<u>נוסחת השונות של משתנה היפרגאומטרי</u>

$X \sim HG(n;a,b)$. המשתנה סופר את מספר הכדורים הלבנים בהוצאה ללא החזרה של n כדורים מתוך כד שבו a כדורים לבנים ו b כדורים שחורים.

$X = \sum_{i=1}^{n} X_i$ כאשר X_i הוא אינדיקטור לכדור לבן בהוצאה ה- i . X_i הם תלויים כי אם בפעם מסוימת קבלנו כדור לבן, אז קטן הסיכוי שבפעם אחרת נקבל כדור לבן.

$$V(X) = \sum_{i=1}^{n} V(X_i) + \sum_{i \neq j} Cov(X_i, X_j)$$

לכל X_i יש הסתברות הצלחה של $\frac{a}{a+b}$.

לכן לכל i מתקיים $V(X_i) = \frac{a}{a+b}\frac{b}{a+b}$ (שונות של אינדיקטור).

ההתפלגויות המשותפות (X_i, X_j) הן זהות עבור כל זוג (i, j).

לכן עבור כל זוג (i, j) מתקיים $Cov(X_i, X_j) = Cov(X_1, X_2)$.

מתקיים $V(X) = nV(X_1) + 2\binom{n}{2}Cov(X_1, X_2)$.

$$Cov(X_1, X_2) = E(X_1X_2) - E(X_1)E(X_2)$$

$E(X_1) = E(X_2) = \frac{b}{a+b}\cdot 0 + \frac{a}{a+b}\cdot 1 = \frac{a}{a+b}$ (תוחלת של אינדיקטור).

המכפלה X_1X_2 שווה ל 0 או ל 1 (כי כל אחד משני המשתנים מקבל את הערך 0 או את הערך 1).

היא שווה ל 1 רק אם גם X_1 וגם X_2 שווים ל 1. זה קורה אם שני הכדורים הראשונים הם לבנים ובסיכוי $\frac{a}{a+b}\cdot\frac{a-1}{a+b-1}$.

לכן $Cov(X_1, X_2) = \frac{a}{a+b}\frac{a-1}{a+b-1} - \left(\frac{a}{a+b}\right)^2$ ומתקבל

$$V(X) = n\frac{a}{a+b}\frac{b}{a+b} + \binom{n}{2}\left[\frac{a}{a+b}\frac{a-1}{a+b-1} - \left(\frac{a}{a+b}\right)^2\right]$$

אם מפתחים את הביטוי אז מקבלים את נוסחת השונות של משתנה $X \sim HG(n;a,b)$.

$$V(X) = n\frac{a}{a+b}\frac{b}{a+b}\left[1-\frac{n-1}{a+b-1}\right]$$

שלומי