Chapter 6
Random Number

Random Numbers

- A number chosen from some specified distribution randomly.
- Random numbers are samples drawn from a uniformly distributed random variable between some satisfied intervals, they have equal probability of occurrence.
- A number chosen from some specified distribution randomly such that selection of large set of these numbers reproduces the underlying distribution is called random number.
- Every number is equally likely to occur and there is no pattern, and thus no way of predicting what number will be next in sequence.
- Most simulations are random number driven.

General Properties of Random Number

1. Uniformity
   - The random numbers generated should be uniform. That means a sequence of random numbers should be equally probable every where.
   - If we divide all the set of random numbers into several numbers of class interval then number of samples in each class should be same.
   - If 'N' number of random numbers are divided into 'K' class interval, then expected number of samples in each class should be equal to ei = N / K.
2. Independent
   - Each random number R_t is an independent sample drawn from a continuous uniform distribution between 0 and 1.
   - The probability density function(pdf) is given by:

$$\text{pdf : } f(x) = \begin{Bmatrix} 1\ for\ 0 \leq x \leq 1 \\ 0\ otherwise \end{Bmatrix}$$

3. Maximum Density: The large samples of random number should be generated in a given range.
4. Maximum Cycle: It states that the repetition of numbers should be allowed only after a large interval of time.

Pseudo Random Numbers

- Here pseudo means false.
- Pseudo implies that the random numbers are generated by using some known arithmetic operation.
- Since, the arithmetic operation is known and the sequence of random numbers can be repeatedly obtained, the numbers cannot be called truly random.
- However, the pseudo random numbers generated by many computer routines, very closely fulfill the requirement of desired randomness.

Pseudo Random Numbers

- If the method for generating random number or the random number generator is defective then generated pseudo random numbers may have following departures from idle randomness:

1. The generated numbers may not be uniformly distributed
2. The generated numbers may not be continuous
3. The mean of the generated numbers may be too high or too low
4. The variance may be too high or too low.

5. There may be cyclic patterns in the generated There may be cyclic patterns in the generated numbers, like

a) Auto correction between numbers

b) a group of numbers continuously above the mean, followed by group continuously below of mean.

Thus, before employing a pseudo random number generator, it should be properly validated, by testing the generated random numbers for randomness.

Generation of random number

- In computer simulation, where a very large number of random numbers is generally required, the random numbers can be obtained by the following methods:

1. Random numbers may be drawn from the random number tables stored in the memory of the computer.
2. Using electronics devices-Very expensive
3. Using arithmetic operation

Requirements of a good pseudo random generator

1. The sequence of generated random numbers must follow uniform distribution.
2. The sequence of random numbers generated must be statistically independent.
3. The sequence must be non-repeating for any desired length. Although theoretically not possible, a long repeatability cycle is adequate for practical purposes.
4. Generation of random numbers must be fast because in simulation studies, a large number of random numbers are required. A slow generator will greatly increase the time and cost fro simulation studies.
5. The generator must require less computer memory as will as computational resources.

Algorithm for generating Random Numbers

1. Linear Congruential Method

- A sequence of integers X1, X2, X3, ……….. are produced between zero and m-1 by using the recursive relation as follows:

$$X(i+1) = (a\,X(i) + c) \bmod m, \text{ for } i = 0, 1, 2, 3, 4, \ldots\ldots\ldots\ldots$$

- The initial random integer X(0) is known as seed, a is called multiplier, c is increment and m is the modulus.

a. If a = 1 in above expression, the expression reduces to additive congruential method

i.e. $X(i+1) = (X(i) + c) \bmod m$

b. If c = 0 in above equation, the expression reduces to multiplicative congruential method,

i.e. $X(i+1) = aX(i) \bmod m$

c. If $a>1$ and $c>0$ in above expression, then it represents mixed type congruential method. For this type we use

$$X(i+1) = (a\,X(i) + c) \bmod m, \text{ for } i = 0, 1, 2, 3, 4, \ldots\ldots\ldots.$$

2. Combined Linear Congruential Method: Combined linear congruential method uses the combination of two or more multiplicative congruential generators so as to provide good statistical properties and a longer period.

