Chapter 5

Probability Concept and Random Number Generation

Probability concepts in Simulation-Stochastic variable

Probability theory is applied to situations where uncertainty exists. In our daily life, we have many situations where uncertainty plays a vital role. The chance of rain tomorrow is 40%. This means, the probability of rain 0.4. The description of activities can be of two type: Deterministic and Stochastic. The process in which the outcome of an activity can be described completely in terms of its input is deterministic. On the other hand, when the outcome of an activity is random i.e. there can be various possible outcomes, the activity is said to be stochastic. In case of an automatic machine, the output per hour is deterministic, while in a repair shop, the number of machine repaired will vary from hour to hour in random fashion. The term random and stochastic are interchangeable.

Discrete Probability Function

If a random variable is discrete variable, its probability distribution is called discrete probability distribution. If a random variable X can take X_i (i= 1, 2………………..N) countably infinite no of values with the probability of value X_i being $P(X_i)$, the set of numbers $P(X_i)$ is said to be a probability distribution or probability mass function of random variable X. The number $P(X_i)$ must satisfy the following two conditions.

1. $P(X_i) \geq 0$ for all values of i
2. $\sum_{X=1}^{n} P(Xi) = 1$

Example:
If we flip a coin two times then we can have four possible outcomes/ Sample Space: HH, HT, TH, TT. Let X be a random variable that represents the number of heads in above experiment. The random variable X can only take values 0, 1 or 2. So it is discrete random variable. The probability distribution for this statistical experiment appears below.

Number of Head: X	0	1	2
P(X)	0.25	0.5	0.25

Continuous Probability Function

If a random variable is a continuous variable, its probability distributions is called continuous probability distribution. If a random variable is continuous and not limited to a discrete value, it will have infinite number of values in any interval. Such variable is defined by a function F(X) called probability density function(PDF). The probability that a variable X falls in the range x1 and x2 is given by

$$\int_{x1}^{x2} F(X)dx \quad F(X) \geq 0$$

Random variable and Random Number

A random variable is a function that assigns a real number to each outcome of the experiment.
Example:
Consider the experiment of rolling a pair of dice. Then S= {(1,1), (1,2), (1,3)………………..(6,6)}

If x is random variable corresponding to the sum of the two dice, then X assigns the value 2, 3, 4, …………12 to the above outcome of experiment.
There are two types of random variable: Discrete and Continuous.
A random variable X is said to be discrete if the number of possible values of X is finite or countably infinite i.e. possible values of x may be in the range X1, X2……………..X_N. A random variable X is said to be continuous if its range space is an interval or a collection of interval. A continuous variable can assume any value over a continuous range.

Random number is simply a value taken on by a random variable. **Random numbers** are numbers that occur in a sequence such that two conditions are met: (1) the values are uniformly distributed over a defined interval or set, and (2) it is impossible to predict future values based on past or present ones.

Properties of random number

- **The generated random numbers should satisfy the property of uniformity.**
- **The generated random numbers should satisfy the property of independence.**
- The random number should be replicable.
- It should take a long time before the number starts to repeat.
- The routine should be fast
- The routine should not require a lot of storage.

Pseudo Random number

Pseudo means false but her pseudo means that the random number are generated by using some known arithmetic operation and known mathematical formula. Since the arithmetic operation/ formula is known and the sequence of random numbers can be repeatedly obtained, the numbers can't be truly random. Truly random number are generated by physical generators. For example, tossing a coin. However, the pseudo random numbers generated by many computer routines very closely fulfilled the requirement of desire randomness.
If the method of random number generation i.e. random number generator, is defective, the generated pseudo random numbers,

- May not be uniformly distributed.
- May not be continuous.
- The mean of generated numbers may be too high or too low
- The variance may be too high or too low.

Generation of Random number

Random number can be generated by the following methods.

1. Random numbers may be drawn from the random number tables stored in the computer memory. It is very slow process and the number considerably occupy space of computer memory.
2. An electronic device may be constructed as part of a digital computer to generate truly random number. This is however considered as expensive.
3. Pseudo random numbers may be generated by using mathematical formulas and arithmetic operation. This method commonly specifies a procedure, where starting with an initial number, the second number is generated and from that third number and so on. A number of recursive procedure are used for generating random number.
   One of the method for generating PRN is Mid square method. It starts with a fixed initial value, say 4-digit integer, called seed. The number is squared and the middle four digit of this square

become the second number. The middle digit of this second number are then squared again to generate third random number and so on. We may also have to add zero to make the digit's length eight if necessary. Finally, we get realization from the uniform (0,1) distribution after placement of decimal points i.e. after division by 10000.

