

Distribuciones de probabilidades de variables aleatorias discretas

Juan Espinoza B.
Facultad de Agronomía – Universidad de Concepción

Introducción

Como se ha señalado en artículos anteriores, los eventos experimentales de mayor interés con frecuencia son numéricos, es decir, realizamos un experimento y observamos el valor numérico de alguna variable. Si repetimos el experimento n veces, obtenemos una muestra de datos cuantitativos. Como ejemplo, supongamos que un producto fabricado en una empresa, se vende en lotes de 20 cajas, cada una de las cuales contiene 12 artículos. A fin de verificar la calidad del producto, el jefe de control de procesos de la empresa selecciona al azar cuatro de entre los 240 artículos de un lote y determina si los artículos están defectuosos o no. Si más de uno de los artículos muestreados resulta defectuoso, se rechazará todo el lote.

La selección de cuatro artículos fabricados de entre 240 produce un espacio muestral Ω que contiene $\binom{240}{4}$ eventos simples, cada uno de los cuales corresponde a una posible combinación de cuatro artículos que podrían seleccionarse del lote. El evento de interés para el jefe de control de procesos es la observación de la variable X número de artículos defectuosos entre los cuatro que se prueban, por lo tanto existe una relación funcional entre los eventos simples E_i de Ω y los valores que puede asumir X .

El evento $x = 0$ es la colección de todos los eventos simples que no contienen artículos defectuosos. De forma similar, el evento $x = 1$ es la colección de todos los eventos simples en lo que se observa un artículo defectuoso. Puesto que el valor que X puede asumir es un evento numérico, que varía de forma aleatoria de una repetición del experimento a otra, se dice que X es una **variable aleatoria**.

Conceptos básicos

Una **variable aleatoria** es una función que cuantifica los resultados de un experimento aleatorio. Una **variable aleatoria discreta** X sólo puede asumir una cantidad de valores susceptible de contarse. En estadística la **distribución de probabilidad** para una variable aleatoria discreta X es una tabla, gráfica o fórmula que da la probabilidad $P(X = x)$ asociada a cada posible valor de X . Si consideremos el experimento de lanzar dos veces una moneda balanceada y se observa el número X de caras. Calculemos la distribución de probabilidad para X . Sean C_i y S_i la observación de una cara y un sello, respectivamente, en el i – ésimo lanzamiento, para $i = 1, 2$. Los cuatro eventos simples y los correspondientes valores de X se muestran en la tabla 1.

Evento Simple	Descripción	P(E _i)	Número de caras X
E ₁	C ₁ C ₂	1/4	2
E ₂	C ₁ S ₂	1/4	1
E ₃	S ₁ C ₂	1/4	1
E ₄	S ₁ S ₂	1/4	0

El evento $x = 0$ es la colección de todos los eventos simples que producen un valor de $x = 0$ en este caso es el único evento E₄, luego la probabilidad de que x asuma el valor cero es: $P(x = 0) = p(0) = P(E_4) = \frac{1}{4}$. El evento $x = 1$ contiene dos eventos simples, E₂ y E₃. Por tanto, $P(x = 1) = p(1) = P(E_2) + P(E_3) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$. Por último, $P(x = 2) = p(2) = P(E_1) = \frac{1}{4}$. La tabla 2, muestra la distribución de probabilidad para x .

X	p(x)
0	1/4
1	1/2
2	1/4
	$\sum_x p(x) = 1$

Como se verá más adelante, esta distribución de probabilidad también puede calcularse con

la fórmula $p(x) = \frac{\binom{2}{x}}{4}$, usando la distribución binomial.

Requisitos para una distribución de probabilidad discreta

1. $0 \leq p(x) \leq 1$
2. $\sum_{todax} p(x) = 1$

Esperanza. Sea x una variable aleatoria discreta con distribución de probabilidades $p(x)$. Entonces el **valor esperado** o **medio de** x es $\mu = E(x) = \sum_{todax} xp(x)$

Varianza. Sea x una variable aleatoria discreta con distribución de probabilidades $p(x)$. Entonces, la **varianza** de x es $\sigma^2 = E[(x - \mu)^2] = E(x^2) - \mu^2$. La **desviación estándar** de x es la raíz cuadrada positiva de la varianza de x .

