Hypothesis Testing for Sampling

Anastassia Kornilova

Let $x_1 \cdots x_N$ be a population with an unknown mean μ. Suppose we pick a sample without replacement of $y_1 \cdots y_n$. We want to test if the sample sum will have the same sign as the population sum, by using a hypothesis test. Let:

$$S = \sum_{i=1}^{N} x_i = n\mu \quad S_n = \sum_{j=1}^{n} y_j$$

Let the null hypothesis be $S < 0$.

By the bound defined in Serfling in *Probability Inequalities for Sum In Sampling Without Replacement*:

$$P(|S_n - S| > t) \leq 2\exp(-2t^2 * N/(n(N-n-1)(b-a)^2))$$

where b and a represent bounds on the population. Flipping the bounds and expanding the absolute value:

$$P(-t < S_n - S < t) \geq 2\exp(-2t^2 * N/(n(N-n-1)(b-a)^2))$$

Rearranging the terms, leads to a confidence level expression for $n\mu$.

$$P(S_n - t < S < S_n + t) \geq 2\exp(-2t^2 * N/(n(N-n-1)(b-a)^2))$$

If we want this to represent an $\alpha\%$ coincidence interval, we set the right hand side equal to α and solve for t.

$$\alpha = 2\exp(-2t^2 * N/(n(N-n-1)(b-a)^2))$$

$$t = \sqrt{\frac{n(N-n-1)(b-1)^2}{-2N} * \log(\alpha/2)}$$

Using this value of t, we can test if the $\alpha\%$ confidence interval contains S, the hypothesized value. If it doesn't, then performing the significance test will lead to a p-value $< \frac{100-\alpha}{100}$. The value of S is unknown, so we can not test if $S_n - t < S$. If $S_n - t > 0$, then $S_n - t > S$. Since S can represent any negative value, it follows that if $S_n - t > 0$, we can reject H_0.