Example: if we take seed Z_0 = 1234, then we will get the sequence of numbers as 0.1234, 0.5227, 0.3215, 0.3362, 0.3030, 0.1809……………………...

Generate the random number sequence of number for Z_0 = 2100 using mid square method.

Quality of an efficient random number Generator

- It should have a sufficiently long cycle i.e. it should be sufficiently long sequence of random numbers before beginning to repeat the sequence.
- The random numbers generated should be replicable i.e. by specifying a starting condition, it should be possible to obtain the same set of random numbers. Many time common random numbers are required for the comparison of two systems.
- The generated random number should fulfil the requirement of uniformity and independence.
- The random number generator should be fast and cost effective.
- It should be portable to different computer and ideally to different programming language.

Testing number for randomness

A sequence of random number is considered to be random if

1. The number are uniformly distributed i.e. every number has an equal chance of occurrence.
2. The number are not serially auto-correlated i.e. there is no correlation between adjacent pair or number, or the appearance of one number doesn't influence the appearance of next number.

There are a number of test, which are used to ensure that random numbers are uniformly distributed and are not serially auto-correlated.

Uniformity Test(Frequency Test)

The test of uniformity or frequency test is a basic test that should always be performed to validate a random number generator. The uniformity test counts how often numbers in a given range occur in the sequence to ensure that the number are uniformly distributed. Two frequency test are available and they are

1. Kolmogorov-Smirnov test i.e. K-S test
2. Chi-Square test

Both of these test compare the generated random number with the theoretical uniform distribution. The algorithms of testing a random number generator are based on some statistics theory, i.e. testing the hypotheses. The basic ideas are the following, using testing of uniformity as an example.

We have two hypotheses one says the random number generator is indeed uniformly distributed. We call this H_o, known in statistics as *null hypothesis*. The other hypothesis says the random number generator is not uniformly distributed. We call this H_1, known in statistics as *alternative hypothesis*. We are interested in testing result of H0, reject it, or fail to reject it.

K-S Test

This test compares the cdf (Continuous distributed function) of uniform distribution *F(x)* to the empirical cdf, $S_N(X)$ of the sample of *N* observations. The largest deviation between F(X) and $S_N(X)$ is determined and is compared with the critical value, which is available as function of N.

Procedure of applying K-S uniformity test

1. Rank the data from smallest to largest

$$R_{(1)}, \leq R_{(2)}, \leq \dots \leq R_{(N)}$$

2. Number are computed from the empirical distribution $S_N(X)$ i.e. i/N
3. Compute the deviations

$$D^{+} = \max_{1 \leq i \leq N} \left\{ \frac{i}{N} - R_{(i)} \right\}$$

$$D^{-} = \max_{1 \leq i \leq N} \left\{ R_{(i)} - \frac{i-1}{N} \right\}$$

4. Compute the largest deviation i.e. D is also know as sample distribution

$$D = \max(D^{+}, D^{-})$$

1. Determine the critical value, D_α, from Table below for the specified significance level α and the given sample size *N*.

TABLE 8.3-13: CRITICAL VALUES $d_{\alpha;n}$ OF THE MAXIMUM ABSOLUTE DIFFERENCE BETWEEN SAMPLE AND POPULATION RELIABILITY FUNCTIONS

Sample Size, N	Level of Significance, α				
	0.20	0.15	0.10	0.05	0.01
3	0.565	0.597	0.642	0.708	0.828
4	0.494	0.525	0.564	0.624	0.733
5	0.446	0.474	0.474	0.565	0.669
10	0.322	0.342	0.368	0.410	0.490
15	0.266	0.283	0.304	0.338	0.404
20	0.231	0.246	0.264	0.294	0.356
25	0.21	0.22	0.24	0.27	0.32
30	0.19	0.20	0.22	0.24	0.29
35	0.18	0.19	0.21	0.23	0.27
40	0.17	0.18	0.19	0.21	0.25
45	0.16	0.17	0.18	0.20	0.24
50	0.15	0.16	0.17	0.19	0.23
over 50	$\frac{1.07}{\sqrt{N}}$	$\frac{1.14}{\sqrt{N}}$	$\frac{1.22}{\sqrt{N}}$	$\frac{1.36}{\sqrt{N}}$	$\frac{1.63}{\sqrt{N}}$

2. If the sample statistic *D* is greater than the critical value D_α , the null hypothesis that the sample data is from a uniform distribution is rejected; if $D \leq D_\alpha$, then there is no evidence to reject it.

