

Probabilidad: Relación entre Distribución Hipergeométrica y de Bernoulli para $n \in \{2, 3\}$ *

García Prado, Sergio
sergio@garciparedes.me

20 de septiembre de 2017

1. Distribución Hipergeométrica

La distribución hipergeométrica $X \sim H(n, N, D)$ se corresponde con una distribución de probabilidad para variables aleatorias discretas. Esta modeliza la situación de obtener k casos favorables, de n observaciones, sobre una población de N individuos, donde D representan la situación de éxito y $N - D$ la de fracaso. Por tanto, modeliza la misma situación que la distribución binomial, solo que en este caso se presupone que cada observación se realiza **sin reemplazamiento**, mientras que en la distribución binomial se asume reemplazamiento.

En las siguientes secciones se demuestra que una distribución hipergeométrica $X \sim H(n, N, D)$ es equivalente a la suma de n distribuciones de bernoulli dependientes entre sí. Por tanto, $X = X_1 + \dots + X_i + \dots + X_n$ donde $X_i \sim B(\frac{D}{N})$ representa la i -ésima observación.

2. Demostración para $n = 2$

A continuación se demuestra el caso para dos variables:

$$X \sim H(2, N, D) \tag{1}$$

$$X = X_1 + X_2 \tag{2}$$

$$X_1 \sim B\left(\frac{D}{N}\right) \qquad X_2 \sim B\left(\frac{D}{N}\right) \tag{3}$$

$$P[X_1 = 1] = \frac{D}{N} \tag{4}$$

$$P[X_2 = 1] = P[X_2 = 1, X_1 = 0] + P[X_2 = 1, X_1 = 1] \tag{5}$$

$$= P[X_2 = 1 \mid X_1 = 0] * P[X_1 = 0] + P[X_2 = 1 \mid X_1 = 1] * P[X_1 = 1] \tag{6}$$

$$= \dots \tag{7}$$

$$= \frac{D}{N-1} * \left(1 - \frac{D}{N}\right) + \frac{D-1}{N-1} * \frac{D}{N} \tag{8}$$

$$= \frac{D}{N} \tag{9}$$

*URL: <https://github.com/garciparedes/probability-hypergeometric-to-bernoulli>

3. Demostración para $n = 3$

A continuación se demuestra el caso para tres variables (nótese caso de que ya haya ocurrido un éxito puede ser tratado de la misma manera sin importar si el éxito se ha dado en X_1 o en X_2):

$$X \sim H(3, N, D) \quad (10)$$

$$X = X_1 + X_2 + X_3 \quad (11)$$

$$X_1 \sim B\left(\frac{D}{N}\right) \quad X_2 \sim B\left(\frac{D}{N}\right) \quad X_3 \sim B\left(\frac{D}{N}\right) \quad (12)$$

$$(13)$$

$$P[X_1 = 1] = \frac{D}{N} \quad (14)$$

$$P[X_2 = 1] = P[X_2 = 1, X_1 = 0] + P[X_2 = 1, X_1 = 1] \quad (15)$$

$$= P[X_2 = 1 \mid X_1 = 0] * P[X_1 = 0] + P[X_2 = 1 \mid X_1 = 1] * P[X_1 = 1] \quad (16)$$

$$= \frac{D}{N-1} * \left(1 - \frac{D}{N}\right) + \frac{D-1}{N-1} * \frac{D}{N} \quad (17)$$

$$= \frac{D}{N} \quad (18)$$

$$P[X_3 = 1] = P[X_3 = 1, X_2 = 0, X_1 = 0] + P[X_3 = 1, X_2 = 0, X_1 = 1] \quad (19)$$

$$+ P[X_3 = 1, X_2 = 1, X_1 = 0] + P[X_3 = 1, X_2 = 1, X_1 = 1] \quad (20)$$

$$= P[X_3 = 1 \mid X_2 = 0, X_1 = 0] * P[X_2 = 0, X_1 = 0] \quad (21)$$

$$+ P[X_3 = 1 \mid X_2 = 0, X_1 = 1] * P[X_2 = 0, X_1 = 1] \quad (22)$$

$$+ P[X_3 = 1 \mid X_2 = 1, X_1 = 0] * P[X_2 = 1, X_1 = 0] \quad (23)$$

$$+ P[X_3 = 1 \mid X_2 = 1, X_1 = 1] * P[X_2 = 1, X_1 = 1] \quad (24)$$

$$= P[X_3 = 1 \mid X_2 = 0, X_1 = 0] * P[X_2 = 0 \mid X_1 = 0] * P[X_1 = 0] \quad (25)$$

$$+ P[X_3 = 1 \mid X_2 = 0, X_1 = 1] * P[X_2 = 0 \mid X_1 = 1] * P[X_1 = 1] \quad (26)$$

$$+ P[X_3 = 1 \mid X_2 = 1, X_1 = 0] * P[X_2 = 1 \mid X_1 = 0] * P[X_1 = 0] \quad (27)$$

$$+ P[X_3 = 1 \mid X_2 = 1, X_1 = 1] * P[X_2 = 1 \mid X_1 = 1] * P[X_1 = 1] \quad (28)$$

$$= \frac{D}{N-2} * \left(1 - \frac{D}{N-1}\right) * \left(1 - \frac{D}{N}\right) \quad (29)$$

$$+ 2 * \frac{D-1}{N-2} * \frac{D}{N-1} * \left(1 - \frac{D}{N}\right) \quad (30)$$

$$+ \frac{D-2}{N-2} * \frac{D-1}{N-1} * \frac{D}{N} \quad (31)$$

$$= \dots \quad (32)$$

$$= \frac{D}{N} \quad (33)$$

Referencias

[RdT18] María Pilar Rodríguez del Tío. Probabilidad, 2017/18.