... + X_n$

$E(R) = p_1 + p_2 + ... + p_n$ con $0 < p_i < 1$ para todo i

$$E(R^2) = {p_1}^2 + {p_2}^2 + ... + {p_n}^2$$

$$Var(R) = p_1(1-p_1) + p_2(1-p_2) + ... + p_n(1-p_n)$$

$$G_R(t) = \left[1 + (t-1)p_1\right]\left[1 + (t-1)p_2\right]...\left[1 + (t-1)p_n\right]$$

Antes de encontrar la forma general de la función de probabilidad vamos a calcular manualmente las tablas de probabilidad para los casos en que se suman dos y tres variables aleatorias Bernoulli. Como siempre, recurrimos al cálculo de probabilidades correspondiente a variables aleatorias independientes.

4.2.1. Suma de dos variables aleatorias independientes con distribución Bernoulli

$R = X_1 + X_2$ Dominio $R = \{0;1;2\}$

- $P(R=0) = (1-p_1)(1-p_2) = 1-(p_1+p_2)+p_1p_2$
- $P(R=1) = p_1(1-p_2)+(1-p_1)p_2 = p_1 - p_1p_2 + p_2 - p_1p_2 = (p_1+p_2)-2p_1p_2$
- $P(R=2) = p_1p_2$

4.2.2. Suma de tres variables aleatorias independientes con distribución Bernoulli

$R = X_1 + X_2 + X_3$ Dominio $R = \{0;1;2;3\}$

- $P(R=0) = (1-p\ )(1-p_2)(1-p_3) = [1-(p_1+p_2)+p_1p_2](1-p_3) =$

$1-(p_1+p_2)+p_1p_2-p_3+(p_1p_3+p_2p_3)-p_1p_2p_3 =$

$1-(p_1+p_2+p_3)+(p_1p_2+p_1p_3+p_2p_3)-p_1p_2p_3$

- $P(R=1) = p_1(1-p_2)(1-p_3)+(1-p_1)p_2(1-p_3)+(1-p_1)(1-p_2)p_3 =$

$p_1[1-(p_2+p_3)+p_2p_3]+p_2[1-(p_1+p_3)+p_1p_3]+p_3[1-(p_1+p_2)+p_1p_2] =$

$p_1-(p_1p_2+p_1p_3)+p_1p_2p_3+p_2-(p_1p_2+p_2p_3)+p_1p_2p_3+p_3-(p_1p_3+p_2p_3)+p_1p_2p_3 =$

$(p_1+p_2+p_3)-2(p_1p_2+p_1p_3+p_2p_3)+3p_1p_2p_3$

- $P(R=2) = p_1p_2(1-p_3)+p_1(1-p_2)p_3+(1-p_1)p_2p_3 =$

$p_1p_2-p_1p_2p_3+p_1p_3-p_1p_2p_3+p_2p_3-p_1p_2p_3 =$

$(p_1p_2+p_1p_3+p_2p_3)-3p_1p_2p_3$

- $P(R=3) = p_1p_2p_3$

Armamos las siguientes tablas para sintetizar la información obtenida:

r	P(R=r)
0	$1-(p_1+p_2)+p_1p_2$
1	$(p_1+p_2)-2p_1p_2$
2	p_1p_2

r	P(R=r)
0	$1-(p_1+p_2+p_3)+(p_1p_2+p_1p_3+p_2p_3)-p_1p_2p_3$
1	$(p_1+p_2+p_3)-2(p_1p_2+p_1p_3+p_2p_3)+3p_1p_2p_3$
2	$(p_1p_2+p_1p_3+p_2p_3)-3p_1p_2p_3$
3	$p_1p_2p_3$

4.2.3. La generalización a través de operadores

Las expresiones anteriores se reducen notoriamente si introducimos un símbolo operador estadístico Z que se define de la siguiente manera:

