Distribuciones de variables discretas

Kevin Piterman

5 de Mayo, 2020

Resumen.
X variable aleatoria discreta, o sea el rango de X que notamos por R_X, es un conjunto finito o numerable.

$$R_X = \{x_1, x_2, ..., x_n, ...\}.$$

Función de probabilidad puntual de X:

$$p_X(x_n) = P(X = x_n).$$

Función de distribución acumulada de X:

$$F_X(t) = P(X \leq t) = \sum_{x \leq t, x \in R_X} p_X(x).$$

Recordemos algunas distribuciones:

Tenemos un experimento que tiene cierta probabilidad p de éxito, donde $0 \leq p \leq 1$.

Sea X la variable aleatoria que nos dice si tuvimos éxito o no en este experimento. Concretamente esta variable vale 0 si no tuvimos éxito, y 1 si tuvimos éxito. Luego $R_X = \{0, 1\}$.

$$p_X(0) = 1 - p,$$

$$p_X(1) = p.$$

Esta variable tiene **distribución Bernoulli** y se denota $X \sim Be(p)$.

Supongamos que realizamos nuestro experimento n veces, y contamos en X la cantidad de éxitos que tuvimos. Entonces el rango de X es $R_X = \{0, 1, 2, ..., n\}$.

$$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}$$

Esta variable tiene **distribución Binomial**, y se denota $X \sim Bi(n, p)$.

Sea X la variable aleatoria que cuenta la cantidad de veces que tuve que realizar mi experimento hasta tener éxito.

$$R_X = \{1, 2, ...\}$$
$$p_X(k) = (1-p)^{k-1} p$$

Esta variable tiene **distribución geométrica** y se denota $X \sim Ge(p)$.

Sea X la variable aleatoria que cuenta la cantidad de veces que tuve que realizar mi experimento hasta tener exactamente r éxitos, donde $r \geq 1$ es un natural.

$$R_X = \{r, r+1, r+2, ...\}.$$

$$p_X(k) = \binom{k-1}{r-1} (1-p)^{k-r} p^r.$$

Esta variable tiene **distribución Binomial Negativa** y se denota $X \sim BN(r, p)$.

Supognamos ahora que tenemos una población de tamaño N (finito), donde tenemos B casos "favorables" (o sea las bolitas que voy a querer sacar, por ejemplo las de un color determinado). Estamos suponiendo que todos los elementos de mi población son **distinguibles** entre sí. Supongamos entonces que elegimos m elementos de nuestra población SIN reposición. Notar que $m \leq N$. Sea X la variable aleatoria que cuenta la cantidad de casos favorables (o buenos) que obtuvimos. Es claro que X es una variable aleatoria discreta.

$$R_X = \{\max\{0, m - (N - B)\}, ..., \min\{B, m\}\}.$$

$$p_X(k) = \frac{\binom{B}{k}\binom{N-B}{m-k}}{\binom{N}{m}}.$$

Esta variable tiene **distribución Hipergeométrica** y se denota $X \sim \mathcal{H}(N, B, m)$.

Ejemplo 1. Un verdulero compra cajones con 50 manzanas cada uno. Adoptó como política inspeccionar al azar 6 manzanas del cajón, y si encuentra alguna en mal estado lo rechaza. Se sabe que el 20% de los cajones trae 10 manzanas en mal estado y el 80% solo trae 3. Sea X la cantidad de cajones que acepta en un pedido de 30 cajones. Hallar la distribución de X. ¿Cuántos cajones espera que rechace en este pedido?

Solución. El rango de X es $R_X = \{0, 1, ..., 30\}$ y X es una Binomial de parámetros $n = 30$ y p (la probabilidad de aceptar un cajón). Calculemos p. Notemos que $p = P(\text{sacr 6 manzanas buenas})$.

Podemos calcular lo que vale p si nos restringimos a cada cajón. Sea CM los cajones que traen 10 manzanas malas, y sea CB los cajones que traen solo 3 malas. Entonces:

$$P(CM) = 0.2,$$

$$P(CB) = 0.8.$$

Por la fórmula de probabilidad total,

$$p = P(6 \text{ manzanas buenas}|CM)P(CM) + P(6 \text{ manzanas buenas}|CB)P(CB)$$