   Note: the level α is the probability of rejecting the H_0 null while H_0 null is true.

Example:

The sequence of number 0.24, 0.89,0.11, 0.61, 0.23, 0.86, 0.41, 0.64, 0.50, 0.65 has been generated. Use the K-S test with a level of significance $\alpha = 0.05$ and $D_{0.05} = 0.410$ to determine if the hypothesis that the number are uniformly distributed on the interval [0,1] can be rejected.

R_i	0.11	0.23	0.24	0.41	0.50	0.61	0.64	0.65	0.86	0.89
i / N	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	1
(I / N) – R_i	-	-	0.06	-	0	-	0.06	0.15	0.04	0.11
R_i - (i-1)/N	0.11	0.13	0.04	0.11	0.10	0.11	0.04	-	0.06	-

From the above table we get,
$D^+ = \text{Max}(i / N – R_i) = 0.15$
$D^- = \text{Max}(R_i - (i-1)/N) = 0.13$
$D=\text{Max}(D^+, D^-) =0.15$
The critical value $D_{0.05}$ is 0.410 (i.e. the critical value of D for $\alpha = 0.05$ and N=10 is 0.410).
Since $D<D_\alpha$, there is no chance to reject the hypothesis that the given random numbers are uniform.
(or, since the computed value is less than critical value, the given random numbers are uniform at 95% level of significance)

The sequence of number 0.54, 0.73, 0.98, 0.11 and 0.68 has been generated. Use the K-S test with $\alpha = 0.05$ to determine if the hypothesis that the number are uniformly distributed on the interval [0,1] can be rejected.

R_i	0.11	0.54	0.68	0.73	0.98
i / N	0.2	0.4	0.6	0.8	1
(I / N) – R_i	0.09	-	-	0.07	0.02
R_i - (i-1)/N	0.11	0.34	0.28	0.13	0.18

From the above table we get,
$D^+ = \text{Max}(i / N – R_i) = 0.09$
$D^- = \text{Max}(R_i - (i-1)/N) = 0.34$
$D=\text{Max}(D^+, D^-) = 0.34$
The critical value $D_{0.05}$ is 0.565 (i.e. the critical value of D for $\alpha = 0.05$ and N=5 is 0.565).
Since $D<D_\alpha$, there is no chance to reject the hypothesis that the given random numbers are uniform.
(or, since the computed value is less than critical value, the given random numbers are uniform)

Chi-Squared Test

The Chi-Squared test is very important and useful statistical test to determine how often certain observed data fit the theoretical expected data.

$$\text{Chi-Square} = X_0^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

Where,
n= Number of class
O_i = Number (frequency) of times the observed data flalls in each class i for i=1,2,3……..n
E_i = Expected number of occurrence in each class i
For the uniform distribution, E_i, the expected number in each class is given by $E_i = N/n$ where N = total number of observation. The sampling distribution of X_0^2 is approximately the Chi- squared distribution with n-1 degree of freedom. Chi-square is a characteristics of distribution which is a measure of its randomness. The statistics chi-square is computed by subtracting the number of the random number in each

class from the expected number, squaring the difference, adding the squares and dividing the sum by the expectation.

Example:
If we have a sequence of 4000 (i.e. N=4000), 3 digits random number from 000 to 999. Then we can have n=10 classes in the range 00-99, 100-199, ………………………..900-999. Then the expected number of occurrence in each class is given by $E_i = N/n = 4000/10 = 400$. Now we have to measure how far the observed frequency deviates from the expected value i.e. if deviation from 400 is too much then we would suspect about non uniformity. So by how much deviation is accepted in sequence to be uniformly distributed is **given by Chi-square test.**
The Chi^2 statistics can be calculated as follows.

i	Class i	O_i (Number of observed occurrence) *(Assumed frequency)*	E_i (Number of expected occurrence)	(O_i - Ei)	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
1	00-99	425	400	25	625	1.5625
2	100-199	372	400	-28	784	1.96
3	200-299	395	400	-5	25	0.0625
4	300-399	415	400	15	225	0.5625
5	400-499	340	400	-60	3600	9
6	500-599	370	400	-30	900	2.25
7	600-699	410	400	10	100	0.25
8	700-799	382	400	-18	324	0.81
9	800-899	365	400	-35	1225	3.0625
10	900-999	394	400	-6	36	0.09

