

Wyprowadzenie wariancji rozkładu łącznego $X_1/n_1 - X_2/n_2$ pod warunkiem $p_1 = p_2$, gdzie X_1 i X_2 to niezależne zmienne losowe z rozkładu hipergeometrycznego, a $p_1 = M_1/N_1$, $p_2 = M_2/N_2$.

Zapiszmy łączną wariancję rozważanej zmiennej losowej, korzystając z własności wariancji oraz tego, że $Cov(X_1, X_2) = 0$ z niezależności X_1 i X_2

$$Var\left(\frac{X_1}{n_1} - \frac{X_2}{n_2}\right) = Var\left(\frac{X_1}{n_1}\right) + Var\left(\frac{X_2}{n_2}\right) = \frac{1}{n_1^2}Var(X_1) + \frac{1}{n_2^2}Var(X_2). \quad (1)$$

Wiemy, że wariancje X_1 i X_2 są równe

$$Var(X_1) = n_1 p_1 (1 - p_1) (N_1 - n_1) / (N_1 - 1), \quad (2)$$

$$Var(X_2) = n_2 p_2 (1 - p_2) (N_2 - n_2) / (N_2 - 1). \quad (3)$$

Pamiętając, że zakładamy równość $p_1 = p_2$ zastąpmy oba parametry jednym p . Po podstawieniu otrzymujemy

$$\begin{aligned} Var\left(\frac{X_1}{n_1} - \frac{X_2}{n_2}\right) &= \frac{1}{n_1} p (1 - p) \frac{N_1 - n_1}{N_1 - 1} + \frac{1}{n_2} p (1 - p) \frac{N_2 - n_2}{N_2 - 1} = \\ &= p (1 - p) \left(\frac{N_1 - n_1}{n_1 (N_1 - 1)} + \frac{N_2 - n_2}{n_2 (N_2 - 1)} \right). \end{aligned} \quad (4)$$

Teraz pozostaje jedynie zastanowić się jak możemy wyliczyć parametr p . Czy to już oczywiste, że $p = (X_1 + X_2) / (n_1 + n_2)$, czy trzeba/można to jakoś pokazać?