Preguntas random que me hacen

Autor: Andrés Miniguano Trujillo
Escuela Politécnica Nacional
Email: andres.miniguano@epn.edu.ec
Creación: 24 de mayo de 2020

Aplicación de la distribución hipergeométrica

En este documento usarás la distribución hipergeométrica para determinar la probabilidad de una serie de eventos. Esto lo harás como una generalización de las técnicas de conteo que has aprendido hasta ahora.

Problema: Un concesionario dispone de 15 vehículos, de los cuales 5 de ellos estarían con falla. Si 4 clientes diferentes entran y compran 4 autos de los 15, (a) ¿cuál es la probabilidad de que los cuatro compren un vehículo defectuoso?, (b) ¿Cuál es la probabilidad de al menos una persona compre el vehículo defectuoso? y (c) Si solamente se compra un auto, ¿cuál es la probabilidad de que este sea defectuoso?

Empieza preguntándote ¿qué información dispones? Tienes que hay $N = 15$ autos, de los cuales sabes que hay $m = 5$ con fallas y también $n = 4$ clientes. Esto es importante tenerlo en mente más adelante.

Vas a empezar con la pregunta (a):

> ¿Cuál es la probabilidad de que los cuatro compren un vehículo defectuoso?

Sabes que en un concesionario, si entra un cliente y se lleva un auto, entonces habrá un auto menos en el inventario. De manera similar, nota que tampoco es importante el orden en el cual se seleccionan los autos: los clientes entran y quieren un auto, no el primero, el segundo, el décimo, etc.

Así, denota con la letra E al evento

$$E : \text{4 clientes diferentes entran y compran 4 autos defectuosos.}$$

Alternativamente, pudes expresar E como

$$E : n \text{ clientes diferentes entran y compran } k \text{ autos defectuosos,}$$

donde la letra k indica el número de autos defectuosos comprados. Nota que es importante nombrar al número de clientes y el número de autos defectuosos con letras distintas, puesto que indican magnitudes diferentes para tu problema. Podrías preguntarte ¿y si compran solo 3 defectuosos? Allí cambiaría completamente el resultado.

E es un evento que contiene todas las formas en las cuales se pueden elegir los autos, estas pueden representarse en conjuntos de cuatro elementos donde cada elemento es un auto de los 15 disponibles. Puesto que te interesa determinar su probabilidad, usarás la regla de Laplace:

$$P(E) = \frac{|E|}{|\Omega|},$$

donde $|E|$ indica el número de elementos del conjunto E (también denominada cardinalidad), Ω es el espacio muestral y $|\Omega|$ es su cardinalidad.

Así pasas a la encrucijada: ¿cómo calcular el número de elementos en E? Esto se puede hacer usando una selección de elementos sin considerar el orden y sin reemplazo y el principio de multiplicación. ¿Cómo?

Primero, recuerda que el coeficiente binomial $\binom{\alpha}{\beta}$ te indica el número de *combinaciones sin repetición*, β elementos de un conjunto de α elementos, donde tampoco importa el orden. Hay $m = 5$ autos defectuosos, de los cuales vas a seleccionar $k = 4$, ¿cuántas formas hay de hacer esto? Justamente $\binom{m}{k}$. Por otro lado, hay $N - m = 10$ autos sin falla, de los cuales puedes admitir $n - k = 0$ en tu selección, el número de combinaciones de estos será $\binom{N-m}{n-k}$. Empleando el *principio de la multiplicación* tienes que el número de elementos en E está dado por

$$
\begin{aligned}
|E| &= \binom{m}{k} \times \binom{N-m}{n-k} = \binom{5}{4} \times \binom{15-5}{4-4} \\
&= \frac{5!}{4!1!} \times \frac{10!}{0!10!} = \frac{5 \times 4 \times 3 \times 2}{4 \times 3 \times 2} \times \frac{1}{1} = 5.
\end{aligned}
$$

Ahora calcula el número de elementos en Ω. Para hacerlo primero debes determinar ¿qué es el espacio muestral? Ω es el siguiente evento:

$$\Omega : \text{cuatro clientes diferentes entran y compran un auto cada uno.}$$

Nota que aquí no te interesa conocer las características de los autos, solo sabes que se compran los 4 y de paso sabes que hay 15 autos. Aquí te debe estar sonando lo siguiente: Entonces para determinar el tamaño de Ω basta saber el número de formas en las cuales puedo seleccionar los cuatro autos. Eso está nuevamente dado por el coeficiente binomial:

$$|\Omega| = \binom{N}{n} = \binom{15}{4} = \frac{15!}{4!11!} = 1365.$$

Con esta información, ya puedes determinar la probabilidad del evento E:

$$P(E) = \frac{|E|}{|\Omega|} = \frac{5}{1365} = \frac{1}{273} \approx 0.003663.$$

Lo anterior significa que si repitieses el experimento 1000 veces, entonces en *aproximadamente* 4 oportunidades ocurriría que los 4 autos estarían defectuosos.

Aquí hago una acotación para introducir el concepto de una variable hipergeométrica de parámetros N, m y n:

> Se dice que una variable aleaotria X sigue una distribución hipergeométrica de parámetros N, m y n si su función de masa de probabilidad está dada por:
>
> $$p_X(k) = \frac{\binom{m}{k} \times \binom{N-m}{n-k}}{\binom{N}{n}},$$
>
> donde N indica el tamaño de la población, m es el número de éxitos (o elementos de interés), n es el número de elementos seleccionados de la población y k es el número de aciertos dentro del conjunto de interés.