Total = 19.61

We know,

$$\text{Chi-Square} = X_0^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

$$= 19.61$$

Now to conclude how small or large a computed Chi-square value can we accept for verifying the uniformity, we have to use statistical table which gives critical value of Chi-square. We need degree of freedom(d_f) to read Chi-square table where $d_f = n-1$.
Here df=10-1=9.
Now use the Chi^2 table and find the row of entries for 9.

Percentage Points of the Chi-Square Distribution

Degrees of Freedom	Probability of a larger value of x ^{2}								
	0.99	0.95	0.90	0.75	0.50	0.25	0.10	0.05	0.01
9	2.088	3.325	4.168	5.899	8.343	11.39	14.68	16.92	21.67

Let level of significance **α =0.05**
So from the above Chi^2 table, at 95% confidence level the acceptable value for 9 degree of freedom is 16.92 which is less than 19.61 i.e. ($X^2_0 > X^2_{\alpha, n-1}$). hence the given set of random number is **not acceptable** so far as it's uniformity in distribution is concerned i.e. not uniformly distributed

Percentage Points of the Chi-Square Distribution

Degrees of Freedom	Probability of a larger value of x [2]								
	0.99	0.95	0.90	0.75	0.50	0.25	0.10	0.05	0.01
1	0.000	0.004	0.016	0.102	0.455	1.32	2.71	3.84	6.63
2	0.020	0.103	0.211	0.575	1.386	2.77	4.61	5.99	9.21
3	0.115	0.352	0.584	1.212	2.366	4.11	6.25	7.81	11.34
4	0.297	0.711	1.064	1.923	3.357	5.39	7.78	9.49	13.28
5	0.554	1.145	1.610	2.675	4.351	6.63	9.24	11.07	15.09
6	0.872	1.635	2.204	3.455	5.348	7.84	10.64	12.59	16.81
7	1.239	2.167	2.833	4.255	6.346	9.04	12.02	14.07	18.48
8	1.647	2.733	3.490	5.071	7.344	10.22	13.36	15.51	20.09
9	2.088	3.325	4.168	5.899	8.343	11.39	14.68	16.92	21.67
10	2.558	3.940	4.865	6.737	9.342	12.55	15.99	18.31	23.21
11	3.053	4.575	5.578	7.584	10.341	13.70	17.28	19.68	24.72
12	3.571	5.226	6.304	8.438	11.340	14.85	18.55	21.03	26.22
13	4.107	5.892	7.042	9.299	12.340	15.98	19.81	22.36	27.69
14	4.660	6.571	7.790	10.165	13.339	17.12	21.06	23.68	29.14
15	5.229	7.261	8.547	11.037	14.339	18.25	22.31	25.00	30.58
16	5.812	7.962	9.312	11.912	15.338	19.37	23.54	26.30	32.00
17	6.408	8.672	10.085	12.792	16.338	20.49	24.77	27.59	33.41
18	7.015	9.390	10.865	13.675	17.338	21.60	25.99	28.87	34.80
19	7.633	10.117	11.651	14.562	18.338	22.72	27.20	30.14	36.19
20	8.260	10.851	12.443	15.452	19.337	23.83	28.41	31.41	37.57
22	9.542	12.338	14.041	17.240	21.337	26.04	30.81	33.92	40.29
24	10.856	13.848	15.659	19.037	23.337	28.24	33.20	36.42	42.98
26	12.198	15.379	17.292	20.843	25.336	30.43	35.56	38.89	45.64
28	13.565	16.928	18.939	22.657	27.336	32.62	37.92	41.34	48.28
30	14.953	18.493	20.599	24.478	29.336	34.80	40.26	43.77	50.89
40	22.164	26.509	29.051	33.660	39.335	45.62	51.80	55.76	63.69
50	27.707	34.764	37.689	42.942	49.335	56.33	63.17	67.50	76.15
60	37.485	43.188	46.459	52.294	59.335	66.98	74.40	79.08	88.38

Consider the following 100 two-digits value in table below. Use the Chi-square test with $\alpha = 0.05$ to test whether the data shown below are uniformly distributed, where $X_{0.05, 9} = 16.9$.

