RUNS TEST IN RANDOM NUMBER GENERATION

Pankaj Kumar
p_kumar2@me.iitr.ac.in

Test for Random number Generations

Test for Uniformity:

Test for Independence : RUNS TEST

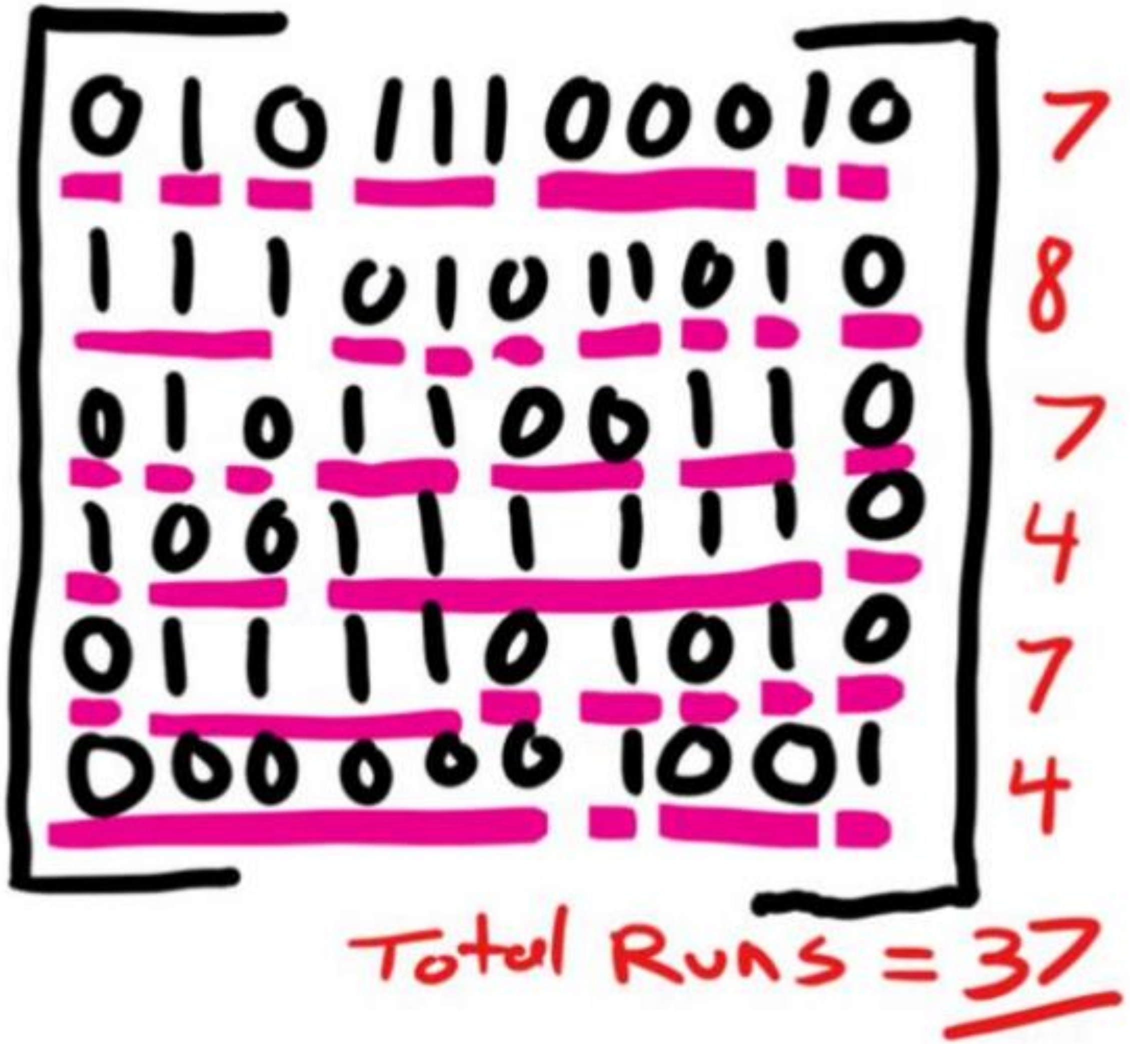
RUNS TEST

- The runs test examines the arrangement of numbers in a sequence to test the hypothesis of independence.
- The runs test is a statistical test that is used to check the randomness of data generated by a random number generator. It is a non-parametric test that uses runs of data to decide whether the presented data is random or tends to follow a pattern

