

Wstęp

Już na początku XVIII wieku ukazały się pierwsze prace związane z rozkładem dwumianowym (Bernoulliego). Na przestrzeni tych kilku wieków teoria rachunku prawdopodobieństwa i statystyki znacząco się rozwinęła. W przypadku dyskretnym, najczęściej testowane są proporcje populacji. Albo czy dana próbka ma jakąś konkretną proporcję, albo czy dwie próbki mają tę samą. Obecnie znana jest powszechnie teoria dotycząca testowania hipotez, gdy populacja jest nieskończona (a raczej na tyle duża, że możemy ją w przybliżeniu uznać za nieskończoną). Wtedy proporcja to parametr p rozkładu dwumianowego. Jednak przypadek nieskończonej populacji nie wyczerpuje tematu testowania proporcji. Gdy populacja jest bardzo mała, albo gdy próbka jest niewiele mniejsza od całej populacji rozkład dwumianowy nie jest dobrym modelem. Natomiast dobrze taką sytuację modeluje rozkład hipergeometryczny, ponieważ uwzględnia on rozmiar populacji. Załóżmy, że N będzie rozmiarem populacji, n rozmiarem próbki, a M ilością osobników w populacji z daną cechą (której proporcje będziemy testować). Wtedy zmienna losowa X z rozkładu hipergeometrycznego określa ilość osobników z daną cechą w próbce. Funkcja prawdopodobieństwa określona jest wzorem

$$h(k; n, M, N) = P(X = k) = \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}}, \quad (1)$$

gdzie $L \leq k \leq U$, $L = \max\{0, M - N + n\}$ i $U = \min\{n, M\}$.

Bibliografia

- [1] F. Y. Edgeworth. On the value of a mean as calculated from a sample.
Journal of the Royal Statistical Society, 81(4):624–632, Jul. 1918.