TEC Tecnológico de Costa Rica

Distribuciones de Probabilidad Discretas: Binomial y geométrica.

Escuela de Matemática

Semana 8

Giovanni Sanabria Brenes

Luis Ernesto Carrera Retana

Erick Chacón Vargas

Mario Marín Sánchez

Rebeca Solís Ortega

1. Distribución de Probabilidad Binomial

Existen diversos experimentos aleatorios para los cuales el espacio muestral solamente admite dos valores:

- Cualitativos, ocurre cuando se cumple o no se alguna condición.
- Cuantitativos, ocurre cuando se da solamente un éxito o un fracaso (1 o 0).

Para este tipo de distribución no hay diferencia entre estos dos tipos, así que basta con que los consideremos como éxito y fracaso, con valores asignados de 1 y 0, respectivamente.

Cuando un experimento como éste se ejecuta una sola vez se dice que es un ensayo tipo *Bernoulli*, la variable aleatoria solo toma dos valores que corresponden con $X(\text{éxito}) = 1$ y $X(fracaso) = 0$ y la distribución de probabilidad también es bastante simplificada con valores $f_X(1) = p$, y $f_X(0) = 1 - p$.

Dado un experimento que consiste de una secuencia de n ensayos independientes tipo Bernoulli, donde la probabilidad, p, de éxito no cambia entre ensayo y ensayo del experimento. Si la variable aleatoria que interesa cuantificar es el número de éxitos en los n ensayos, la variable recibe el nombre de *Binomial.*

Por ejemplo:

1. Se lanza una moneda el aire en 8 ocasiones y se toma como variable aleatoria el número de veces que cae corona. Aquí éxito es que salga corona, por lo que la variable: número de éxitos (escudos) en 8 intentos, es una variable binomial.
2. Se registran los próximos 100 nacimientos en un hospital y se toma como variable aleatoria el número de mujeres que nacen. Aquí éxito es que el nacimiento sea mujer. Por lo tanto la variable, número de éxitos (nacimientos femeninos) en 100 intentos, es una variable binomial.

Definición 1.

Un experimento se clasifica como Binomial si cumple las siguientes propiedades:

1. El experimento consiste de n repeticiones de otro experimento(ensayo de Bernoulli).
2. Cada uno de los intentos es idéntico y puede resultar en uno de dos posibles resultados, éxito E con probabilidad p o fracaso F con probabilidad $1 - p$. Esta probabilidad se mantiene constante entre repeticiones.
3. Las repeticiones de los intentos son independientes, esto es el resultado de una de las pruebas no incide sobre el resultado de las otras.
4. La variable aleatoria observada X es el número de éxitos obtenidos en los n intentos.
   Si X sigue una distribución binomial se suele escribir que

$$X \sim B(n, p)$$

En una variable binomial se tienen por parámetros la probabilidad de éxito en cada ensayo, p, y el total de ejecuciones, n, y como variable aleatoria X, el número de éxitos digamos x. La distribución de probabilidad la denotaremos por $b(x;n,p)$. Para la distribución de probabilidad acumulada usamos la notación $B(x;n,p)$.

Teorema 1.

Si X es una variable aleatoria discreta que sigue una distribución de probabilidad Binomial de parámetros n y p entonces se tiene:

1. $Rango\, X = \{0,1,2,\ldots,n\}$,
2. $b(x;n,p) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & \text{si } x = 0,1,2\ldots,n \\ 0 & \text{en cualquier otro caso} \end{cases}$,
3. $B(x;n,p) = \begin{cases} 0 & \text{si } x < 0 \\ \sum_{i=0}^{r} \binom{n}{i} p^i (1-p)^{n-i} & \text{si } r \leq x < r+1 \\ 1 & \text{si } x \geq n \end{cases}$.

La primera es inmediata debido a que X es el número de éxitos en n intentos. Los posibles valores de X van de 0 a n.