‘RUN’

- *A "run" is a consecutive series of values that consistently increase or decrease.*

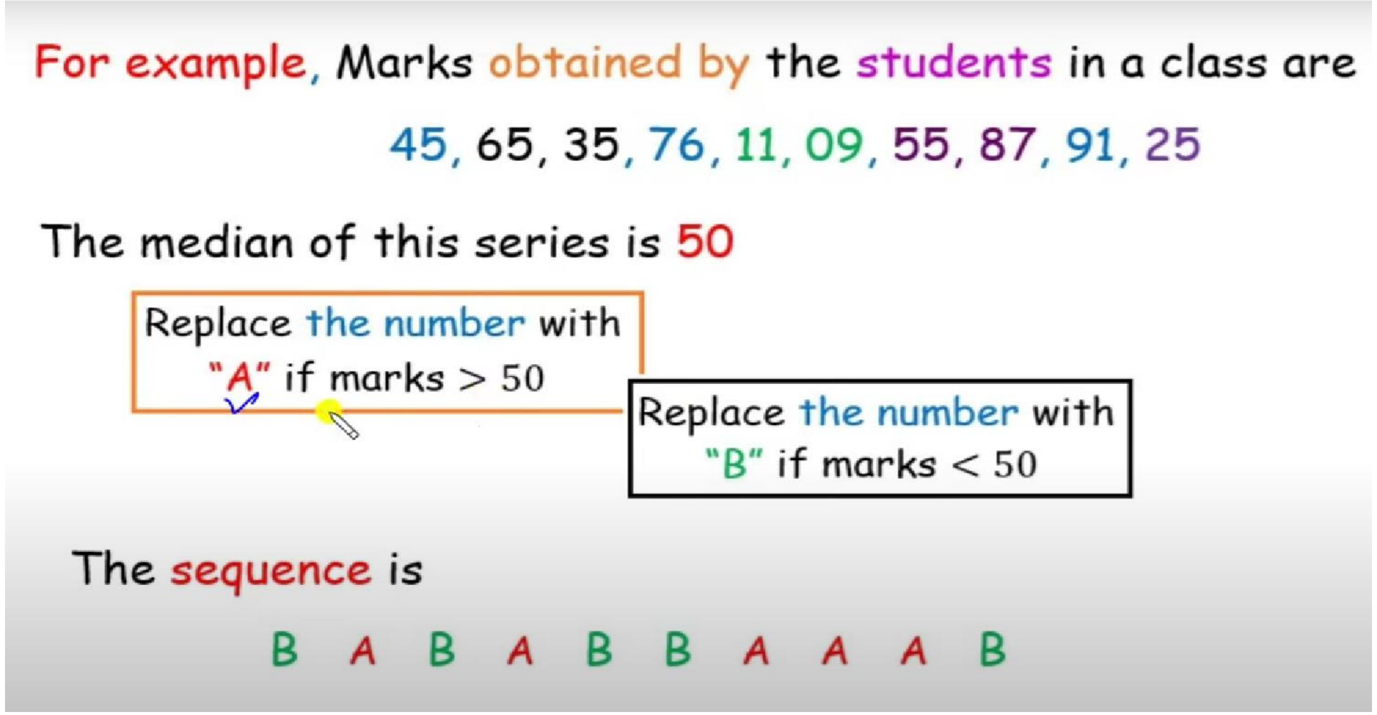
- *A “run” is defined as a series of increasing values or decreasing values. The number of increasing, or decreasing, values is the length of the run.*


01011100010 7
11101011010 8
0101100110 7
10011111110 4
01111101010 7
0000001001 4
Total Runs = 37

PROCEDURE TO COUNT NO. OF RUNS

There are several ways to define runs, however, in all cases the formulation must produce a dichotomous sequence of values.

