

Theory of Runs

by

Yi-Ling Lin

and

Dr. Ananda Jayawardhana

February 22, 2001

Definition

Given an ordered sequence of two or more types of symbols, a run is an unbroken sequence of similar events or like events.

Example

Consider the following sequence. It has 3 runs of symbol “a” and 2 runs of symbol “b”.

a a b a b b a a

In a sequence of two kinds of symbols, The total number of runs may be used as a measure of randomness in order of the appearance of the objects in the sequence.

Two many runs may indicate some pattern while too few runs may indicate some clustering.

Notation

Consider only two types of symbols a's and b's.

Let n_1 = Number of a's,

n_2 = Number of b's,

r_1 = Number of runs of a's,

r_2 = Number of runs of b's, and

$r = r_1 + r_2$.

If $r_1 = r_2$, then the sequence can begin with a run of either a's or b's.

If $r_1 = r_2 + 1$, then the sequence must begin with a run of a's.

If $r_1 = r_2 - 1$, then the sequence must begin with a run of b's.

Null hypothesis:

H_o : Randomness in the appearance of the two symbols.

There are $n_1 - 1$ spaces between n_1 a's and these a's can be separated into r_1 runs by selecting and widening $r_1 - 1$ of these spaces, and then filling them with runs of b's. The $r_1 - 1$ spaces

can be selected in $\binom{n_1 - 1}{r_1 - 1}$ different ways.

Example:

Let $n_1 = 5$, $n_2 = 4$, $r_1 = 3$, and $r_2 = 2$.

a a a a a 4 spaces

a a a a a 3 runs,
2 spaces

a a a a a

a a a a a

a a a a a

a a a a a

a a a a a

$$\binom{n_1 - 1}{r_1 - 1} = \binom{5 - 1}{3 - 1} = 6$$

Each of these gaps can be filled in

$$\binom{n_2 - 1}{r_2 - 1} = \binom{4 - 1}{2 - 1} = 3 \text{ ways.}$$

b bbb

bb bb

bbb b

If $r_1 = r_2 \pm 1$, then by the multiplication rule, the total number of ways of obtaining r_1 runs of a's and r_2 runs of b's is

$$\binom{n_1 - 1}{r_1 - 1} \binom{n_2 - 1}{r_2 - 1}.$$

Joint Distribution

However if $r_1 = r_2$, then the number of such arrangements will be $2 \binom{n_1 - 1}{r_1 - 1} \binom{n_2 - 1}{r_2 - 1}$ because either a's or b's can have the first run.

Then the joint distribution of R_1 and R_2 is given by

$$\begin{aligned}
 & P(R_1 = r_1, R_2 = r_2) \\
 &= \frac{c \binom{n_1 - 1}{r_1 - 1} \binom{n_2 - 1}{r_2 - 1}}{\binom{n_1 + n_2}{n_1}}, \text{ where } r_i = 1, \\
 & 2, \dots, n_i; i=1,2 \text{ and } c = \begin{cases} 1 & \text{if } r_1 = r_2 \pm 1 \\ 2 & \text{if } r_1 = r_2. \end{cases}
 \end{aligned}$$

Marginal Distribution

Marginal probability density function (p.d.f.) of R_1 is

$$P(R_1 = r_1) = \sum_{r_2 \in \{r_1, r_1+1, r_1-1\}} P(R_1 = r_1, R_2 = r_2).$$

$$= \frac{\binom{n_1 - 1}{n_1 - r_1} \binom{n_2 + 1}{r_1}}{\binom{n_1 + n_2}{n_1}}, \text{ where}$$

$$1 \leq r_1 \leq n_1.$$

Probability distribution of $R = R_1 + R_2$, the total number of runs

When r is even,

$$P(R = 2k) = P(R_1 = k, R_2 = k)$$

$$= \frac{2 \binom{n_1 - 1}{k - 1} \binom{n_2 - 1}{k - 1}}{\binom{n_1 + n_2}{n_1}}$$

When r is odd,

$$\begin{aligned} P(R = 2k + 1) &= P(R_1 = k, R_2 = k + 1) \\ &\quad + P(R_1 = k + 1, R_2 = k) \end{aligned}$$

$$= \frac{\binom{n_1 - 1}{k - 1} \binom{n_2 - 1}{k}}{\binom{n_1 + n_2}{n_1}} + \frac{\binom{n_1 - 1}{k} \binom{n_2 - 1}{k - 1}}{\binom{n_1 + n_2}{n_1}}$$