La demostración de la segunda de las aseveraciones se basa en que

$$b(x;n,p) = P(X = x)$$

Es decir, se quiere hallar la probabilidad que el número de éxitos X sea x en n intentos. Resulta que la ocurrencia de exactamente x éxitos es la conjunción de que en x cualesquiera de los ensayos ocurra éxito y en los $n-x$ restantes ocurra fallo. Esto equivale a la probabilidad de que se forman palabras con x letras E (una por cada éxito) y $n-x$ letras F (una por cada fracaso. Por la regla del producto e independencia, la probabilidad de forma una de estas palabras es siempre igual a

$$P(\underbrace{EE\ldots E}_{x}\underbrace{FF\ldots F}_{n-x}) = \underbrace{P(E)\ldots P(E)}_{x}\underbrace{P(F)\ldots P(F)}_{n-x} = p^x(1-p)^{n-x}$$

Como cada palabra de esta forma tiene tiene una probabilidad $p^x(1-p)^{n-x}$ y hay $C(n,x)$ palabras entonces:

$$b(x;n,p) = P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$$

La tercera parte es simplemente una adaptación de la definición de la función de probabilidad acumulada.

Vale la pena destacar que este tipo de cálculos es bastante laborioso, no obstante se dispone de tablas que resumen algunos de los valores más frecuentes.

Ejemplo 1.

Veinte palomas vuelan hacia 3 nidos. Cada una de ellas se ubicará en forma aleatoria en alguno de los nidos, además una paloma se ubica en forma independiente de lo que hicieron o harán las otras. Calcule las siguientes probabilidades.

1. Exactamente 4 palomas se ubican en el primer nido.
2. A lo sumo cuatro palomas se ubican en el primer nido.
3. Al menos cuatro palomas se ubican en el primer nido.

Solución

Dadas las condiciones indicadas, el arribo de cada paloma a algún nido es un ensayo de Bernoulli donde éxito es que la paloma se ubique en el nido 1 y fracaso es que no lo haga. La probabilidad de éxito es $1/3$ y la variable aleatoria que indica el número de palomas que quedan ubicadas en el primer nido sigue una distribución tipo binomial: $b(x;20,1/3)$, y las respuestas a los problemas son:

1. $P[X=4] = b(4;20,1/3) = \binom{20}{4}\left(\frac{1}{3}\right)^4\left(\frac{2}{3}\right)^{16}$
2. A lo sumo cuatro palomas significa ninguna, una, dos, tres o cuatro es decir:
$$b(0;20,1/3)+b(1;20,1/3)+b(2;20,1/3)+b(3;20,1/3)+b(4;20,1/3) = \sum_{i=0}^{4}\binom{20}{i}\left(\frac{1}{3}\right)^i\left(\frac{2}{3}\right)^{20-i},$$
por lo tanto:
$$P[X \leq 4] = 0.1515.$$
3. Al menos cuatro puede verse como el complemento de a lo sumo tres, por tanto la respuesta puede ser
$$1 - P[X \leq 3] = 1 - B(4;20,1/3).$$

Ejemplo 2.

Suponga que el 20 por ciento de los componentes fabricados por una planta no pasan un control de calidad. Cuál es la probabilidad de que en una muestra de 15 componentes consecutivos a lo sumo 8 no pasen la prueba.

Solución

Este es un experimento de tipo binomial, que tiene 15 repeticiones del mismo experimento, con una probabilidad de éxito de 0.2. Si Y es la variable aleatoria que indica el número de componentes defectuosos, interesa calcular el valor de la expresión:

$$P[Y \leq 8] = \sum_{k=0}^{8}\binom{15}{k}(0.2)^k(0.8)^{15-k} = B(8;15,0.2),$$

recurriendo a tablas se obtiene que la probabilidad del evento indicado es 0.9992.