Distribución Bernoulli

Una prueba o experimento Bernoulli tiene uno de dos resultados mutuamente excluyentes, que generalmente se denotan S (éxito) y F (fracaso). Por ejemplo, al seleccionar un objeto para control de calidad puede ocurrir que sea defectuoso o no lo sea. El espacio muestral Ω , de un experimento Bernoulli consta de dos resultados $\Omega = \{ \text{éxito, fracaso} \}$

$$\text{sea } X(w) = \begin{cases} 1 & \text{si } w \text{ es éxito} \\ 0 & \text{si } w \text{ es fracaso} \end{cases}$$

Supongamos que p es la probabilidad de observar un éxito en un ensayo Bernoulli, donde $0 < p < 1$.

La distribución de probabilidad de la variable aleatoria (v. a) . Bernoulli, está dada por:

$$P(x = 1) = p, \quad P(x = 0) = 1 - p = q$$

La distribución de probabilidad de la variable aleatoria de Bernoulli X está dada por:

$$X \sim \text{Bernoulli}(p) \quad P(X = x) = p^x q^{1-x}; \quad x = 0, 1$$

Si $X \sim \text{Bernoulli}(p)$, entonces

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

$$\begin{aligned} \text{Var}(X) &= E(X^2) - E(X)^2 = E(X^2) - \mu^2 = \\ &= 0^2 \cdot q + 1^2 \cdot p - p^2 = p(1 - p) = pq \end{aligned}$$

$$\text{Var}(X) = pq$$

Ejemplo

Si $X \sim \text{Bernoulli}(1/3) \Leftrightarrow P(x) = (1/3)^x (2/3)^{1-x}$
entonces $p = 1/3$ y $q = 2/3$, luego
 $E(X) = 1/3$ y $\text{Var}(X) = 2/9$

Una variable aleatoria de Bernoulli, por sí sola, tiene poco interés en las aplicaciones de ingeniería y ciencias. En cambio, la realización de una serie de pruebas de Bernoulli conduce a varias distribuciones de probabilidad discretas bien conocidas y útiles.

La distribución de probabilidad binomial

Muchos experimentos en la vida real consisten en efectuar una serie de pruebas de Bernoulli y son análogos al lanzamiento de una moneda no balanceada un número n de veces. Suponga que el 20% de los artículos que produce una máquina son defectuosos. En este caso seleccionar una muestra aleatoria de 10 artículos y analizarlos para determinar si son defectuosos sería análogo a lanzar una moneda no balanceada 10 veces, siendo la probabilidad de obtener una cara (artículo defectuoso) en una sola prueba igual a 0,20.

Las encuestas de opinión pública o preferencias de los consumidores que generan dos respuestas (sí o no, aprueba o desaprueba, etc.) también son análogas al experimento de lanzar una moneda no balanceada si el tamaño N de la población es grande y el tamaño de la muestra n es relativamente pequeño, digamos $0,10N$ o menos. Todos estos experimentos son ejemplos particulares de un **experimento binomial**.

Experimento binomial es aquel experimento generado por n ensayos Bernoulli independientes, donde cada ensayo tiene igual probabilidad de éxito p , $0 < p < 1$

$$\Omega = \Omega_1 \times \Omega_2 \times \dots \times \Omega_n; \Omega_i = \{ \text{éxito, fracaso} \}.$$

Sea X = número de éxitos en n ensayos Bernoulli, entonces el recorrido de X es $R_X = \{ 0, 1, 2, \dots, n \}$.