Example

$$n_1=5, n_2=5$$

$$\begin{array}{cccccccccc} a & a & a & a & a & b & b & b & b & b \\ b & b & b & b & b & a & a & a & a & a \end{array} \quad r=2$$

$$\begin{array}{cccccccccc} a & b & a & b & a & b & a & b & a & b \\ b & a & b & a & b & a & b & a & b & a \end{array} \quad r=10$$

$$r \in \{2, 3, \dots, 10\}$$

$$P(R=2) = \frac{2 \binom{5-1}{1-1} \binom{5-1}{1-1}}{\binom{5+5}{5}} = \frac{2}{252}$$

$$P(R=3) = \frac{\binom{5-1}{1} \binom{5-1}{1-1} + \binom{5-1}{1-1} \binom{5-1}{1}}{\binom{5+5}{5}}$$

$$= \frac{8}{252}$$

Probability Distribution

R	2	3	4	5	6	7	8	9	10
$P(R)$	$\frac{2}{252}$	$\frac{8}{252}$	$\frac{32}{252}$	$\frac{48}{252}$	$\frac{72}{252}$	$\frac{48}{252}$	$\frac{32}{252}$	$\frac{8}{252}$	$\frac{2}{252}$

$$P(R=2) = P(R=10) = \frac{2}{252} = 0.00793$$

$$P(R=3) = P(R=9) = \frac{8}{252} = 0.03174$$

$$P(R=2 \text{ or } R=10) = 0.01586$$

$$P[(R=2) \cup (R=3) \cup (R=9) \cup (R=10)]$$

$$= 0.07934$$

H_o : Randomness in the appearance of the two symbols.

H_a : H_o is not true.

Rule:

If the level of significance, $a = 0.05$, then reject H_o if $R = 2$ or 10 .

Rule:

If the level of significance, $a = 0.10$, then reject H_o if $R = 2, 3, 9$, or 10 .

Notice that we cannot achieve the exact level of significance.

Tests of Randomness: Runs above and below the median

On 23 successive checks the sample averages of a process were recorded as below or above the standard average.

a b a a b b a b b a a a b a a b b a a b a b b
($R=14$)

1. H_o : Arrangement is random
 H_a : Arrangement is not random
2. Level of significance: $\alpha = 0.01$
3. Rejection criterion: Reject H_o if $R < 6$ or $R > 19$
4. Calculation: $R=14$
5. Decision: Do not reject H_o . The arrangement is random.

R	$P(R = r)$
2	0.0000015
3	0.0000155
4	0.0001627
5	0.0007729
6	0.0036610
7	0.0103729
8	0.0292883
9	0.0549155
10	0.1025089
11	0.1332615
12	0.1722149
13	0.1578637
14	0.1435124
15	0.0922580
16	0.0585765
17	0.0256272
18	0.0109831
19	0.0030509
20	0.0008136
21	0.0001220
22	0.0000163
23	0.0000007

Large Sample Approximation

$Z = \frac{R - E(R)}{\sqrt{Var(R)}}$ has an approximate standard normal distribution.

Claim: (a) $E(R) = \frac{2n_1 n_2}{n_1 + n_2} + 1$

(b) $Var(R) = \frac{2n_1 n_2 (2n_1 n_2 - n_1 - n_2)}{(n_1 + n_2)^2 (n_1 + n_2 - 1)}$

Proof: Define

$$I_k = \begin{cases} 1 & \text{if } k^{\text{th}} \text{ element } \neq (k-1)^{\text{th}} \text{ element} \\ 0 & \text{otherwise} \end{cases}$$

and $I_1 \equiv 0$.

Then, $R = 1 + \sum_{k=1}^n I_k$, where $n = n_1 + n_2$.

Consider the sequence

a a b a b b b a a

$$\begin{aligned} R &= 1 + I_1 + I_2 + \dots + I_9 \\ &= 1 + 0 + 0 + 1 + 1 + 1 + 0 + 0 + 1 + 0 = 5 \end{aligned}$$

Each I_k ($k > 1$) is a Bernoulli random variable but they are not independent.