Sobre ese mismo ejemplo:

- La probabilidad que exactamente 8 componentes no pasen la prueba:

$$b(8,15,0.2) = B(8,15,0.2) - B(7,15,0.2) = 0.9992 - 0.9958.$$

- la probabilidad que fallen al menos 8 es:

$$P[Y \geq 8] = 1 - P[Y \leq 7] = 1 - B(7,15.0.2) = 1 - 0.9958.$$

Algunas veces se acepta que un experimento se porta como un binomial, aunque cumpla en forma parcial las reglas citadas en la definición. Por ejemplo en problemas de elección sin reposición la probabilidad de cada experimento está influenciada por los resultados de los anteriores. No obstante cuando el número de intentos es relativamente pequeño respecto al espacio muestral el comportamiento de la variable aleatoria puede aproximarse como si fuera binomial.

Ejemplo 3.

Suponga que en una ciudad viven un millón de personas de los cuales sólo 800 000 son nativos de la ciudad. Si se toma una muestra de 10 ciudadanos al azar cual es la probabilidad de que a lo sumo dos de ellos no sean nativos.

Solución

Si bien la secuencia de 10 experimentos consecutivos de escoger un ciudadano y que no sea nativo tienen probabilidades diferentes también es cierto que estas probabilidades prácticamente son iguales, en un caso como este podemos asumir que el comportamiento de la variable aleatoria Y que indica el número de no nativos en la muestra se aproxima por una binomial. Así la probabilidad solicitada es:

$$P[Y \leq 2] = \sum_{k=0}^{2} \begin{pmatrix} 10 \\ k \end{pmatrix} (0.2)^k (0.8)^{10-k} = B(2;10,0.2),$$

recurriendo a tablas se obtiene que la probabilidad del evento indicado es 0.6778.

1.0.1. Media y Varianza para Binomiales

Si una variable X es Bernoulli, es decir si X es 0 o 1, si $X = 1$ hay éxito y si $X = 0$ hay fracaso, entonces la media se puede calcular en forma muy sencilla pues.

$$\mu_X = 0(1-p) + 1(p) = p,$$

mientras que la varianza es:

$$VAR(X) = (0-p)^2(1-p) + (1-p)^2 p = p(1-p).$$

Si una variable Y es binomial de parámetros n, p entonces Y puede verse como una secuencia de ensayos de Bernoulli es decir $Y = X_1 + \cdots + X_n$. Donde X_i es 1 si en el intento i hay éxito, y es 0 si en este intento hay fracaso. De acuerdo a lo anterior:

$$E(X_i) = p, Var(X_i) = p(1-p)$$

Para calcular la media y varianza de Y se utilizan las propiedades de esperanza y varianza:

$$E(Y) = E(X_1) + E(X_2) + ... + E(X_n) = \underbrace{p + p + ... + p}_{n} E = np$$

Similar la varianza. Esto se resume en el siguiente teorema.

Teorema 2.

Si X sigue una distribución binomial de parámetros n, p entonces:

$$E[X] = np \text{ y } VAR(X) = np(1-p).$$

Este último resultado también se puede obtener recurriendo a la definición de la media.

2. Distribución de Probabilidad Geométrica

Definición 2.

Se dice que un experimiento es geométrico cuando consiste de repetir un ensayo de Bernoulli, con repeticiones independientes hasta que ocurra el primer éxito.

Ejemplo de este tipo de distribuciones son: lanzar una moneda tantas veces como sea necesario hasta obtener el primer escudo, o extraer bolitas con reposición de una celda que tiene 4 rojas y 5 azules hasta obtener la primer bolita azul.

Teorema 3.

Si X es una variable aleatoria discreta que sigue una distribución de probabilidad Geométrica, con una probabilidad de éxito de p entonces se tiene:

1. El espacio muestral tiene la forma

$$\Omega = \{E, FE, FFE, FFFE\dots\}$$

2. El rango de X es $1,2,3,\dots$ y

$$f_X(x) = (1-p)^{x-1}p$$

3. La media esta distribución se calcula por

$$\mu_X = \sum_{i=1}^{\infty} i(1-p)^{i-1}p = -p\sum_{i=1}^{\infty}\frac{d(1-p)^i}{dp} = -p\frac{d(\sum_{i=1}^{\infty}(1-p)^i)}{dp} = -p\frac{d(1/p)}{dp} = \frac{1}{p}$$

Si X sigue una distribución geométrica se suele escribir que

$$X \sim G(p)$$

Note que en algunos casos interesa como variable aleatoria el total de fracasos previos al primer éxito, en ese caso deben hacerse las adaptaciones pertinentes en rango y en media.