La distribución de probabilidad para X es:

$$P(X = x) = \binom{n}{x} p^x q^{n-x} \quad x = 0, 1, 2, \dots, n$$

Observaciones

$$1) \quad (a + b)^n = \sum_{k=0}^n \binom{n}{k} a^k b^{n-k} \quad (p + q)^n = 1 = \sum_{x=0}^n \binom{n}{x} p^x q^{n-x}$$

$$2) \quad \text{Si } X \sim \text{Binomial}(n, p), \text{ entonces} \\ E(X) = np \quad \text{Var}(X) = npq$$

Ejemplo

Un agente de seguros vende pólizas a 5 individuos, todos de la misma edad. De acuerdo con las tablas actuariales, la probabilidad de que un individuo con esa edad viva 30 años más es de $3/5$. Determinar la probabilidad de que dentro de 30 años vivan:

1. Los cinco individuos.
2. Al menos tres.
3. Sólo dos.
4. Al menos uno.

Solución

Estamos frente a una variable Bernoulli ya que dentro de 30 años se pueden presentar dos situaciones que la persona este viva ($p = 3/5$) o que haya muerto ($q = 2/5$). Al considerar los 5 individuos, estamos frente a una variable aleatoria X binomial con $n = 5$, $p = 0,6$ $X \sim B(5, 0,6)$.

$$P(X = x) = \binom{5}{x} 0,6^x 0,4^{5-x} \quad x = 0, 1, 2, 3, 4, 5$$

1. Debemos calcular $P(X = 5)$

$$P(X = 5) = \binom{5}{5} 0,6^5 0,4^0 = 0,07776$$

2. Debemos calcular $P(X \geq 3)$, o lo que es lo mismo,

$$1 - P(X < 3) = 1 - P(X = 0) - P(X = 1) - P(X = 2)$$

$$= 1 - 0,01024 - 0,0768 - 0,2304 = 0,68256$$
3. $P(X = 2) = 0,2304$
4. $P(X \geq 1) = 1 - P(X = 0) = 1 - 0,01024 = 0,98976$

Ejemplo

Suponiendo que la probabilidad de tener un hijo varón es 0,51. Hallar la probabilidad de que una familia con seis hijos tenga:

1. Por lo menos un niño.
2. Por lo menos una niña.

Solución

Si X es la variable número de hijos varones en una familia de seis hijos, estamos frente a una variable binomial con

$$n = 6 \text{ y } p = 0,51.$$

1. Debemos calcular $P(X \geq 1) = 1 - P(X = 0)$

$$1 - P(X = 0) = 1 - \binom{6}{0} (0,51)^0 (0,49)^6 = 0,9862$$

2. Debemos calcular $P(X \leq 5) = 1 - P(X = 6)$

$$1 - P(X = 6) = 1 - \binom{6}{6} (0,51)^6 (0,49)^0 = 0,9824$$

La distribución Poisson

Fue introducida en 1837 por el matemático francés Simeón Dennis Poisson (1781-1840). La distribución de Poisson se utiliza como distribución de las ocurrencias de un fenómeno en una unidad de tiempo. También se utiliza como modelo del número de defectos o disconformidades que ocurren en una unidad de producto. En realidad, cualquier fenómeno aleatorio que ocurre por unidad (de área, de volumen, de tiempo etc.) se puede aproximar bien, en la mayoría de los casos, por la ley de Poisson. También se utiliza en el diseño de límites de control de los diagramas p en control de calidad.

Características de una variable aleatoria Poisson

1. El experimento consiste en contar el número x de veces que ocurre un evento en particular durante una unidad de tiempo dado, o en un área o volumen (o peso, distancia o cualquier otra unidad de medida) dada.

2. La probabilidad de que un evento ocurra en una unidad dada de tiempo, área o volumen es la misma para todas las unidades.
3. El número de eventos que ocurren en una unidad de tiempo o área o volumen es independiente del número de los que ocurren en otras unidades.
4. El número medio (o esperado) de eventos en cada unidad se denota por la letra griega lambda, λ .