Note

- If question asks you to generate random numbers using Linear Congruential Method and provides you with multiplier, increment, modulus and seed values then always use the original formula.

  i.e. $X(i+1) = (a\,X(i) + c) \bmod m, \text{ for } i = 0, 1, 2, 3, 4, \ldots\ldots\ldots.$

Numerical

Let multiplier = 13, increment = 1 and modulus value = 19. Use congruential method to generate random numbers taking seed value = 1.

sol^n: Given a = 13, c = 1, m = 19 and X(0) = 1

We have, X(i+1) = (aX(i) + c) mod m

For i = 0, X(1) = (aX(0) + c) mod m

= (13 * 1 + 1) mod 19

= 14 mod 19 = 14

For i = 1, X(2) = (aX(1) + c) mod 19

= (13*14+1) mod 19

= 12

And so on.

Condition to stop Iteration

1. If question provides condition, do accordingly.
2. If condition not provided:

   a. Stop if same number repeats

   b. Else go and find all random numbers

Numerical

Use Linear Congruential Method to generate a sequence of three two digit random integers.

Given seed value = 32, multiplier = 8, increment = 47, modulus value = 100.

Numerical

Use Linear Congruential Method to generate a sequence of three two digit random integers.

Given seed value = 32, multiplier = 8, increment = 47, modulus value = 100.

Sol^n Given $X(0) = 32$, $a = 8$, $c = 47$, $m = 100$

$X(1) = (8*32+47) \bmod 100 = 3$ (Not OK)

$X(2) = (8*3+47) \bmod 100 = 71$ (OK)

$X(3) = (8*71+47) \bmod 100 = 15$ (OK)

$X(4) = (8*15+47) \bmod 100 = 67$ (OK)

Numerical

Use Multiplicative Congruential Method to generate a sequence of four three digit random integers.

Given seed value = 117, multiplier = 8, increment = 47, modulus value = 1000.

Test For Random Numbers

1. **Frequency test**: Uses the Kolmogorov-Smirnov(KS) or the chi-square test to compare the distribution of the set of numbers generated to a uniform distribution.
2. **Runs test**: Tests the runs up and down or the runs above and below the mean by comparing the actual values to expected values. The statistic for comparison is the chi-square.
3. **Autocorrelation test**: Tests the correlation between numbers and compares the sample correlation to the expected correlation of zero.
4. **Gap test**. Counts the number of digits that appear between repetitions of a particular digit and then uses the Kolmogorov-Smirnov(KS) test to compare with the expected number of gaps.
5. **Poker test**. Treats numbers grouped together as a poker hand. Then the hands obtained are compared to what is expected using the chi-square test.

Kolmogorov-Smirnov(KS) Test

- It is a test for random number developed by A.N. Kolmogorov and N.V. Smirnov.
- It is used to test the uniformity of random numbers i.e. whether random numbers are uniformly generated or not.
- This test is designed for continuous distributions where the Observed Cumulative Distribution Function(CDF) is compared with empirical CDF.

KS Test Algorithm

1. Define the hypothesis.
2. For provided data, rank data from smallest to greatest. Let $R(i)$ denote the i^{th} smallest observation among the N observations.

   i.e. $R(1) \le R(2) \le R(3) \le \ldots \le R(N)$

3. Compute D^+ and D^- as

$$D^+ = \max \left\{ \frac{i}{N} - R(i) \right\}$$

$$D^- = \max \left\{ R(i) - \frac{(i-1)}{N} \right\}$$

for $1 \le i \le N$

KS Test Algorithm

4. Compute $D = \max\{D^+, D^-\}$

5. Determine the critical value D_α from the table for specific value of α and sample size N.

6. If $D > D_\alpha$, the null hypothesis is rejected. This means the random numbers are $\boxed{\text{not uniform.}}$

If $D \leq D_\alpha$, the null hypothesis is not rejected. This means the random numbers $\boxed{\text{are uniform}}$