0.34	0.90	0.25	0.89	0.87	0.44	0.12	.21	.46	.67
.83	.76	.79	.64	.70	.81	.94	.74	.22	.74
.96	.99	.77	.67	.56	.41	.52	.73	.99	0.02
.97	.30	.17	.82	.56	.05	.45	.31	.78	.05
.79	.71	.23	.19	.82	.93	.65	.37	.39	.42
.99	.17	.99	.46	.05	.66	.10	.42	.18	.49
.37	.51	.54	.01	.81	.28	.69	.34	.75	.49
.72	.43	.56	.97	.30	.94	.96	.58	.73	.05
.06	.39	.84	.24	.40	.64	.40	.19	.79	.62
.18	.26	.97	.88	.64	.47	.60	.11	.29	.78

Solution:
Defining the intervals of equal length as 0 - 0.1, 0.1 - 0.2,……………………0.9 – 1.0, we will get n=10 interval classes.
We have observed from the table that frequency of occurrence of 0.0 to 0.1 range is 8, 0.1 to 0.2 range is 8 and so on.

(Note: the interval is closed so in first interval the permitted values are from 0 to 0.1, similarly for second interval the permitted values are from 0.11 to 0.2 and so on)

Now Expected no of occurrences for each class i, $E_i = N/n = 100/10 = 10$

Class i	O_i	E_i	$O_i - E_i$	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
0 - 0.1	8	10	-2	4	0.4
0.1 – 0.2	8	10	-2	4	0.4
0.2 – 0.3	10	10	0	0	0
0.3 – 0.4	9	10	-1	1	0.1
0.4 – 0.5	12	10	2	4	.4
0.5 – 0.6	8	10	-2	4	.4
0.6 – 0.7	10	10	0	0	0
0.7 – 0.8	14	10	4	16	1.6
0.8 – 0.9	10	10	0	0	0
0.9 - 1	11	10	1	1	0.1

Total=3.4

So from the above table,

$$\text{Chi-Square} = X_0^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

$$= 3.4$$

Degree of freedom, $d_f = n-1 = 10-1 = 9$

The critical value of X_0^2 obtained from Chi^2 table for $\alpha = 0.05$ and 9 degree of freedom is 16.9. since $X_0^2 \leq X^2_{0.05,\,9}$ the null hypothesis of a uniform distribution is not rejected i.e. it can be accepted.

#Determine the Chi-square. Is it acceptable at 95% confidence level?

36, 91, 51, 2, 54, 6, 58, 6, 58, 2, 54, 1, 48, 97, 43, 22, 83, 25, 79, 95, 42, 57, 73, 17, 2, 42, 95, 38, 79, 29, 65, 9, 55, 97, 39, 83, 31, 77, 17, 62, 3, 49, 90, 37, 13, 17, 58, 11, 51, 92, 33, 78, 21, 66, 9, 54, 49, 90, 35, 84, 26, 74, 22, 62, 12, 90, 36, 83, 32, 75, 31, 94, 34, 87, 40, 7, 58, 5, 56, 22, 58, 77, 71, 10, 73, 23, 57, 13, 36, 89, 22, 68, 2, 44, 99, 27, 81, 26, 85,62

Note: use $\alpha = 0.05$
Interval =(0-10), (10-20),……….(90-100)

Independence Test

Sometime the random number generator passes the K-S test and Chi-square tests for uniformity, but the number generated my not be independent. Hence there is need of independence test. There are many methods to check the independence of generators. Two among them are

1. Auto correlation test
2. Run test

Auto Correlation test

The uniformity test of random number is only a necessary test for randomness but not sufficient one. A sequence of numbers may be perfectly uniform but still not random. For example, the sequence 1,2, 3,4,5,6,7,8,9,1,2,3,4………and so on would give a perfectly uniform distribution with Chi-square value as zero. But also the sequence can't be regarded as random because the number are not independent as the occurrence of one number decides the occurrence of next i.e. 3 determine 4. This defect is called serial auto correlation of adjacent pair of random numbers.

The chi square test for serial auto correlation uses a 10 * 10 matrix. The 10 classes described in the uniformity test are represented both along the rows and the column giving a large set of groups. Thus to reduce the number of groups, instead of 10, random numbers are divided into smaller number of classes like three or four. Example, three classes well be

i. Less than or equal to 0.33
ii. Less than or equal to 0.66
iii. Less than or equal to 1.0

With three classes in a row and three classes in column there will be 9 groups.