La distribución geométrica es un caso particular del siguiente experimento:

"Se ejecutan, de manera independiente, ensayos de Bernolli hasta completar k éxitos"

Si la variable aleatoria de interés es el total de repeticiones del experimento hasta obtener los k éxitos, entonces el rango es $k, k+1, k+2, \dots$ y la densidad de probabilidad es

$$f_X(i) = \binom{i-1}{k-1}(1-p)^{i-k}p^k$$

Note que en ese caso se requiere que el último ensayo sea éxito y de los $i-1$ ensayos anteriores haya $k-1$ éxitos

Ejemplo 4.

Un restaurante ofrece servicio de lunes a viernes. Se ha estimado que la probabilidad de que el restaurante no pueda atender todos los pedidos de un día es del 15% ¿Cuál es la probabilidad de que en una semana dada, el jueves sea el primer día que no pueda atender todos sus pedidos?

Solución

$$\begin{aligned} Y &: \text{ \# de días de la semana hasta el día que no se atiendan todos los pedidos} \\ Y &\sim G(0.15) \\ P(Y=4) &= 0.85^3 0.15 = 0.09211875 \end{aligned}$$

Ejemplo 5.

Juan se dedica a pegar piso cerámico y se ha determinado que la probabilidad de que pegue un cuadro de cerámica bien es del 70%. El rector de una universidad lo ha contratado para que pegue el piso cerámico de 15 servicios sanitarios que posee la institución, cada uno lleva 16 cuadros de cerámica. El rector considera que el piso pegado de un servicio sanitario es aceptable si tiene a lo sumo dos cuadros de cerámica mal pegados. Juan pegará el piso del primer servicio sanitario, y cuando termine pasa al segundo, y asi sucesivamente. Cada vez que termine un servicio, el rector inspecciona la obra y si no es aceptable, lo despide. ¿Cuál es la probabilidad de que Juan pegue la cerámica solamente de 10 servicios sanitarios?

Solución Para resolver este problema se requiere definir dos variables aleatorias. Primero requerimos saber la probabilidad de que el piso de un servicio sanitario sea pegado aceptablemente, para ello considere la variable

$$X : \text{\# de cuadros mal pegados de 16 cuadros de un servicio sanitario}$$

Note que $X \sim B(16, 0.7)$ por lo tanto, la probabilidad de que el piso de un servicio sanitario sea pegado aceptablemente es

$$p = P(X \leq 2) = \sum_{x=0}^{2} C(16, x)\, 0.3^x 0.7^{16-x} = 0.0993597$$

Ahora bien, de acuerdo a lo indica juan pegará la cerámica de baño en baño, hasta que haya un baño cuya cerámica no sea pegada aceptablemente, es decir si se define

$$Y : \text{\# de servicios sanitarios arreglados por Juan.}$$

note que por ejemplo si $Y = 5$ significa que Juan arreglo 5 sanitarios, los primeros 4 servicios fueron arreglados aceptableme y el quinto no, y Juan fue despedido. Así, Y es el número de intentos hasta hallar un éxito (un servicio sanitario no aceptable). La probabilidad de éxito es $1 - p = 1 - 0.0993597$, por lo tanto

$$Y \sim G(1 - 0.0993597)$$

y entonces

$$P(Y = 10) = (0.0993597)^9 (1 - 0.0993597)$$

3. Ejercicios propuestos

1. Para probar un nuevo sistema en una central telefónica en una empresa, se realiza una prueba diaria la cual consiste en marcar un número específico tantas veces como sea necesario hasta obtener comunicación. El sistema se considera eficiente durante un día específico si al realizar dicha prueba telefónica se logra comunicación con a lo sumo 3 intentos. Si la probabilidad de obtener comunicación está garantizada en un 25 % de las veces por parte de la empresa que instaló la central telefónica. Determine la probabilidad de que al observar el sistema por 15 días hábiles, en al menos 8 de ellos el sistema sea hallado eficiente.