Para n = 2 Para n = 3

$Z = p_1 + p_2$ $Z = p_1 + p_2 + p_3$

$Z^2 = p_1p_2$ $Z^2 = p_1p_2 + p_1p_3 + p_2p_3$

$Z^3 = p_1p_2p_3$

Debe quedar claro que los superíndices refieren al orden del operador y no son potencias algebraicas más que en el sentido aplicable para los símbolos operadores usuales. En general, para cualquier valor de n el símbolo operador Z se define como:

$$Z = \sum_{i=1}^{n} p_i$$

$$Z^2 = \sum_{i=1}^{n} \sum_{\substack{j=1 \\ i \neq j}}^{i} p_i p_j$$

$$Z^3 = \sum_{i=1}^{n} \sum_{\substack{j=1 \\ i \neq j}}^{i} \sum_{\substack{k=1 \\ i \neq j \\ j \neq k}}^{j} p_i p_j p_k$$

.....

$Z^n = p_1p_2p_3....p_n$ y se define $Z^{m+n} = 0$ para todo $m > 0$

Es entonces que nuestras tablas se pueden rescribir utilizando el operador Z

r	P(R=r)
0	$1-Z+Z^2$
1	$Z-2Z^2$
2	Z^2

R	P(R=r)
0	$1-Z+Z^2-Z^3$
1	$Z-2Z^2+3Z^3$
2	Z^2-3Z^3
3	Z^3

La generalización para la probabilidad de que la variable aleatoria R tome algún valor r será el término que acompañe a la potencia r-ésima de t en el desarrollo de la FGP:

$$G_R(t)=[1+(t-1)p_1][1+(t-1)p_2]...[1+(t-1)p_n]$$

Nos preguntamos entonces, ¿qué término acompaña a t^r ?

Para encontrarlo, hacemos la distributiva de los n factores de la función generatriz, siempre expresando los resultados en potencias de $t-1$:

$$G_R(t)=1$$
$$+(t-1)(p_1+p_2+...+p_n)$$
$$+(t-1)^2(p_1p_2+p_1p_3+....+p_{n-1}p_n)$$
$$+(t-1)^3(p_1p_2p_3+p_1p_2p_4+....+p_{n-2}p_{n-1}p_n)$$
$$+.....$$
$$+(t-1)^n p_1p_2p_3....p_n$$

Expresión que resulta muy cómodo escribir en términos del operador Z:

$G_R(t)=1+(t-1)Z+(t-1)^2Z^2+...+(t-1)^rZ^r+(t-1)^{r+1}Z^{r+1}+....+(t-1)^nZ^n$ desde donde aplicando la propiedad 2 enunciada en la página 3 pueden obtenerse fácilmente los primeros momentos factoriales y a partir de allí su esperanza y varianza. En efecto:

$$G'_R(1)=Z$$

$$G''_R(1)=2Z^2$$

y luego $E(R)=Z \quad Var(R)=2Z^2+Z(1-Z)$

No hay que perder de vista que buscamos el coeficiente que acompaña a t^r y que se va a encontrar en aquellos términos que tienen potencia igual o superior a $(t-1)^r$. Desarrollando, a través del Binomio de Newton cada uno de esos términos y localizando el factor que acompaña a t^r en cada caso resulta:

$$(t-1)^r Z^r = \left[\sum_{s=0}^{r}\binom{r}{s}t^{r-s}(-1)^s\right]Z^r$$ el término que acompaña a t^r corresponde al valor $s=0$

$$+(t-1)^{r+1}Z^{r+1} = \left[\sum_{s=0}^{r+1}\binom{r+1}{s}t^{r+1-s}(-1)^s\right]Z^{r+1}$$ el término que acompaña a t^r corresponde al valor $s=1$

$$+(t-1)^{r+2}Z^{r+2} = \left[\sum_{s=0}^{r+2}\binom{r+2}{s}t^{r+2-s}(-1)^s\right]Z^{r+2}$$ el término que acompaña a t^r corresponde al valor $s=2$

y así sucesivamente. Es decir que en la FGP tendremos:

$$G_R(t) = ... + \left[\binom{r}{0}(-1)^0 Z^r + \binom{r+1}{1}(-1)^1 Z^{r+1} + \binom{r+2}{2}(-1)^2 Z^{r+2} + ...\right]t^r + ...$$

