Lecture 47: Runs: the null distribution of the number of runs

Statistical Inference–II ▪ Units IV–V: Nonparametric Inference

Kapil Kumar

Department of Statistics
Chaudhary Charan Singh University, Meerut

M.Sc. Statistics, Semester III
Academic Session 2026–2027

This lecture

Previous lecture

Lecture 46: One-sided bounds; duality with the sign test; principles of nonparametric testing.

Today

- Tests for Randomness Based on Runs
  - Conditional Distribution of the Number of Runs

Runs

A *run* is a maximal uninterrupted sequence of identical symbols. For example,

$$A\,A\,B\,B\,A\,B\,B$$

contains four runs: AA, BB, A and BB.

Runs tests detect clustering or excessive alternation.

- Too few runs indicate clustering or persistence;
- too many runs indicate rapid alternation.

Conditional distribution of the number of runs

Suppose a sequence contains n_1 symbols of type A, n_2 symbols of type B, and $n = n_1 + n_2$. Conditional on (n_1, n_2), all $\binom{n}{n_1}$ arrangements are equally likely under randomness.

Lemma 47.1 (Compositions)

The number of ways of dividing r identical symbols, placed in a row, into k non-empty runs is $\binom{r-1}{k-1}$.

Proof.

A division into k runs is obtained by choosing $k-1$ of the $r-1$ gaps between adjacent symbols at which to cut, and every choice gives a different division. □

Null distribution of the number of runs

Theorem 47.2 (Null distribution of the number of runs)

Under randomness, conditionally on (n_1, n_2),

$$\mathbb{P}(R = 2k) = \frac{2\binom{n_1-1}{k-1}\binom{n_2-1}{k-1}}{\binom{n}{n_1}},$$

$$\mathbb{P}(R = 2k+1) = \frac{\binom{n_1-1}{k}\binom{n_2-1}{k-1} + \binom{n_1-1}{k-1}\binom{n_2-1}{k}}{\binom{n}{n_1}},$$

where a binomial coefficient with an invalid lower index is interpreted as zero.

Proof of Theorem: Null distribution of the number of runs

Proof.

Runs of A's and B's must alternate, and an arrangement is completely determined by the symbol it starts with and by the lengths of its successive runs.

Even case, $R = 2k$. There are k runs of each type. The sequence may start with A or with B (2 ways); the n_1 A's can be split into k runs in $\binom{n_1-1}{k-1}$ ways and the n_2 B's in $\binom{n_2-1}{k-1}$ ways (Compositions Lemma).

Odd case, $R = 2k+1$. Either there are $k+1$ runs of A and k runs of B (the sequence must then start and end with A), giving $\binom{n_1-1}{k}\binom{n_2-1}{k-1}$ arrangements, or k runs of A and $k+1$ runs of B, giving $\binom{n_1-1}{k-1}\binom{n_2-1}{k}$ arrangements.

Dividing by the total number $\binom{n}{n_1}$ of equally likely arrangements gives the result. □

After this lecture

Practice (Lecture Notes, Chapter 3)

- Short-answer questions (Section 3.14): Q5.
- Exercises (Section 3.15): 3.8.

Next lecture

Lecture 48: Mean and variance of the number of runs; the one-sample runs test.