Let us consider the following random numbers
49, 95, 82, 19, 41, 31, 12, 53, 62, 40, 87, 83, 26, 01, 91, 55, 38, 75, 90, 35, 71, 57, 27, 85, 52, 08, 35, 57, 88, 38, 77, 86, 29, 18, 09, 96, 58, 22, 08, 93, 85, 45, 79, 68, 20, 11, 78, 93, 21, 13, 06, 32, 63, 79, 54, 67, 35, 18, 81, 14, 62, 13, 76, 74, 76, 45, 29, 36, 80, 78, 95, 25, 52
Determine whether the hypothesis of independence can be rejected, where $\alpha = 0.05$
Solution,
these N = 73 random numbers, giving (N-1)=72 pairs, are grouped in 9 classes with expectation $E_i = N/n = 72/9 \sim 8$

Class	O_i	E_i	$O_i - E_i$	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
R1≤ 0.33 & R2≤ 0.33					
R1≤ 0.67 & R2≤ 0.33					
R1≤ 1.0 & R2≤ 0.33					
R1≤ 0.33 & R2≤ 0.67					
R1≤ 0.67 & R2≤ 0.67					
R1≤ 1.0 & R2≤ 0.67					
R1≤ 0.33 & R2≤ 1.0					
R1≤ 0.67 & R2≤ 1.0					
R1≤ 1.0 & R2≤ 1.0					

Total=?

From the above table,

$$\text{Chi-Square} = X_0^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

$$= ?$$

The observation O_i in different class is determined by taking the pair of random numbers. Pair 0.49 and 0.95 falls in class R1≤0.67 & R2≤1.0, similarly next pair 0.95 and o.82 falls in range R1≤1.0 &R≤1.0 and so on. Since the total number of pair is 72.
Since there are two random variables, the degree of freedom $d_f = n-2 = 9-2 = 7$. The critical value of Chi square for 7 degree of freedom at α=0.05 level of significance is 14.07.
since $X_0^2 \leq X^2_{0.05,\,7}$ the null hypothesis of independence is not rejected i.e. it can be accepted.

Run Test

Tests the run up and down or the runs above and below the mean by comparing the actual value to expected value. The run test examines the arrangements of numbers in a sequence to test the hypothesis of independence.
A run is defined as a succession of similar events proceeded and followed by a different event.
Example: in a sequence of tosses of a coin, we have H, T, T, H, H, T, T, T, H, T. we get six runs as 1, 2, 2, 3, 1, 1
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 number of too few runs, it is not likely a random sequence. Example, 0.08, 0.18, 0.23, 0.36, 0.42, …., the sequence has only

one run, i.e. an up run, so not likely a random sequence. Also, if a sequence has too many runs, it is not likely to be random. Example, 0.08, 0.93, 0.15, 0.96, 0.26, 0.84, 0.28, 0.79, 0.36, 0.57. it has total 9 runs, five up and four down, so it is not likely a random sequence.

Example:

Based on the runs up and runs down, determine whether the following sequence of 40 numbers is such that the hypothesis of independence can be rejected where $\alpha = 0.05$

.41	.68	.89	.94	.74	.91	.55	.62	.36	.27
.19	.72	.76	.08	.54	.02	.01	.36	.16	.28
.18	.01	.95	.69	.18	.47	.23	.32	.82	.53
.31	.42	.73	.04	.83	.45	.13	.57	.63	.29

Solution,

Assigning a "+" or "-" in the number based on whether they are followed by a larger or smaller number. The last number is followed by no events so it will neither get a "+ "nor a "-". The sequence of runs up and runs down is below.

+ + + - + - + - - - + + - + - - + - + - - + - - + - + + - - + + - + - - + + -

Total number of runs in this sequence, a=26

For N=40, the mean an variance is given by

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

$$= \frac{2 * 40 - 1}{3}$$

$$= 26.33$$

And

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

$$= \frac{16 * 40 - 29}{90}$$

$$= 6.79$$

Then, converting it to a standard normal distribution

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

$$= \frac{26 - 26.33}{\sqrt{6.79}}$$

$$= -0.13$$

Now, the critical value , $Z_{\alpha/2} = Z_{0.025} = 1.96$ (from below z score table). since $Z_0 >= -Z_{\alpha/2}$, and $Z_0 <= Z_{\alpha/2}$, so the independence of the numbers can't be rejected on the basis of this test. i.e. null hypothesis of independence can't be rejected.