   Solucin: Este es un problema complejo, primero se debe encontrar la probabilidad de que el sistema resulte eficiente un da, y luego calcular la probabilidad de que en 15 das haya al menos 8 en que el sistema result eficiente. La primera parte es geomtrica y la segunda es binomial.

   La probabilidad de éxito es $p = 1/4$. Si X es el número de intentos,

$$P[\text{eficiente}] = P[X \leq 3] = \frac{1}{4} + \left(\frac{3}{4}\right)\left(\frac{1}{4}\right) + \left(\frac{3}{4}\right)^2\left(\frac{1}{4}\right) = \lambda$$

   La probabilidad de observar el sistema por 15 días y que en al menos 8 de ellos sea eficiente es binomial

$$\begin{aligned} P[y \geq 8] &= 1 - P[x \leq 7] \\ &= 1 - B(7; 15, \lambda) \end{aligned}$$

2. Una tienda que vende suministros de cómputo vende dos tipos de discos compactos, el normal y otro tipo llamado extra. El 70 % de los clientes de la tienda buscan el tipo extra.

   a) Entre 10 clientes seleccionados al azar, que desean comprar un disco, ¿cuál es la probabilidad de que por lo menos 5 busquen el tamaño extra?

      Solución

      El número de discos extra buscados por los próximos diez clientes sigue una distribución binomial. As la respuesta se obtiene de calcular:

      $P[X > 4] = 1 - B(4; 10, 0.7) = 1 - 0.0473 = 0.9527$

   b) Si en la tienda hay en este momento 5 discos de cada tipo, ¿cuál es la probabilidad de que los próximos 10 clientes que busquen un disco puedan comprar lo que desean?

      Solución

      $b(5; 10, 0.7) = 0.103$

3. El 45 % de los miembros de una población tienen edades superiores a los 40 años. Se selecciona al azar una muestra de 550 personas, ¿cuál es la probabilidad de que:

   a) entre 200 y 250 tengan más de 40 años?

   Solución
   Esto no es binomial, pero se puede aceptar cierta flexibilidad, cada vez que se selecciona una persona los porcentajes pueden variar, pero en este caso la variacin es tan pequea que se puede aproximar de la siguiente manera:

   $$B(250;550,0.45)-B(199;550,0.45)$$

   b) al menos 150 sean mayores de 40 años?

   Solución
   Si Y es la cantidad de personas con edades superiores a los 150 aos, $Y \sim b(x;150,0.45)$ y $P[200 \leq Y \leq 250]$ es:

   $$1-B(149;550,0.45)$$

4. Se lanza una moneda cargada, 0.6 escudo, hasta observar 2 coronas. Describa la distribución de probabilidad asociada con este experimento.

   Solución Este tipo de distribución se llama *binomial negativa*, repetir un ensayo tipo Bernoulli con probabilidad de éxito p, hasta obtener k éxitos. Para la variable X que indica el número de repeticiones necesarias el evento $X = n$ es una combinación de dos situaciones: Primero que el último sea un éxito y segundo que entre los primeros $n-1$ ensayos haya $k-1$ éxito. Así

   $$P[X=n] = bn(x;k,p) = \binom{n-1}{k-1} p^{k-1}(1-p)^{n-k} p = \binom{n-1}{k-1} p^{k}(1-p)^{n-k}$$
   $$x = k,\ k+1,\ \cdots$$

   La respuesta a la pregunta dada es $bn(x;2,0.6) = \binom{n-1}{1} p^2(1-p)^{n-2}$, para $x = 2,3,\ldots$

5. Suponga que el 30 % de los estudiantes del TEC usan con regularidad el servicio de la soda comedor. Se selecciona al azar una muestra de 500 estudiantes. Cuál es la probabilidad de que:

   a) entre 125 y 200 usen el servicio de soda comedor.

   b) más de 300 usen el servicio de soda comedor.