n	α 0.01	α 0.05	α 0.1	α 0.15	α 0.2
1	0.995	0.975	0.950	0.925	0.900
2	0.929	0.842	0.776	0.726	0.684
3	0.828	0.708	0.642	0.597	0.565
4	0.733	0.624	0.564	0.525	0.494
5	0.669	0.565	0.510	0.474	0.446
6	0.618	0.521	0.470	0.436	0.410
7	0.577	0.486	0.438	0.405	0.381
8	0.543	0.457	0.411	0.381	0.358
9	0.514	0.432	0.388	0.360	0.339
10	0.490	0.410	0.368	0.342	0.322
11	0.468	0.391	0.352	0.326	0.307
12	0.450	0.375	0.338	0.313	0.295
13	0.433	0.361	0.325	0.302	0.284
14	0.418	0.349	0.314	0.292	0.274
15	0.404	0.338	0.304	0.283	0.266
16	0.392	0.328	0.295	0.274	0.258
17	0.381	0.318	0.286	0.266	0.250
18	0.371	0.309	0.278	0.259	0.244
19	0.363	0.301	0.272	0.252	0.237
20	0.356	0.294	0.264	0.246	0.231
25	0.320	0.270	0.240	0.220	0.210
30	0.290	0.240	0.220	0.200	0.190
35	0.270	0.230	0.210	0.190	0.180
40	0.250	0.210	0.190	0.180	0.170
45	0.240	0.200	0.180	0.170	0.160
50	0.230	0.190	0.170	0.160	0.150
OVER 50	$\frac{1.63}{\sqrt{n}}$	$\frac{1.36}{\sqrt{n}}$	$\frac{1.22}{\sqrt{n}}$	$\frac{1.14}{\sqrt{n}}$	$\frac{1.07}{\sqrt{n}}$

Numerical –KS Test

Perform uniformity test using KS test with a level of significance $\alpha = 0.05$ on the following five generated numbers.

0.44, 0.81, 0.14, 0.05, 0.93

Soln:

Let H_0 represent the null hypothesis.

H_0: the generated numbers are uniform

Arranging given data in ascending order

i	1	2	3	4	5
R(i)	0.05	0.14	0.44	0.81	0.93

Calculating D^+ and D^-

(i)	$R(i)$	$\frac{i}{N}$	$\frac{(i-1)}{N}$	$\frac{i}{N} - R(i)$	$R(i) - \frac{(i-1)}{N}$
1	0.05	0.2	0	0.15	0.05
2	0.14	0.4	0.2	0.26	-0.06
3	0.44	0.6	0.4	0.16	0.04
4	0.81	0.8	0.6	-0.01	0.21
5	0.93	1	0.8	0.07	0.13

Note: While determining D^+ and D^-, no need to ~~calculate~~ consider negative values.

So, we have $D^+ = \max\left\{\frac{i}{N} - R_{(i)}\right\}$

$= 0.26$

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

$= 0.21$

Calculation of D

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

$= 0.26$

The critical value D_α for $\alpha = 0.05$ and $N = 5$ from table is 0.565

Since $D < D_\alpha$, the null hypothesis is not rejected. So we can conclude that the random numbers are uniform.

Numerical

K-S test is to be performed to test the uniformity of following random numbers with a level of significance of $\alpha = 0.05$.

0.24, 0.89, 0.11, 0.61, 0.23, 0.86, 0.41, 0.64, 0.50, 0.65

i	R(i)	i/n	(i-1)/n	(i/n)-R(i)	R(i)-((i-1)/n)
1	0.11	0.1	0	-0.01	0.11
2	0.23	0.2	0.1	-0.03	0.13
3	0.24	0.3	0.2	0.06	0.04
4	0.41	0.4	0.3	-0.01	0.11
5	0.50	0.5	0.4	0	0.10
6	0.61	0.6	0.5	-0.01	0.11
7	0.64	0.7	0.6	0.06	0.04
8	0.65	0.8	0.7	0.15	-0.05
9	0.86	0.9	0.8	0.04	0.06
10	0.89	1	0.9	0.11	-0.01

Chi-Square Test

- It is a type of frequency test
- It is a test used to check the randomness of a distribution
- This statistical test is used to determine how often certain observed data fit the theoretically expected data.
- This method compares the observed frequency with the theoretical. So it determines how often certain observed data fit the theoretically expected data.

Chi-Square Test

- The Chi-Square test uses the following sample statistics
- The total N observations are divided into n number of classes.