La distribución de probabilidad Poisson

La distribución de probabilidades par una variable aleatoria Poisson está dada por

$$P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!} \quad \lambda > 0, \quad x = 0, 1, 2, 3, \dots$$

Donde, λ = es el número de eventos en una unidad dada de tiempo, área o volumen.
 $e = 2,71828\dots$

la media y la varianza de la variable aleatoria Poisson son

$$\text{Media } \mu = E(X) = \lambda, \quad \text{Varianza } \sigma^2 = \text{Var}(X) = \lambda$$

Ejemplo

El número de accidente por semana en una fábrica sigue una distribución Poisson de parámetro $\lambda = 2$. Calcular:

1. La probabilidad de que en una semana haya algún accidente.
2. La probabilidad de que haya 4 accidentes en dos semanas.
3. La probabilidad de que haya 2 accidentes en una semana y otros dos en la semana siguiente.
4. Si sabemos que ha habido un accidente hallar la probabilidad de que en esa semana no haya más de tres accidentes.

Solución

1. Si X es la v. a. número de accidentes por semana, sabemos que se ajusta a un modelo Poisson con $\lambda = 2$, entonces

$$P(X > 0) = 1 - P(X = 0) = 1 - e^{-2} \frac{2^0}{0!} = 0,8646$$

2. Si X es el número de accidentes en la primera semana e Y es el número de accidentes en la segunda semana, estamos ante dos variables independientes Poisson de parámetro $\lambda = 2$, debemos calcular $P(X + Y = 4)$, pero la variable $X + Y$ es una variable Poisson de parámetro $\lambda = 2+2 = 4$, luego se tiene que:

$$P(X + Y = 4) = \frac{e^{-4} 4^4}{4!} = 0,1954$$

3. Debemos calcular la probabilidad de la intersección de dos sucesos independientes

$$P(X = 2) \cap P(Y = 2) = P(X = 2) \cdot P(Y = 2) = \frac{e^{-2} 2^2}{2!} \cdot \frac{e^{-2} 2^2}{2!} = 0,0733$$

4. Debemos calcular la siguiente probabilidad condicionada

$$\begin{aligned} P(X \leq 3 / X \geq 1) &= \frac{P(1 \leq X \leq 3)}{P(X \geq 1)} \\ &= \frac{\frac{e^{-2} 2}{1!} + \frac{e^{-2} 2^2}{2!} + \frac{e^{-2} 2^3}{3!}}{1 - \frac{e^{-2} 2^0}{0!}} = 0,8348 \end{aligned}$$

Aproximación de la distribución Binomial usando la distribución Poisson

La distribución de probabilidad Poisson está relacionada con una distribución de probabilidad binomial cuando n es grande y $\mu = np \leq 7$, y puede utilizarse como aproximación

Ejemplo

La proporción de alumnos de un distrito universitario con calificación de sobresaliente es de 0,005%. Determinar la probabilidad de que entre 5000 alumnos seleccionados al azar haya dos con calificación media sobresaliente.

Solución

Teóricamente la variable X definida por el número de alumnos (de entre los 5000) que tienen una calificación media sobresaliente es una variable binomial con $n = 5000$ y $p = 0,00005$. La probabilidad pedida es $P(X = 2)$, pero deberíamos calcular

$$\begin{aligned} P(X = 2) &= \binom{5000}{2} (0,00005)^2 (1 - 0,00005)^{5000-2} \\ &= 0,024334938 \end{aligned}$$

que es difícil de calcular; sin embargo como $np = 0,25 < 5$ y $p < 0,1$, podemos aproximar la variable binomial $X \sim (5000, 0,00005)$ por una variable Poisson de parámetro $\lambda = np = 0,25$. Los cálculos se muestran a continuación.

$$P(X = 2) = e^{-0,25} \frac{0,25^2}{2!} = 0,0243337524$$

Se observa que difieren en la sexta cifra decimal, es decir alrededor de las millonésimas.