$$\begin{aligned} \text{Let } P(I_k = 1) &= \left(\frac{n_1}{n} \right) \left(\frac{n_2}{n-1} \right) + \left(\frac{n_2}{n} \right) \left(\frac{n_1}{n-1} \right) \\ &= \frac{2n_1 n_2}{n(n-1)} = (\text{say } p). \end{aligned}$$

$$\begin{aligned} E(R) &= 1 + 0 + E \left[\sum_{k=2}^n I_k \right] = 1 + (n-1)p \\ &= 1 + \frac{2n_1 n_2}{n} \end{aligned}$$

$$\begin{aligned}
V(R) &= E\left\{\left[R - E(R)\right]^2\right\} \\
&= E\left[\left\{\sum_{k=2}^n [I_k - E(I_k)]\right\}^2\right] \\
&= \sum_{k=2}^n E[I_k - E(I_k)]^2 + \\
&\quad \sum_{2 \leq j \neq k \leq n} E\left\{\left[I_j - E(I_j)\right]\left[I_k - E(I_k)\right]\right\} \\
&= \sum_{k=2}^n Var(I_k) + 2 \sum_j \sum_{k < j} Cov(I_j, I_k)
\end{aligned}$$

If a random variable $X \sim Bernoulli(p)$, then $E(X) = p$ and $Var(X) = p(1-p)$.

Hence $E(I_k) = \frac{2n_1 n_2}{n(n-1)}$ and

$$Var(I_k) = \frac{2n_1 n_2}{n(n-1)} \left(1 - \frac{2n_1 n_2}{n(n-1)}\right).$$

$$COV(I_j, I_k) = E(I_j I_k) - E(I_j)E(I_k)$$

for $j \neq k$

Consider $P(I_j \cdot I_{j+1} = 1) = P(I_j = 1, I_{j+1} = 1)$
 for $j = 2, 3, \dots, n-1$.

$$P(I_j \cdot I_{j+1} = 1) =$$

$$P\{(j-1)^{th} \text{ element is } a, j^{th} \text{ element is } b, \text{ and } (j+1)^{th} \text{ element is } a\}$$

+

$$P\{(j-1)^{th} \text{ element is } b, j^{th} \text{ element is } a, \text{ and } (j+1)^{th} \text{ element is } b\}$$

$$\begin{aligned} &= \left(\frac{n_1}{n} \right) \left(\frac{n_2}{n-1} \right) \left(\frac{n_1-1}{n-2} \right) + \left(\frac{n_2}{n} \right) \left(\frac{n_1}{n-1} \right) \left(\frac{n_2-1}{n-2} \right) \\ &= \frac{n_1 n_2}{n(n-1)} \end{aligned}$$

There are $2(n-2)$ such quantities.

$$E(I_j I_{j+1}) = 1 \cdot P(I_j I_{j+1} = 1) + 0 \cdot P(I_j I_{j+1} = 0)$$

Consider $P(I_j I_k = 1)$ for $k \neq j$ or $j+1$ and

$2 \leq k, j \leq n$.

$$\begin{aligned}
P(I_j I_k = 1) &= \\
P\{_{(j-1)^{th} \text{ element is } a, j^{th} \text{ element is } b, (k-1)^{th} \text{ element is } a, \text{ and } k^{th} \text{ element is } b}\} + \\
P\{_{(j-1)^{th} \text{ element is } a, j^{th} \text{ element is } b, (k-1)^{th} \text{ element is } b, \text{ and } k^{th} \text{ element is } a}\} + \\
P\{_{(j-1)^{th} \text{ element is } b, j^{th} \text{ element is } a, (k-1)^{th} \text{ element is } a, \text{ and } k^{th} \text{ element is } b}\} + \\
P\{_{(j-1)^{th} \text{ element is } b, j^{th} \text{ element is } a, (k-1)^{th} \text{ element is } b, \text{ and } k^{th} \text{ element is } a}\} \\
&= \frac{4n_1 n_2 (n_1 - 1)(n_2 - 1)}{n(n-1)(n-2)(n-3)}
\end{aligned}$$

$$\begin{aligned}
V(R) &= (n-1) \frac{2n_1 n_2}{n(n-1)} \left(1 - \frac{2n_1 n_2}{n(n-1)} \right) + 2(n-2) \frac{n_1 n_2}{n(n-1)} \\
&\quad + [(n-1)(n-2) - 2(n-2)] \left[\frac{4n_1 n_2 (n_1 - 1)(n_2 - 1)}{n(n-1)(n-2)(n-3)} \right] \\
&= \frac{2n_1 n_2 (2n_1 n_2 - n_1 - n_2)}{(n_1 + n_2)^2 (n_1 + n_2 - 1)}
\end{aligned}$$