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

where n = number of classes

O_i = Observed number of occurrence(frequency) in each class

E_i = Expected number of occurrence(frequency) in each class

N = total number of observations

Chi-Square Test

Note:

1. For chi square test, degree of freedom = n-1
   - where n represents number of classes
2. Chi Square test is usually recommended when $E_i \geq 5$

   i.e number of expected occurrence/frequency in each class ≥ 5

Chi-Square Test - Procedure

1. Divide the sample range into n intervals. Note: 3 values in each interval at least.
2. Determine:
   a. Expected number of values E_i in each interval i under the null hypothesis.
   b. Observed number of values O_i in each interval.

3. Compute the statistic χ^2 value.
4. Find the standard value of χ^2 based on given confidence level/ level of significance(α) and degree of freedom. For simple chi-square test

   degree of freedom = n - 1

5. Draw conclusion accordingly. If $\chi^2_{calculated} < \chi^2_{standard}$ the null hypothesis is not rejected. Else the null hypothesis is rejected.

Chi-Square Test (Example)

The two Digit random numbers generated by a multiplicative congruential method are given below. Determine Chi-Square. Is it acceptable at 95% confidence level?

36, 91, 51, 02, 54, 06, 58, 06,58,02, 54, 01, 48, 97, 43, 22, 83, 25, 79, 95, 42, 87, 73, 17, 02, 42, 95, 38, 79, 29, 65, 09, 55, 97, 39, 83, 31, 77,17, 62, 03, 49, 90, 37, 13, 17, 58, 11, 51, 92, 33, 78, 21, 66, 09, 54, 49, 90, 35, 84, 26, 74, 22, 62, 12, 90,36, 83, 32, 75, 31, 94, 34, 87, 40, 07, 58, 05, 56,22, 58,77, 71, 10, 73,23,57,13,36,89,22,68,02,44,99,27,81,26,85, 22

sol^n : Let H_0 represents null hypothesis where H_0: the numbers are acceptable for given confidence level

Here, Total number of samples (N) = 100

Let us divide these data into 10 classes i.e. n =10

E_i=N/n=100/10=10

`

Classes	Observed Frequency(Oi)	$(O_i - E_i)$	$(O_i - E_i)^2$	
0 < r ≤ 10	13	3	9	0.9
10 < r ≤ 20	7	-3	9	0.9
20 < r ≤ 30	12	2	4	0.4
30 < r ≤ 40	13	3	9	0.9
40 < r ≤ 50	7	-3	9	0.9
50 < r ≤ 60	13	3	9	0.9
60 < r ≤ 70	5	-5	25	2.5
70 < r ≤ 80	10	0	0	0
80 < r ≤ 90	12	2	4	0.4
90 < r ≤ 100	8	-2	4	0.4

So $\chi^2 = 8.2$

i.e. $\chi^2_{calculated} = 8.2$

Given, confidence level = 95% = 0.95

So level of significance(α) = 1 – 0.95 = 0.05

Similarly degree of freedom in this case = n – 1

= 10 - 1 = 9

Now we use Chi-Square value from table for α = 0.05 and degree of freedom =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
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

From Chi-Square Table, $\chi^2_{standard} = 16.92$

Since $\chi^2_{calculated} < \chi^2_{standard}$ the null hypothesis is not rejected. So we conclude that the numbers are acceptable for 95% confidence level.

K-S test vs Chi-Square Test

K-S test	Chi-Square Test
Done for smaller samples.	Done for larger samples.
Difference between observed and expected CDFs(Cumulative Distribution Function)	Difference between observed and expected PDFs(Probability Density Function)
Uses each observed sample without grouping	Group observations

Numerical – 2074 Bhadra

7. What are the properties of Random number? Using Chi-Square test the uniformity at 90% for the given random numbers. Degree of freedom for 6 = 10.645, 7 = 12.017, 8 = 13.362 9 = 14.684, 10 = 15.987.

20	34	43	42	14	10	33	17	6	11
15	16	4	1	35	22	9	46	37	57
51	49	40	27	59	5	44	19	41	55
53	29	3	31	48	8	56	28	12	7

Classes	Observed Frequency(Oi)	$(O_i - E_i)$	$(O_i - E_i)^2$	
0 < r ≤ 6	5	1	1	0.25
6 < r ≤12	6	2	4	1
12 < r ≤ 18	4	0	0	0
18 < r ≤ 24	3	-1	1	0.25
24 < r ≤ 30	3	-1	1	0.25
30 < r ≤ 36	4	0	0	0
36 < r ≤ 42	4	0	0	0
42 < r ≤ 48	4	0	0	0
48 < r ≤ 54	4	0	0	0
54 < r ≤ 60	3	-1	1	0.25
				∑=2

Note: E_i=N/n

- Here degree of freedom = n-1 = 10-1=9

Here $\chi^2_{standard}$= 14.684

Since $\chi^2_{calculated} < \chi^2_{standard}$ the null hypothesis is not rejected.