Extrayendo factor común Z^r :

$$G_R(t) = ... + Z^r\left[\binom{r}{0} - \binom{r+1}{1}Z + \binom{r+2}{2}Z^2 - ....\right]t^r + ...$$

De lo que se concluye que: $P(R=r) = Z^r\left[\binom{r}{0} - \binom{r+1}{1}Z + \binom{r+2}{2}Z^2 - ....\right]$

Pero recordando la convergencia de la siguiente serie, fácilmente demostrable:

$$\left(\frac{1}{z+1}\right)^r = 1 - rz + \binom{r+1}{2}z^2 + \binom{r+2}{3}z^3 - ... = \sum_{s=0}^{\infty}\binom{r+s-1}{s}(-z)^s$$

$$\left(\frac{1}{z+1}\right)^{r+1} = \sum_{s=0}^{\infty}\binom{r+s}{s}(-z)^s = \binom{r}{0} - \binom{r+1}{1}Z + \binom{r+2}{2}Z^2 - ....$$

resulta evidente que

$$\boxed{P(R=r) = Z^r\left(\frac{1}{1+z}\right)^{r+1}}$$

donde se debe tener en cuenta que las potencias de Z superiores a n valen cero.

Si quisiéramos obtener la probabilidad de que R tome al menos el valor r entonces basta resolver la suma:

$$P(R \geq r) = \sum_{s=r}^{\infty} Z^s \left(\frac{1}{1+z}\right)^{s+1} = \frac{1}{1+z}\sum_{s=r}^{\infty}\left(\frac{z}{1+z}\right)^s = \frac{1}{(1+z)}\left[\left(\frac{z}{1+z}\right)^r + \left(\frac{z}{1+z}\right)^{r+1} + \left(\frac{z}{1+z}\right)^{r+2} + \ldots\right]$$

$$= \frac{1}{1+z}\left(\frac{z}{1+z}\right)^r \sum_{s=0}^{\infty}\left(\frac{z}{1+z}\right)^s$$ donde la suma infinita representa una serie geométrica convergente, por lo que:

$$P(R \geq r) = \frac{1}{(1+z)}\frac{1}{1-\frac{z}{1+z}}\left(\frac{z}{1+z}\right)^r$$ e inmediatamente se deduce: $\boxed{P(R \geq r) = \left(\frac{z}{1+z}\right)^r}$

Aprendidas estas fórmulas el lector encontrará entonces la solución al problema planteado en la introducción y podrá determinar que la probabilidad de que exactamente cuatro de las mujeres tomen al menos una taza de té es 0,334125 mientras que la probabilidad de que a lo sumo cuatro de ellas tomen al menos una taza de té es 0,56854. Además, el número esperado de mujeres en la reunión que toma más de una taza de té es 4,25 y su varianza 28,7875.

Referencias bibliográficas

- CASAS SÁNCHEZ, J. M. (1996): Inferencia estadística para economía y administración de empresas. Madrid. Editorial Centro de Estudios Ramón Areces, S.A.

- LANDRO, A (2010). Acerca de la probabilidad – Parte I . Buenos Aires. Ediciones Cooperativas.

- LEVIN, R. (2004). Estadística para administración y economía. México. Editorial Prentice Hall

- MILLER, M (2012). Mathematics and Statistics for financial risk management . Nueva Jersey. John Wiley & Sons.

- WEBSTER, A.L. (2000). Estadística aplicada a los negocios y la economía. Madrid. Editorial Mc Graw Hill.