

Distribuciones Discretas

El comportamiento de una variable aleatoria queda descrito por su recorrido, su función de probabilidad, su función de distribución de probabilidad, su esperanza y su varianza, sin importar si se la representa en forma gráfica mediante un histograma, en forma tabular, o con una fórmula.

A menudo, las observaciones que se generan en diversos experimentos estadísticos tienen similares tipos de comportamiento. En consecuencia, las variables aleatorias discretas asociadas con esos experimentos se pueden describir esencialmente con la misma distribución de probabilidad, y por tanto todas ellas se pueden representar mediante una fórmula única. De hecho, se necesita sólo un puñado de distribuciones de probabilidades importantes para describir muchas de las variables aleatorias discretas que se encuentran en la práctica.

Estas distribuciones sirven de modelo para analizar nuestra realidad. Como todo modelo, es una representación simplificada, y muchas veces debemos asimilar un determinado problema al modelo que más se acerca a nuestro experimento. Por ejemplo, si nuestro problema a estudiar fueran los resultados de lanzar tres monedas, podríamos estudiarlo desde una mirada similar simplificada: lanzaremos tres veces una misma moneda y calcularemos las características resultantes a través de algún modelo estadístico conocido que nos permita calcularlas.

Modelos especiales

Hay algunas variables aleatorias discretas que se usan frecuentemente para distintos usos prácticos. Estos modelos reciben nombres especiales, que se utilizan en todo el mundo para distintas áreas de aplicación.

Los modelos principales que estudiaremos son: Binomial, Hipergeométrico, Binomial Negativo y Geométrico. Todas estas distribuciones se basan en las siguientes premisas:

- Cada prueba tiene sólo dos posibles resultados: genéricamente los llamamos *éxito* y *fracaso*,
- Denotamos con p a la $P(\text{éxito})$ y q a la $P(\text{fracaso})$,
- El resultado de cada prueba es independiente del resultado de las demás pruebas,
- La probabilidad de éxito no cambia de una prueba a otra.

Distribución Binomial

Se tiene un número fijo de pruebas n , con la siguiente característica propia de esta distribución:

Nos interesa sólo el número total de éxitos X y no el orden en que hayan ocurrido.

Por lo que la variable aleatoria X representa la cantidad de éxitos ocurridos en un experimento en el que se han prefijado la cantidad de pruebas a realizarse.

Cuando se cumplen las condiciones anteriores, X tiene la *distribución de probabilidad Binomial* con parámetros n y p , donde n es el número de intentos (pruebas) y p la probabilidad de obtener un éxito en cada prueba. Los valores posibles de X son desde cero hasta " n ", o sea, su $R_x = \{0, 1, 2, \dots, n\}$. La función de probabilidad resultante es:

$$f(k) = P(X=k) = {}_nC_k p^k q^{(n-k)}, \text{ para } k \text{ en } \{0, 1, 2, \dots, n\}$$

Se puede determinar que:

La **Esperanza** de la distribución binomial es: $E(X) = \mu = n \cdot p$

y su **Varianza** es: $V(X) = \sigma^2 = n \cdot p \cdot q$

Ejemplos:

1. En una familia de tres hijos ¿cuál es la probabilidad de que como máximo 2 sean niñas?

De la realidad surge que la probabilidad de ser niña es de 0,5 al igual que la probabilidad de ser niño, no ser niña, también es de 0,5. El sexo de cada hijo es independiente del de los demás. Podríamos considerar que nuestro éxito es que el sexo del hijo sea niña (éxito significa cumplir un objetivo, y nuestro objetivo es analizar cuál es la probabilidad de que han nacido dos niñas).

Pensemos nuestro problema ajustándolo a una distribución Binomial:

Me interesa analizar *cuántas niñas hay en los tres hijos de una familia dada*. Si ser niña es un éxito, y tres hijos son tres pruebas, entonces nuestro problema se transforma en analizar la variable X que mostrará *cuántos éxitos hay en tres pruebas*.

El recorrido de nuestra variable X es $R_x = \{0, 1, 2, 3\}$. O sea, en este caso, podemos pensar que $n = 3$, $p = 0,5$ y $q = 0,5$. Ahora bien, nuestra pregunta era *cuál es la probabilidad de que como máximo 2 sean niñas*, por lo que queremos calcular la probabilidad que $X \leq 2$.