Entonces $P(6 \text{ manzanas buenas}|CM)$ está contando sacar 6 elementos "buenos" de un cajón de 50 elementos, si sacamos 6 y el total de buenos es 40. Por lo tanto, lo que estamos calculando es la probabilidad de que cierta hipergeométrica sea igual a 6. Concretamente, sea $Z \sim \mathcal{H}(50, 40, 6)$, entonces

$$P(6 \text{ manzanas buenas}|CM) = P(Z = 6) = \frac{\binom{40}{6}}{\binom{50}{6}}.$$

Análogamente, si $Y \sim \mathcal{H}(50, 47, 6)$, entonces

$$P(6 \text{ manzanas buenas}|CB) = P(Y = 6) = \frac{\binom{47}{6}}{\binom{50}{6}}.$$

Finalmente, reemplazamos en la fórmula de probabilidad total y obtenemos que

$$p = \frac{\binom{40}{6}}{\binom{50}{6}}0.2 + \frac{\binom{47}{6}}{\binom{50}{6}}0.8.$$

Con este concluimos cuál es la distribución de X.

Sea Y la variable aleatoria que cuenta la cantidad de cajones que rechazamos. Entonces $Y = 30 - X$ o bien $Y \sim Bi(30, 1 - p)$. La cantidad de cajones que se espera rechazar es la esperanza de Y. Luego,

$$\mathbb{E}(Y) = \mathbb{E}(30 - X) = \mathbb{E}(30) - \mathbb{E}(X) = 30 - 30p = 30(1 - p).$$

Ejemplo 2. Tenemos dos cajas con libros para leer. Cada caja contiene N libros. Elegimos una caja al azar y extraemos un libro para leer. Continuamos haciendo esto hasta el momento en donde queremos sacar un libro de una de las cajas y vemos que está vacía. ¿Cuál es la probabilidad de que queden exactamente k libros en la otra caja en ese momento?

Solución. Sea Y la variable aleatoria que nos cuenta la cantidad de libros que quedaron en la otra caja cuando nos dimos cuenta de que una estaba vacía. $R_Y = \{0, 1, ..., N\}$. Debemos calcular $P(Y = k)$. Sea $C1$ el evento que dice que se vació la caja 1, y sea $C2$ el evento que dice que se vació la caja 2. Recordemos que estamos trabajando SABIENDO que hay una caja vacía, pero no cuál. Por lo tanto, todas las probabilidades debemos interpretarlas condicionadas a que sucede el evento $C1 \cup C2$. De esta forma

$$P(Y = k) = P(Y = k \text{ y se vació el primero}) + P(Y = k \text{ y se vació el segundo}) = P((Y = k) \cap C1) + P((Y = k) \cap C2).$$

y como la elección de las cajas es al azar, ambos sumandos en la fórmula anterior son iguales.

$$P(Y = k) = 2P((Y = k) \cap C1).$$

Notemos que la probabilidad $P((Y = k) \cap C1)$ es la probabilidad de que haya elegido $N + 1$ veces la caja 1 y $N - k$ veces la caja 2. Esto lo podemos pensar como la probabilidad de que cierta Binomial Negativa que cuenta la cantidad de ensayos necesarios hasta elegir $N + 1$ veces la caja 1, sea igual a $N - k + N + 1$. Sea $Z \sim BN(N + 1, \frac{1}{2})$ la variable que cuenta la cantidad de elecciones de cajas que tuve que hacer hasta elegir la caja 1 $N + 1$ veces. Entonces $P((Y = k) \cap C1) = P(Z = N + 1 + N - k)$ y así,

$$\begin{aligned}
P(Y = k) &= 2P(Z = N + 1 + N - k) \\
&= 2\binom{N - k + N}{N}\left(\frac{1}{2}\right)^{N+1+N-k} \\
&= \binom{N - k + N}{N}\left(\frac{1}{2}\right)^{2N-k}.
\end{aligned}$$

Nota: en la resolución en clase cometimos un pequeño error al calcular el valor anterior como la probabilidad condicional de $Y = k$ dado $C1$. El problema con esto es que si ya sabemos que se vació la caja 1, entonces es porque la $N + 1$-ésima elección de la caja 1 ya se conoce. Es decir, nos "sobraba" un $p = \frac{1}{2}$ en la Binomial Negativa. Concretamente, $P(Y = k|C1)$ es la probabilidad de haber elegido $N - k$ veces la caja 2 y N veces la caja 1. Es decir, es la probabilidad de que una Binomial de $2N - k$ experimentos con probabilidad $\frac{1}{2}$ cada uno, tenga N éxitos. Luego,

$$P(Y = k|C1) = \binom{2N - k}{N}\left(\frac{1}{2}\right)^{2N-k}$$

Por otro lado, $P(C1) = P(C2) = \frac{1}{2}$, y utilizando probabilidad total concluimos que

$$\begin{aligned}
P(Y = k) &= P(Y = k|C1)P(C1) + P(Y = k|C2)P(C2) \\
&= 2P(Y = k|C1)P(C1) \\
&= P(Y = k|C1) \\
&= \binom{2N - k}{N}\left(\frac{1}{2}\right)^{2N-k}.
\end{aligned}$$

Ejemplo 3. Realizamos un experimento con probabilidad de éxito p unas 5 veces. Si logramos tener éxito a la quinta realización, seguimos haciéndolo hasta fallar.