1.To use Median value:



Above example has 7 runs.

PROCEDURE TO COUNT NO. OF RUNS

2. Normal '+', '-' sign for increment and decrement:

Consider the sequence of integers

6, 1, 6, 1, 4, 6, 7, 6, 10, 7, 7, 3, 3, 1, 8, 2, 5, 9, 8, 5, 10, 4, 4, 4, 6, 3, 9, 10, 8, 7

Flag each number with a + if it is followed by a larger number and with a – if it is followed by a smaller number.

\- + - + + + - + - + - + - + - + + - - + - + + + - + + - -

Each string of '+'s and '–'s forms a run, above sequence has 21 runs

- An **‘up run’** is a sequence of numbers each of which is succeeded by a larger number; a **‘down run’** is a sequence of numbers each of which is succeeded by a smaller number.

- If a sequence of numbers have too few runs, it is unlikely a real random sequence. E.g. 0.08, 0.18, 0.23, 0.36, 0.42, 0.55, 0.63, 0.72, 0.89, 0.91, the sequence has one run, an up run. It is not likely a random sequence.

- If a sequence of numbers have too many runs, it is unlikely a real random sequence. E.g. 0.08, 0.93, 0.15, 0.96, 0.26, 0.84, 0.28, 0.79, 0.36, 0.57. It has nine runs, five up and four down. It is not likely a random sequence.

STEPS TO PERFORM A TEST OF HYPOTHESIS

1. State the null and alternative hypotheses
2. Select the distribution to use
3. Determine the rejection and non-rejection regions
4. Calculate the value of the test statistic
5. Make a decision

APPLYING RUNS TEST

- In testing for independence, the hypotheses are as follows;

H_0: R_i ~ independently

H_a: $R_i \neq$ independently

This null hypothesis, H_0, reads that the numbers are independent. Failure to reject the null hypothesis means that no evidence of dependence has been detected on the basis of this test. This does not imply that further testing of the generator for independence is unnecessary.

Two types of errors are possible when performing a hypothesis test

1. Type I error: H_0 is rejected even though true
2. Type II error: H_0 is accepted even though false

- Level of significance α

  $\alpha = P(\text{reject } H_0 \mid H_0 \text{ true})$

  Frequently, α is set to 0.01 or 0.05

(Hypothesis)

	Actually True	Actually False
Accept	$1 - \alpha$	β (Type II error)
Reject	α (Type I error)	$1 - \beta$

When n1 and n2(here + and – signs) both are less than 20.

- If a is the total number of runs in a truly random sequence, the mean and variance of a is given by

$$\mu_a = \frac{2N-1}{3}$$

and

$$\sigma^2 = \frac{16N-29}{90}$$

- For $N > 20$, the distribution of a is reasonably approximated by a normal distribution, $N(\mu_a, \sigma_a^2$. Converting it to a standardized normal distribution by

$$Z_0 = \frac{a-\mu_a}{\sigma_a}$$

that is

$$Z_0 = \frac{a-[(2N-1)/3]}{\sqrt{(16N-29)/90}}$$

- Failure to reject the hypothesis of independence occurs when $-z_{\alpha/2} \leq Z_0 \leq z_{\alpha/2}$, where the α is the level of significance.