Entonces, resolver el problema anterior se transformó calcular la probabilidad de ocurrencia mediante la Función de Distribución de una variable Binomial. $X \rightarrow B(3, 1/2)$

$$\begin{aligned} P(X \leq 2) &= F(2) = P(X=0) + P(X=1) + P(X=2) = {}_3C_0 p^0 q^{(3-0)} + {}_3C_1 p^1 q^{(3-1)} + {}_3C_2 p^2 q^{(3-2)} = \\ &= 1 \cdot 0,5^0 \cdot 0,5^3 + 3 \cdot 0,5^1 \cdot 0,5^2 + 3 \cdot 0,5^2 \cdot 0,5^1 = \\ &= 1 \cdot 0,125 + 3 \cdot (0,125) + 3 \cdot (0,125) = 0,875 \end{aligned}$$

o bien,

$$P(X \leq 2) = 1 - P(X=3) = 1 - {}_3C_3 p^3 q^{(3-3)} = 1 - 1 \cdot 0,5^3 \cdot 0,5^0 = 1 - 0,125 = 0,875$$

2. Se extraen 4 bolillas CON REEMPLAZO de una urna que tiene 5 blancas y 3 negras. ¿Cuál es la probabilidad de que salgan menos de 2 blancas?

Como la elección se hace con reemplazo, se cumplen los requisitos de un experimento binomial. X es el cantidad de bolillas blancas al tomar 4 veces una bolilla, controlarla y colocarla nuevamente en la urna. Podríamos decir, que ser *una bolilla blanca es un éxito* y en el experimento se realizan *4 pruebas* independientes, por lo que $n = 4$ y $p = 5/8 = 0,625$.

La variable es X : cantidad de bolillas blancas en 4 pruebas y su recorrido es $R_x = \{0, 1, 2, 3, 4\}$, o sea, la variable X sigue una distribución Binomial con parámetros $n = 4$ y $p = 5/8$ que se puede representar como $X \rightarrow B(4, 5/8)$

La probabilidad que necesitamos es $P(X < 2)$ y ésta es igual a:

$$P(X < 2) = F(1) = P(X=0) + P(X=1) = 0,375^4 + 4 \cdot 0,625 \cdot 0,375^3 = 0,1516$$

3. De un lote con 1,000 artículos de los cuales el 10% son defectuosos, se escogen al azar 10. ¿Cuál es la probabilidad de que haya más de 2 defectuosos?

Aunque el muestreo se hace *sin reemplazo*, por la gran cantidad de artículos que hay en el lote comparados con los que se escogen, suponemos que son válidos los supuestos del modelo Binomial, donde X es el número de artículos defectuosos al tomar 10 veces un artículo (nótese que no decimos tomar 10 artículos, sino 10 veces un artículo para así tener pruebas que cumplan con las características de una Binomial) y por ende, la probabilidad que se nos pide es: $P(X > 2)$.

Como el cálculo de las 8 probabilidades necesarias se haría muy pesado, notemos que lo haremos mediante el complemento:

$$P(X > 2) = 1 - P(X \leq 2)$$

Aún así, podemos acceder al resultado consultando la Tabla de la Binomial adecuada con $n = 10$, $p = 0,10$. La Tabla nos dice que $F(2) = P(X \leq 2) = 0,9298$. Por lo tanto, la probabilidad buscada es $P(X > 2) = 0,0702$ [Vea su Tabla y descubra este valor].

Claro que también tenemos la posibilidad de calcular $P(X \leq 2)$ y restarla de uno.

$$P(X \leq 2) = P(X=0) + P(X=1) + P(X=2) = 0,9^{10} + 10 * 0,1^1 * 0,9^9 + 45 * 0,1^2 * 0,9^8 = 0,9298$$

$$\text{y } P(X > 2) = 1 - 0,9298 = 0,0702$$

Distribución de Bernoulli

Este modelo se llama así en honor del matemático y estadístico suizo Daniel Bernoulli, quien lo define en el siglo XVIII. Se trata en realidad de una distribución de probabilidad Binomial con $n = 1$.

Distribución Hipergeométrica

Se tiene un grupo de N objetos de los cuales K son éxitos y el resto $(N-K)$, son fracasos. Se seleccionan n de los N , SIN REEMPLAZO, por lo que se realiza una sola prueba. La variable aleatoria X que nos interesa es el número total de éxitos entre los n seleccionados y no el orden en que salieron. En estas condiciones, la variable aleatoria X se dice que tiene una *distribución de probabilidad Hipergeométrica*. Su recorrido es $R_x = \{0, 1, \dots, n\}$. La función de probabilidad está dada por:

$$P(X=k) = \frac{\binom{K}{k} * \binom{N-K}{n-k}}{\binom{N}{n}} \quad k \in R_x = \{0, 1, \dots, n\}$$

La **esperanza** es: $E(X) = n * \frac{K}{N}$

y la **varianza**: $V(X) = \frac{N-n}{N-1} * n * \frac{K}{N} * (1 - \frac{K}{N})$

N : Cantidad Total de elementos.

n : cantidad de elementos que tomo del total

K : Cantidad de éxitos Totales en N.

k : cantidad de éxitos buscados en los n tomados

Como se ve buscamos cantidad de éxitos en el subconjunto n tomado de un grupo de elementos N en donde hay, y lo sabemos, K éxitos totales. Estas son las características que deben cumplirse para asociar un problema a una *distribución de probabilidad Hipergeométrica*.