La función p_X determina la probabilidad de k éxitos (selecciones aleatorias de un objeto con una característica especial) en n selecciones, *sin reemplazo*, de una población finita de tamaño N que contiene exactamente m objetos con la característica especial. Es importante que contrastes esta distribución con la *distribución binomial*, donde en cambio se estudia la probabilidad de k sucesos en n selecciones con reemplazo.

Nota que en lo anterior podías obtener directamente la probabilidad de E al evaluar $p_X(4)$. Con esto puedes responder la pregunta que dejé suelta arriba «¿y si compran solo 3 defectuosos?». Allí, el resultado sería

$$p_X(3) = \frac{\binom{m}{3} \times \binom{N-m}{n-3}}{\binom{N}{n}} = \frac{\binom{5}{3} \times \binom{10}{1}}{\binom{15}{4}} = \frac{20}{273} \approx 0.007326.$$

Lo anterior significa que si repitieses el experimento 100 veces, entonces en *aproximadamente* 7 oportunidades ocurriría que los 3 autos estarían defectuosos.

Si entras a `https://www.wolframalpha.com` y escribes

```
hypergeometric distribution N = 15, m = 5, n = 4, x = 3
```

puedes determinar esta probabilidad automáticamente.

Ahora vas a responder la pregunta (b):

> ¿Cuál es la probabilidad de al menos una persona compre el vehículo defectuoso?

Aquí introduce el evento F como

F : 4 clientes diferentes entran y compran al menos un auto defectuoso.

El evento F abarca las posibilidades de que hayan tanto uno, como dos, tres o hasta cuatro autos defectuosos. Esto indica que debes determinar la probabilidad acumulada de estos eventos, la cual puedes representar como

$$P(F) = P(X \geq 1) = p_X(1) + p_X(2) + p_X(3) + p_X(4),$$

donde X es la variable aleatoria que has definido previamente. Creo que con esto ya puedes resolver el ejercicio, sin embargo no es la forma más rápida. ¿Recuerdas la propiedad del complemento en probabilidad? Esta propiedad lo que te dice es que si F es un evento, entonces la probabilidad de su complemento, denotado como F^c está dada por

$$P(F^c) = 1 - P(F).$$

El complemento de F es que X sea menor a 1, lo cual significa que X va a tomar el valor de 0 (aquí es importante que notes que el conjunto de valores de X en este caso es $\{0, 1, 2, 3, 4\}$), entonces

$$P(F^c) = p_X(0) = \frac{\binom{m}{0} \times \binom{N-m}{n-0}}{\binom{N}{n}} = \frac{2}{13}$$

y así puedes obtener la probabilidad de F:

$$P(F) = 1 - \frac{2}{13} = \frac{11}{13} \approx 0.8462.$$

Aquí nota que la propiedad del *complemento* fue clave para responder rápidamente a este ejercicio.

Como un dato curioso, si quieres obtener la cardinalidad de F, basta que uses la regla de Laplace y la información del ejercicio anterior:

$$P(F) = \frac{|F|}{|\Omega|} = \frac{11}{13},$$

¿qué conoces aquí?

Finalmente es el turno de la pregunta (c):

Si solamente se compra un auto, ¿cuál es la probabilidad de que este sea defectuoso?

Aquí cambian los parámetros de la variable aleatoria con la que has estado trabajando hasta ahora, puesto que el número de clientes ha bajado de 4 a 1, es decir, ahora $n = 1$. Los otros parámetros se mantienen, es decir $N = 15$ y $m = 5$. Si defines

$$G : \text{1 cliente entra y compra 1 auto defectuoso,}$$

el cálculo de su probabilidad sigue el mismo razonamiento que antes:

$$P(G) = p_X(1) = \frac{\binom{m}{1} \times \binom{N-m}{n-1}}{\binom{N}{n}} = \frac{\binom{m}{1} \times \binom{N-m}{1-1}}{\binom{N}{1}} = \frac{m}{N} = \frac{5}{15} = \frac{1}{3}.$$

Lo anterior es bastante bonito, pues te indica que la distribución hipergeométrica es una generalización de la regla de Laplace.

Para ver esto, fíjate que si no supieses nada de la distribución, podrías simplemente calcular $|G|$ contando el número de formas en las cuales un cliente puede entrar y elegir un auto defectuoso. Como hay 5 autos defectuosos y se elige uno:

$$|G| = \binom{m}{1} = \binom{5}{1} = 5.$$

Por otro lado, el espacio muestral de este experimento cambia a

$$\Omega : \text{1 cliente entra y compra 1 auto.}$$

Hay 15 autos y se selecciona uno, entonces

$$|\Omega| = \binom{N}{1} = \binom{15}{1} = 15$$

y con esto tienes nuevamente los ingredientes para el cálculo de probabilidad mediante la regla de Laplace:

$$P(G) = \frac{|G|}{|\Omega|} = \frac{5}{15} = \frac{1}{3} \approx 0.3333.$$

Esto último indica que si repites este experimento 100 veces, entonces en aproximadamente 33 oportunidades se dará un auto defectuoso al cliente.