Substituting for μ_a and σ_a ==>

$$Z_a = \{a - [(2N-1)/3]\} \,/\, \{\sqrt{(16N-29)/90}\},$$

where $Z \sim N(0,1)$

Acceptance region for hypothesis of independence $-Z_{\alpha/2} \le Z_0 \le Z_{\alpha/2}$

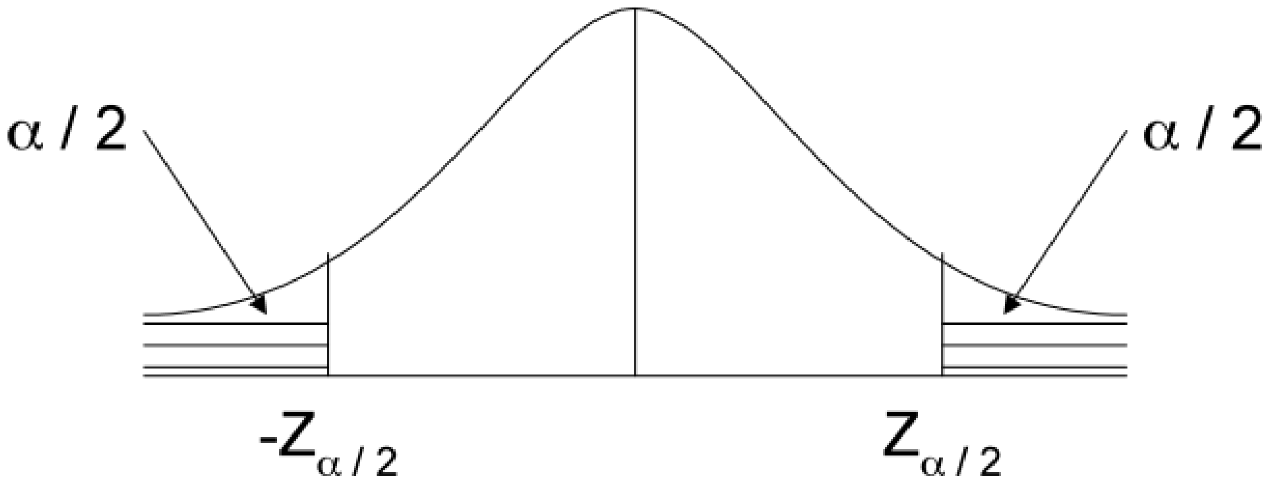


ILLUSTRATION:

- Given a sequence of 40 numbers, check its independence with 0.05 level of significance.

0.41	0.68	0.89	0.94	0.74	0.91	0.55	0.62	0.36	0.27
0.19	0.72	0.75	0.08	0.54	0.02	0.01	0.36	0.16	0.28
0.18	0.01	0.95	0.69	0.18	0.47	0.23	0.32	0.82	0.53
0.31	0.42	0.73	0.04	0.83	0.45	0.13	0.57	0.63	0.29

- H_0 = Sample data is independent
- H_a = Sample data is not independent

ILLUSTRATION:

The sequence of runs up and down is as follows:

$$+ + + - + - + - - - + + - + - - + - + - - + - - + - + + - \quad - + + - + - - + + -$$

There are 26 runs in this sequence. With N=40 and a=26,

$$\mu_a = \{2(40) - 1\} / 3 = 26.33 \text{ and}$$

$$\sigma_a^2 = \{16(40) - 29\} / 90 = 6.79$$

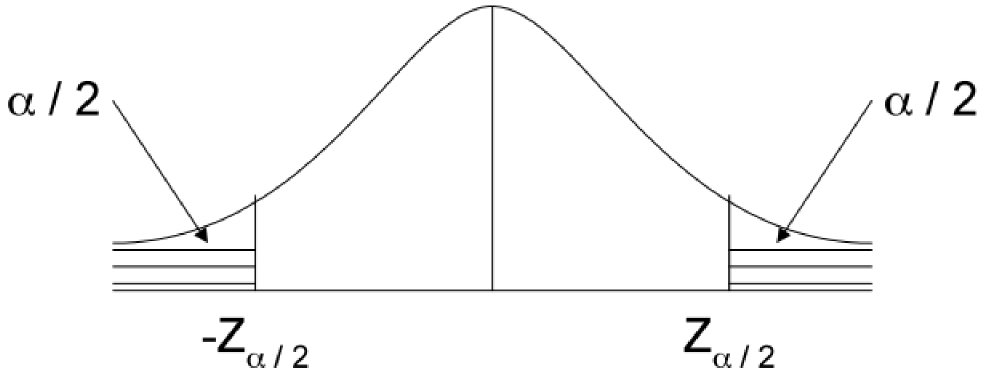
Then,

$$Z_0 = (26 - 26.33) / \sqrt{(6.79)} = -0.13$$

ILLUSTRATION:

$Z_{\alpha/2} = Z_{0.025} = 1.96$

$$-Z_{\alpha/2} \le Z_0 \le Z_{\alpha/2}$$



So, the independence of the numbers cannot be rejected on the basis of this test.