Ejemplos:

1. De un lote de 40 artículos se sabe que hay 5 defectuosos en el mismo. Se seleccionan al azar 4 para probarlos y si fallan la prueba más de 2 se rechaza el lote completo. ¿Cuál es la probabilidad de rechazar el lote?

Dado que el muestreo se hace sin reemplazo, se toman 4 artículos (hay una sola prueba), $n=4$. Tenemos un total de 40 artículos, $N=40$. Hay un total de defectuosos de 5, son 5 éxitos totales, $K=5$. Tenemos una variable aleatoria que asociamos con una distribución de probabilidad Hipergeométrica. No siempre el éxito será algo “bueno”, en este caso un éxito es una pieza defectuosa.

Los parámetros son: $N=40$, $K=5$, $n=4$, X es el número de defectuosos en la muestra y queremos la probabilidad,

$$P(X=k) = \frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}}$$

$$P(X > 2) = P(X=3) + P(X=4) = \frac{\binom{5}{3} \binom{40-5}{4-3}}{\binom{40}{4}} + \frac{\binom{5}{4} \binom{40-5}{4-4}}{\binom{40}{4}} = \frac{10 \cdot 35}{91390} + \frac{5}{91390} \cong 0,00388 \cong 0,004$$

Ésta es la probabilidad de rechazar el lote y es muy baja. Para mejorar el proceso de selección, los ingenieros deciden rechazar el lote cuando haya 2 o más defectuosos. ¿Cuál sería ahora la probabilidad de rechazar un lote? Como los parámetros permanecen iguales, lo que cambia es la probabilidad resultante, que ahora pasa a ser:

$$P(X \geq 2) = P(X=2) + P(X>2) = \frac{\binom{5}{2} \binom{40-5}{4-2}}{\binom{40}{4}} + 0,004 = \frac{10 \cdot 595}{91390} + 0,004 = 0,065 + 0,004 = 0,069$$

Con esta nueva política de rechazar el lote cuando sean 2 o más, ¿cuál es la probabilidad de rechazar un lote que tuviera 6 defectuosos? Los parámetros son, ahora, $N=40$, $K=6$, $n=4$ y queremos la probabilidad:

$$P(X \geq 2) = 1 - P(X=0) - P(X=1) = 1 - \frac{\binom{6}{0} \binom{40-6}{4-0}}{\binom{40}{4}} - \frac{\binom{6}{1} \binom{40-6}{4-1}}{\binom{40}{4}} = 1 - \frac{1 \cdot 46376}{91390} - \frac{6 \cdot 5984}{91390} \cong 0,0997$$

Este es el razonamiento utilizado en los procedimientos de control de calidad para identificar cómo, a partir de una muestra reducida, tomar una decisión de aceptar o rechazar un lote. Fíjense que en este caso, con un pequeño incremento en la cantidad de piezas defectuosas supuestas en el lote, puedo definir una regla sencilla para rechazar el lote (si detecto dos o más defectuosas en mi control, lo rechazo).

- En el salón de tercer año de una escuela hay 35 alumnos, de los cuales 10 son niñas y 25 niños. Se nombra un comité de 7 alumnos que represente al salón. La selección se hace al azar. ¿Qué probabilidad hay de que en el comité haya mayoría de niñas?

Si consideramos como variable aleatoria X : cantidad de niñas, al seleccionar 7 alumnos de un grupo de 35 alumnos de los cuales 10 son niñas, se cumplen las hipótesis de una Hipergeométrica. Los parámetros son: $N=35$, $K=10$, $n=7$, y X es la cantidad de niñas en el comité. Si el comité es de 7 alumnos y tiene que haber mayoría de niñas, deseamos que la cantidad de niñas sea más que 3, o sea, 4 ó 5 ó 6 ó 7 niñas en el grupo designado para el Comité.

La probabilidad pedida es entonces:

$$P(X > 3) = P(X=4) + P(X=5) + P(X=6) + P(X=7) =$$

$$= \frac{\binom{10}{4} \binom{35-10}{7-4}}{\binom{35}{7}} + \frac{\binom{10}{5} \binom{35-10}{7-5}}{\binom{35}{7}} + \frac{\binom{10}{6} \binom{35-10}{7-6}}{\binom{35}{7}} + \frac{\binom{10}{7} \binom{35-10}{7-7}}{\binom{35}{7}} = 0.0839$$

Distribución Geométrica

El modelo Geométrico surge con la mismas premisas y características de la *distribución Binomial*, pero con una particularidad al realizar la serie de ensayos:

- No tenemos fijo el número de pruebas, sino que vamos a realizar esos ensayos *hasta obtener el primer éxito*. Este éxito es el único posible dada la naturaleza del experimento. La variable aleatoria que vamos a observar es pues el *número total de pruebas realizadas para alcanzar el éxito*.

