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

García Prado, Sergio
sergio@garciparedes.me

21 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, $X = X_1 + ... + X_i + ... + X_n$ 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[1]:

$$X \sim H(2, N, D) \tag{1}$$
$$X = X_1 + X_2 \tag{2}$$
$$X_1 \sim B(\frac{D}{N}) \qquad X_2 \sim B(\frac{D}{N}) \tag{3}$$

$$\begin{aligned}
P[X_1 = 1] &= \frac{D}{N} && (4)\\
P[X_2 = 1] &= P[X_2 = 1, X_1 = 0] + P[X_2 = 1, X_1 = 1] && (5)\\
&= P[X_2 = 1 \mid X_1 = 0] * P[X_1 = 0] + P[X_2 = 1 \mid X_1 = 1] * P[X_1 = 1] && (6)\\
&= ... && (7)\\
&= \frac{D}{N-1} * (1 - \frac{D}{N}) + \frac{D-1}{N-1} * \frac{D}{N} && (8)\\
&= \frac{D}{N} && (9)
\end{aligned}$$

[*] URL: https://github.com/garciparedes/probability-hypergeometric-to-bernoulli

[1] El proceso de simplificación de cálculo simbólico se ha realizado a través de la plataforma *Wolfram Alpha*

3. Demostración para $n = 3$

A continuación se demuestra el caso para tres variables[1] (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):

$$X \sim H(3, N, D) \tag{10}$$
$$X = X_1 + X_2 + X_3 \tag{11}$$
$$X_1 \sim B(\frac{D}{N}) \qquad X_2 \sim B(\frac{D}{N}) \qquad X_3 \sim B(\frac{D}{N}) \tag{12}$$
$$\tag{13}$$

$$
\begin{aligned}
P[X_1 = 1] &= \frac{D}{N} && (14)\\
P[X_2 = 1] &= P[X_2 = 1, X_1 = 0] + P[X_2 = 1, X_1 = 1] && (15)\\
&= P[X_2 = 1 \mid X_1 = 0] * P[X_1 = 0] + P[X_2 = 1 \mid X_1 = 1] * P[X_1 = 1] && (16)\\
&= \frac{D}{N-1} * (1 - \frac{D}{N}) + \frac{D-1}{N-1} * \frac{D}{N} && (17)\\
&= \frac{D}{N} && (18)\\
P[X_3 = 1] &= P[X_3 = 1, X_2 = 0, X_1 = 0] + P[X_3 = 1, X_2 = 0, X_1 = 1] && (19)\\
&\quad + P[X_3 = 1, X_2 = 1, X_1 = 0] + P[X_3 = 1, X_2 = 1, X_1 = 1] && (20)\\
&= P[X_3 = 1 \mid X_2 = 0, X_1 = 0] * P[X_2 = 0, X_1 = 0] && (21)\\
&\quad + P[X_3 = 1 \mid X_2 = 0, X_1 = 1] * P[X_2 = 0, X_1 = 1] && (22)\\
&\quad + P[X_3 = 1 \mid X_2 = 1, X_1 = 0] * P[X_2 = 1, X_1 = 0] && (23)\\
&\quad + P[X_3 = 1 \mid X_2 = 1, X_1 = 1] * P[X_2 = 1, X_1 = 1] && (24)\\
&= P[X_3 = 1 \mid X_2 = 0, X_1 = 0] * P[X_2 = 0 \mid X_1 = 0] * P[X_1 = 0] && (25)\\
&\quad + P[X_3 = 1 \mid X_2 = 0, X_1 = 1] * P[X_2 = 0 \mid X_1 = 1] * P[X_1 = 1] && (26)\\
&\quad + P[X_3 = 1 \mid X_2 = 1, X_1 = 0] * P[X_2 = 1 \mid X_1 = 0] * P[X_1 = 0] && (27)\\
&\quad + P[X_3 = 1 \mid X_2 = 1, X_1 = 1] * P[X_2 = 1 \mid X_1 = 1] * P[X_1 = 1] && (28)\\
&= \frac{D}{N-2} * (1 - \frac{D}{N-1}) * (1 - \frac{D}{N}) && (29)\\
&\quad + 2 * \frac{D-1}{N-2} * \frac{D}{N-1} * (1 - \frac{D}{N}) && (30)\\
&\quad + \frac{D-2}{N-2} * \frac{D-1}{N-1} * \frac{D}{N} && (31)\\
&= ... && (32)\\
&= \frac{D}{N} && (33)
\end{aligned}
$$

Referencias

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