Z-Value	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.50000	0.49601	0.49202	0.48803	0.48405	0.48006	0.47608	0.47210	0.46812	0.46414
0.1	0.46017	0.45620	0.45224	0.44828	0.44433	0.44038	0.43644	0.43251	0.42858	0.42465
0.2	0.42074	0.41683	0.41294	0.40905	0.40517	0.40129	0.39743	0.39358	0.38974	0.38591
0.3	0.38209	0.37828	0.37448	0.37070	0.36693	0.36317	0.35942	0.35569	0.35197	0.34827
0.4	0.34458	0.34090	0.33724	0.33360	0.32997	0.32636	0.32276	0.31918	0.31561	0.31207
0.5	0.30854	0.30503	0.30153	0.29806	0.29460	0.29116	0.28774	0.28434	0.28096	0.27760
0.6	0.27425	0.27093	0.26763	0.26435	0.26109	0.25785	0.25463	0.25143	0.24825	0.24510
0.7	0.24196	0.23885	0.23576	0.23270	0.22965	0.22663	0.22363	0.22065	0.21770	0.21476
0.8	0.21186	0.20897	0.20611	0.20327	0.20045	0.19766	0.19489	0.19215	0.18943	0.18673
0.9	0.18406	0.18141	0.17879	0.17619	0.17361	0.17106	0.16853	0.16602	0.16354	0.16109
1.0	0.15866	0.15625	0.15386	0.15151	0.14917	0.14686	0.14457	0.14231	0.14007	0.13786
1.1	0.13567	0.13350	0.13136	0.12924	0.12714	0.12507	0.12302	0.12100	0.11900	0.11702
1.2	0.11507	0.11314	0.11123	0.10935	0.10749	0.10565	0.10383	0.10204	0.10027	0.09853
1.3	0.09680	0.09510	0.09342	0.09176	0.09012	0.08851	0.08691	0.08534	0.08379	0.08226
1.4	0.08076	0.07927	0.07780	0.07636	0.07493	0.07353	0.07215	0.07078	0.06944	0.06811
1.5	0.06681	0.06552	0.06426	0.06301	0.06178	0.06057	0.05938	0.05821	0.05705	0.05592
1.6	0.05480	0.05370	0.05262	0.05155	0.05050	0.04947	0.04846	0.04746	0.04648	0.04551
1.7	0.04457	0.04363	0.04272	0.04182	0.04093	0.04006	0.03920	0.03836	0.03754	0.03673
1.8	0.03593	0.03515	0.03438	0.03362	0.03288	0.03216	0.03144	0.03074	0.03005	0.02938
1.9	0.02872	0.02807	0.02743	0.02680	0.02619	0.02559	0.02500	0.02442	0.02385	0.02330
2.0	0.02275	0.02222	0.02169	0.02118	0.02068	0.02018	0.01970	0.01923	0.01876	0.01831
2.1	0.01786	0.01743	0.01700	0.01659	0.01618	0.01578	0.01539	0.01500	0.01463	0.01426
2.2	0.01390	0.01355	0.01321	0.01287	0.01255	0.01222	0.01191	0.01160	0.01130	0.01101
2.3	0.01072	0.01044	0.01017	0.00990	0.00964	0.00939	0.00914	0.00889	0.00866	0.00842
2.4	0.00820	0.00798	0.00776	0.00755	0.00734	0.00714	0.00695	0.00676	0.00657	0.00639
2.5	0.00621	0.00604	0.00587	0.00570	0.00554	0.00539	0.00523	0.00508	0.00494	0.00480
2.6	0.00466	0.00453	0.00440	0.00427	0.00415	0.00402	0.00391	0.00379	0.00368	0.00357

Fig: Z value

#Consider the 50 two-digit values.

0.34	0.9	0.25	.89	.87
.83	.76	.79	.64	.7
.96	.99	.77	.67	.56
.47	.3	.17	.82	.56
.79	.71	.23	.19	.82
.99	.17	.99	.46	.05
.37	.51	.54	.01	.81
.72	.43	.56	.97	.3
.06	.39	.84	.24	.4
.18	.26	.97	.88	.64

Based upon run up and down, determine whether the hypothesis of independence can be rejected, where $\alpha = 0.05$