In case n1>20 or n2>20 or both >20.

- Let n_1 and n_2 be the number of individual observations above and below the mean, let b the total number of runs.
- For a given n_1 and n_2, the mean and variance of b can be expressed as

$$\mu_b = \frac{2n_1n_2}{N} + 1$$

and

$$\sigma_b^2 = \frac{2n_1n_2(2n_1n_2 - N)}{N^2(N-1)}$$

- For either n_1 or n_2 greater than 20, b is approximately normally distributed

$$Z_0 = \frac{b - (2n_1n_2/N) - 1}{\left[\frac{2n_1n_2(2n_1n_2-N)}{N^2(N-1)}\right]^{1/2}}$$

- Failure to reject the hypothesis of independence occurs when $-z_{\alpha/2} \le Z_0 \le z_{\alpha/2}$, where α is the level of significance.

```python
# simple code to implement Runs
# test of randomnes

import random
import math
import statistics

def runsTest(l, l_median):

    runs, n1, n2 = 0, 0, 0

    # Checking for start of new run
    for i in range(len(l)):

        # no. of runs
        if (l[i] >= l_median and l[i-1] < l_median) or \
                (l[i] < l_median and l[i-1] >= l_median):
            runs += 1

        # no. of positive values
        if(l[i]) >= l_median:
            n1 += 1
```

```python
        # no. of negative values
        else:
            n2 += 1

    runs_exp = ((2*n1*n2)/(n1+n2))+1
    stan_dev = math.sqrt((2*n1*n2*(2*n1*n2-n1-n2))/ \
                        (((n1+n2)**2)*(n1+n2-1)))

    z = (runs-runs_exp)/stan_dev

    return z

# Making a list of 100 random numbers
l = []
for i in range(100):
    l.append(random.random())

l_median= statistics.median(l)

Z = abs(runsTest(l, l_median))

print('Z-statistic= ', Z)
```


source: [Runs Test of Randomness in Python – GeeksforGeeks]

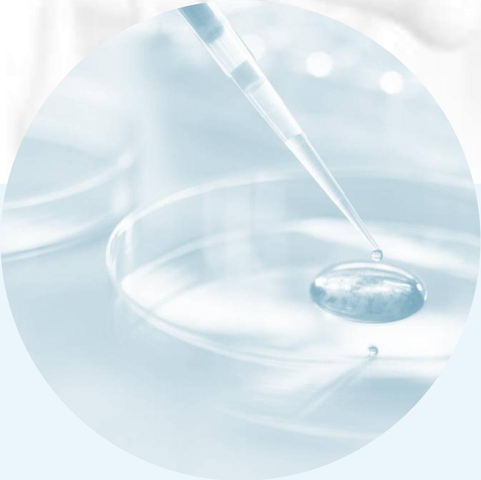
SAMPLE OUTPUTS

```
Z-statistic=  1.407124727947029
Z-statistic=  0.201017818278147
Z-statistic=  1.0050890913907349
Z-statistic=  0.603053454834441
```

In above sample outputs, each time value comes between -1.96 and 1.96, so we fail to reject null hypothesis.. This suggests that the sequence of numbers exhibits a random pattern, and there's no significant evidence to conclude otherwise.

REFERENCES

- 20110404randomno02.pdf (iiserpune.ac.in)
- https://www.eg.bucknell.edu/~xmeng/Course/CS6337/Note/master/node44.html
- Runs Test of Randomness in Python – GeeksforGeeks
- NIST SP 800-22, A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications



THANK YOU