Pensemos en aprobar un examen, un alumno puede rendirlo varias veces pero, una vez que lo aprobó, dudo que alguien lo siga rindiendo. Lo mismo si estoy analizando un ensayo con destrucción: una vez que destruyo mi objetivo, no voy a seguir con las pruebas.

Por ejemplo, si tiramos una moneda hasta que salga la primera cara, el número total de lanzamientos necesarios será la variable aleatoria X que nos interesa. Esta variable aleatoria tiene como posibles valores $\{1, 2, 3, \dots, k\}$. X : cantidad de pruebas hasta alcanzar 1 éxito.

Se puede demostrar que la función de probabilidad resultante es:

$$P(X = n) = q^{(n-1)}p; \quad \text{para } n \in R_x = \{1, 2, 3, \dots, k\}$$

Se puede también verificar que:

La **esperanza** de la distribución Geométrica es: $E(X) = \mu = \frac{(1-p)}{p}$

y su **varianza** es: $V(X) = \sigma^2 = \frac{(1-p)}{p^2}$

Ejemplos:

- En un canal de comunicación, un modem puede equivocar un caracter enviado con una probabilidad $p=0,001$. ¿Qué probabilidad hay de que transmita 5 caracteres o más sin error?

De la consigna surge que se aplica el modelo de *distribución Geométrica*. La probabilidad que buscamos es $P(X \geq 5)$, siendo X : cantidad de caracteres enviados hasta que se produzca un error. Para calcular esta probabilidad hay que encontrar el valor de la *suma de las probabilidades de $X=5, X=6, \dots, X=n$, hasta infinito*, usando $q^{(k-1)} * p$ para cada valor de $X=n$, cosa que sería imposible. Para resolverlo usamos el complemento:

$$\begin{aligned} P(X \geq 5) &= 1 - P(X < 5) = 1 - P(X \leq 4) = \\ &= 1 - P(X=1) - P(X=2) - P(X=3) - P(X=4) \cong 1 - 0,004 \cong 0,996. \end{aligned}$$

Nota: En la distribución Geométrica nunca puedo calcular $P(X > n)$ ni $P(X \geq n)$, debo siempre trabajar con el complemento $1-F(n)$ o $1 - P(X < n)$, ya que el recorrido de la variable es infinito.

En el mismo problema ¿Cuál es el promedio de caracteres transmitidos sin error?

$$\text{La respuesta es } E(X) = 0,999 / 0,001 = 999$$

2. En un proceso de manufactura se sabe que, en promedio, una de cada cincuenta piezas es defectuosa. ¿Cuál es la probabilidad de que la sexta pieza seleccionada al azar sea la primera defectuosa?

Aquí aplicamos el modelo Geométrico con $p=0,02$.

La probabilidad que requerimos es $P(X = 6)$.

De acuerdo con la fórmula, ésta es: $P(X = 6) = 0,98^5 * 0,02^1 \cong 0,018$

Distribución Binomial Negativa

La **Binomial Negativa** es el modelo que aplicamos cuando, en la misma situación que la Geométrica, lo que nos interesa es el número de intentos necesarios para obtener el k -ésimo éxito. Lo último que ocurre en este tipo de experimento es un éxito y los demás $(k-1)$ éxitos anteriores se dan en cualquier orden. Nos interesa conocer cuántas pruebas (n) debemos realizar para lograr esa cantidad k de éxitos. El recorrido de esta variable aleatoria es $R_x = \{k, k+1, k+2, \dots\}$. La función de probabilidades es pues:

$$P(X = n) = {}_{n-1}C_{k-1} p^k q^{(n-k)}, \text{ para } n \in R_x = \{k, k+1, k+2, \dots\}$$

Ejemplo:

1. Una persona se entretiene lanzando al aire una moneda, la recoge y la vuelve a lanzar. ¿Cuál es la probabilidad de que caiga ceca por tercera ocasión, en el sexto lanzamiento?

Este es un problema Binomial Negativo, donde el éxito es que la moneda caiga en ceca. La probabilidad de que suceda ésto, en un solo intento es $p = 0,5$ y $q = 0,5$.

Si X es el número de intentos en que se debe lanzar la moneda para alcanzar tres cecas, X tiene una distribución de probabilidad Binomial Negativa con $k=3$. La probabilidad que se nos pide es $P(X=6)$.

Entonces usando la fórmula, tenemos: $P(X = 6) = {}_5C_2 * 0,5^3 * 0,5^3 = 0,15625$