Solución:

Este es aproximadamente binomial.

Exacta

- $B(200;500,0.3) - B(124;500,0.3)$
- $1 - B(300;500,0.3)$

-

6. Una fábrica de tornillos de precisión empaca sus tornillos en cajas pequeñas de 15 tornillos y en cajas medianas de 80 tornillos.

   Un inspector en control de calidad revisa uno a uno los tornillos para verificar especificaciones e históricamente, de cada 100 tornillos revisados, rechaza 30. Si el inspector selecciona una caja pequeña al azar, ¿cuál es la probabilidad de que en esa caja, entre 7 y 9 tornillos estén buenos.

   Solución:

   $B(9;15,0.3) - B(6;15,0.3)$

7. En un salón hay 8 mujeres y 5 hombres. Para las variables aleatorias de los siguientes experimentos determine: posibles valores, distribución de probabilidad y la probabilidad específica solicitada.

   Solución:

   - Se eligen 6 personas reintegrándolas y tome X como el total de mujeres elegidas, la probabilidad solicitada es $P[X=4]$. Se trata de una binomial $B(x;6,8/13)$ el cálculo de la probabilidad pedida es

   $$\binom{6}{4}\left(\frac{8}{13}\right)^4\left(\frac{5}{13}\right)^2$$

   - Se eligen personas reintegrándolas a la muestra, hasta obtener dos mujeres, la probabilidad solicitada es $P[X=4]$.
     Se trata de una binomial negativa y la probabilidad pedida es:

   $$\binom{3}{1}\left(\frac{8}{13}\right)^2\left(\frac{5}{13}\right)^2$$

8. La probabilidad de que un vehículo pase las pruebas de control de calidad es de 0.93. Determine la probabilidad de que al menos 22 de los próximos 24 vehículos pasen la prueba.

   Solución:

$$B(22,24,0.93) = b(22,24,0.93) + b(23,24,0.93) + b(24,24,0.93)$$

9. Se ha determinado que el 20 % de los estudiantes de un curso son de zona rural. Si se elige un grupo de 25 estudiantes, determine la probabilidad de que en la muestra haya más de 6 pero diez o menos estudiantes de zona rural.

   Solución:

   La distribución para el número de estudiantes de zona rural es binomial $b(x;25,0.2$ y la probabilidad pedida es

$$B(10;25,0.2)-B(6;25,0.2)=\sum_{i=7}^{10}\binom{25}{i}(0.2)^i(0.8)^{25-i}$$

10. Un dado tiene marcadas cuatro caras con un *uno* y las dos restantes con un *cero*. El dado se tira tantas veces como sea necesario hasta obtener una suma acumulada en las caras de k, con k entero positivo.

    a) Especifique el espacio muestral para este experimento aleatorio.

    Solución:

    Note que para lograr una suma k deben hacerse la menos k lanzamientos, luego el espacio muestral es

Lanzamientos	Ejemplo	Probabilidad
k	$\overbrace{11\ldots1}^{k-unos}$	$(2/3)^k$
$k+1$	$\{\overbrace{101\ldots1}^{(k-1)1,(1)0}\ 1$	$\binom{k}{k-1}(1/3)(2/3)^k$
$\vdots$	$\vdots$	$\vdots$
$k+s$	$\{\overbrace{0\ldots011\ldots1}^{(k-1)1(s)0}1$	$\binom{k+s-1}{k}(1/3)^s(2/3)^k$
$\vdots$	$\vdots$	$\vdots$

    b) Determine la probabilidad de que se deban hacer 7 lanzamientos, asuma que $k=5$.

    Solución:

$$P[n=7]=\binom{6}{2}(2/3)^5(1/3)^2$$

11. La probabilidad de que una persona presente una reacción alérgica a una droga es de 0.03. Determine las siguientes probabilidades

    - Que en los próximos 500 pacientes menos de 7 presenten la reacción.

    $$B(6,500,0.03)$$

    - Que el primer paciente en reaccionar sea el cuarentavo.

    $$g(40,0.03) = (1-0.03)^{39}0.03$$