Distribución Binomial Negativa y Geométrica

Consideremos una serie de ensayos Bernoulli independientes y sea r el número exacto de veces que aparece el suceso A de probabilidad p . La variable aleatoria X , que denota el número de fallos antes del r -ésimo éxito, tiene distribución Binomial Negativa. La variable binomial negativa también suele caracterizarse como el tiempo de espera hasta el n -ésimo éxito, y se aplica en estudios de fiabilidad y en situaciones cíclicas donde se alternan éxitos y fracasos.

$$X \sim \text{Bin-Neg}(r, p) \quad R_X = \{r, r+1, r+2, \dots\}$$

La distribución de probabilidades está dada por:

$$P(X = x) = \binom{x-1}{r-1} p^r q^{x-r} \quad x = r, r+1, r+2, \dots$$

Donde

p = probabilidad de éxito en una sola prueba Bernoulli,

$$q = 1 - p$$

x = número de pruebas hasta que se observa el r -ésimo éxito.

la media y la *varianza* de la variable aleatoria binomial negativa son

$$\text{Media } \mu = E(X) = \frac{r}{p}, \quad \text{Varianza } \sigma^2 = \text{Var}(X) = \frac{rq}{p^2}$$

La distribución de probabilidad binomial negativa es una función de dos parámetros, p y r . Para el caso especial en que $r = 1$, la distribución de probabilidad de x se denomina **distribución de probabilidad geométrica**.

$$p(x) = pq^{x-1} \quad x = 1, 2, 3, \dots$$

donde x = número de pruebas hasta que se observa el primer éxito.

$$\text{Media } \mu = E(X) = \frac{1}{p}, \quad \text{Varianza } \sigma^2 = \text{Var}(X) = \frac{q}{p^2}$$

Ejemplo

Un fabricante utiliza fusibles eléctricos en un sistema electrónico. Los fusibles se compran en lotes grandes y se prueban secuencialmente hasta que se observa el primer fusible defectuoso. Suponga que el lote contiene 10% de fusibles defectuosos:

- ¿Qué probabilidad hay de que el primer fusible defectuoso sea uno de los primeros cinco fusibles probados?
- ¿Calcule la media, la varianza y la desviación estándar de x , el número de fusibles probados hasta observarse el primer fusible defectuoso.

Solución

- El número x de fusibles probados hasta observarse el primer fusible defectuoso, es una variable aleatoria geométrica con

$$p = 0,1 \text{ (probabilidad de que un solo fusible sea defectuoso)}$$

$$q = 1 - p = 0,9$$

y

$$\begin{aligned} p(x) &= pq^{x-1} \\ &= (0,1)(0,9)^{x-1} \quad x = 0,1,2,3,\dots \end{aligned}$$

La probabilidad de que el primer fusible defectuoso sea uno de los primeros cinco fusibles probados es

$$\begin{aligned} p(x \leq 5) &= p(1) + p(2) + \dots + p(5) \\ &= (0,1)(0,9)^0 + (0,1)(0,9)^1 + (0,1)(0,9)^2 + (0,1)(0,9)^3 + (0,1)(0,9)^4 = 0,41 \end{aligned}$$

b. La media, la varianza y la desviación estándar de esta variable aleatoria geométrica son

$$\begin{aligned} \mu &= \frac{1}{p} = \frac{1}{0,1} = 10 \\ \sigma^2 &= \frac{q}{p^2} = \frac{0,9}{0,1^2} = 90 \\ \sigma &= \sqrt{\sigma^2} = \sqrt{90} = 9,49 \end{aligned}$$

Bibliografía

1. William Mendenhall/ Terry Sincich. Probabilidad y Estadística para Ingeniería y Ciencias. Editorial Prentice Hall, 1997. Cuarta Edición.
2. Murray R. Spiegel. Estadística. Editorial McGrawHill. 1995.
3. Webster, Allen. Estadística Aplicada. Editorial McGrawHill. 2001.
4. Mora/ Cid/Valenzuela. Probabilidades y Estadística. Universidad de Concepción. 1996.