1. Calcular el valor esperado de experimentos exitosos.
2. ¿Cuántos experimentos espera hacer?
3. ¿Cuantas fracasos espera tener?

Solución. Sea X la variable aleatoria que cuenta la cantidad de éxitos que obtuvimos. El rango de X es $R_X = \{0, 1, 2, 3, ...\}$. Sea A el evento que dice que en la quinta realización tuvimos un éxito. Entonces $P(A) = p$. Por probabilidad total,

$$P(X = k) = P(X = k|A)P(A) + P(X = k|A^c)P(A^c).$$

La idea es que sabemos la distribución de las variable X *condicionada* a si el quinto experimento falló o no.

Si en el quinto tuvimos éxito, entonces la cantidad de éxitos totales se pueden contar como la cantidad que tuvimos en los primeros 4, más 1 del quinto, más la cantidad de éxitos que hay a partir del quinto. Si Y es la cantidad de éxitos en los primeros cuatro experimentos, entonces $Y \sim Bi(4, p)$. Por otro lado, los éxitos que obtenemos a partir del quinto experimento se pueden pensar como una geométrica menos uno, porque contamos hasta sacar un fallido y éste hay que restárselo. Sea $Z \sim Ge(1 - p)$ que cuenta la cantidad de éxitos que tuve hasta obtener un fracaso, más uno.

De esta manera, cuando el quinto fue éxito, la cantidad de éxitos es $Y + 1 + (Z - 1)$, y por lo tanto

$$P(X = k|A) = P(Y + 1 + (Z - 1) = k) = P(Y + Z = k).$$

De manera análoga, vemos que $P(X = k|A^c) = P(Y = k)$ porque ya sabemos que el quinto falla. Como queremos calcular la esperanza de X y no su distribución, utilizaremos el siguiente resultado cuya demostración queda como tarea.

Tarea: Demostrar la siguiente fórmula de “probabilidad total” para la esperanza de variables discretas: si $(A_n)_{n\geq 1}$ es una partición del espacio y X es una variable aleatoria discreta, entonces

$$\mathbb{E}(X) = \sum_n \mathbb{E}(X_n)P(A_n)$$

donde X_n es una variable aleatoria discreta tal que $P(X_n = k) = P(X = k|A_n)$.

Utilizando el resultado anterior con la partición $\{A, A^c\}$, vemos que

$$\mathbb{E}(X) = \mathbb{E}(Y + Z)P(A) + \mathbb{E}(Y)P(A^c)$$

Por la linealidad de la integral, concluimos que

$$\mathbb{E}(X) = \mathbb{E}(Y) + \mathbb{E}(Z)P(A) = 4p + \frac{p}{1 - p}.$$

2. Sea W la cantidad de experimentos realizados. Si sale éxito en el quinto, entonces voy a hacer 5 experimentos más la cantidad que me diga Z. En cambio si el quinto falla, solo voy a hacer 5 experimentos. Notemos que $R_W = \{5, 6, 7, ...\}$.

De esta manera, utilizando probabilidad total,

$$P(W = k) = P(W = k|A)P(A) + P(W = k|A^c)P(A^c).$$

Análogamente al caso anterior, vemos que $P(W = k|A) = P(5 + Z = k)$ y $P(W = k|A^c) = P(5 = k)$.

Luego

$$\mathbb{E}(W) = \mathbb{E}(5 + Z)P(A) + \mathbb{E}(5)P(A^c) = 5 + \frac{p}{1-p}.$$

3. Para este ítem notemos que $W - X$ cuenta exactamente la cantidad de fracasos del experimento. Por lo tanto lo que nos piden es la esperanza de $W - X$, es decir,

$$\mathbb{E}(W - X) = \mathbb{E}(W) - \mathbb{E}(X) = 5 + \frac{p}{1-p} - \left(4p + \frac{p}{1-p}\right) = 5 - 4p.$$