Gap Test

- The gap test is used to determine the significance of the interval between recurrence of the same digit.
- The Gap Test measures the number of digits between successive occurrences of the same digit.
- A gap of length x occurs between the recurrence of some digit.
- If we are only concerned with digits between 0 and 9 then,

$$P(\text{gap of } n) = 0.9^n * 0.1$$

- The theoretical frequency distribution of randomly ordered digits is given by

$$P(\text{gap} \leq x) = F(x) = 0.1\sum_{n=0}^{x} 0.9^n$$
$$= 1 - 0.9^{x+1}$$

Algorithm for Gap Test

Step 1

Specify the Cumulative Distribution Function(CDF) for the theoretical frequency distribution given by

$$P(\text{gap} \leq x) = F(x) = 0.1\sum_{n=0}^{x} 0.9^{n}$$
$$= 1 - 0.9^{x+1}$$

where x = maximum gap length based on the selected class interval width

Step 2

Arrange the observed sample of gaps in a cumulative distribution with these same classes.

Algorithm for Gap Test

- Step 3

Find D, the maximum deviation between F(x) and $S_N(x)$ as

$$D = \max |F(x) - S_N(x)|$$

Where $S_N(X)$ is the observed frequency distribution.

$$S_N(x) = \frac{Number\ of\ gaps \leq x}{Total\ number\ of\ gaps}$$

<u>This value is equal to the cumulative relative frequency of each gap class</u>

Step 4

Determine the critical value D_{α}, from Table(K-S critical value) for the specified value of α and the sample size N.

Step 5

If the calculated value of D is greater than the tabulated value of D_{α} , the null hypothesis of independence is rejected.

Numerical- Gap Test

Based on the frequency with which gaps occur, analyze following 110 digits to test whether they are independent. Use $\alpha = 0.05$

4	1	3	5	1	7	2	8	2	0	7	9	1	3	5	2	7	9	4	1	6	3	3	9	6
3	4	8	2	3	1	9	4	4	6	8	4	**1**	3	8	9	5	5	7	3	9	5	9	8	5
3	2	2	3	7	4	7	0	3	6	3	5	9	9	5	5	5	0	4	6	8	0	4	7	0
3	3	0	9	5	7	9	5	**1**	6	6	3	8	8	8	9	2	9	1	8	5	4	4	5	0
2	3	9	7	1	2	0	3	6	3															

sol^n : Let H_0 represents null hypothesis.

H_0 : The numbers are independent

Here digits are from 0 to 9. So total number of distinct digits = 10

So number of gaps(N) = Number of data values – Number of distinct digits

= 110 -10 = 100

Numerical- Gap Test

Gap Length	Frequency	Relative Frequency	Cumulative Relative frequency	$F(x) = 1 - 0.9^{x+1}$	$\lvert F(x) - S_N(x) \rvert$
0 – 3	35	0.35	0.35	0.3439	0.0061
4 – 7	22	0.22	0.57	0.5695	0.0005
8 – 11	17	0.17	0.74	0.7176	0.0224
12 – 15	9	0.09	0.83	0.8147	0.0153
16 – 19	5	0.05	0.88	0.8784	0.0016
20 – 23	6	0.06	0.94	0.9202	0.0198
24 – 27	3	0.03	0.97	0.9497	0.0223
28 – 31	0	0	0.97	0.9657	0.0043
32 – 35	0	0	0.97	0.9775	0.0075
36 – 39	2	0.02	0.99	0.9852	0.0043
40 – 43	0	0	0.99	0.9903	0.0003
44 – 47	1	0.01	1	0.9936	0.0064

n	α 0.01	α 0.05	α 0.1	α 0.15	α 0.2
1	0.995	0.975	0.950	0.925	0.900
2	0.929	0.842	0.776	0.726	0.684
3	0.828	0.708	0.642	0.597	0.565
4	0.733	0.624	0.564	0.525	0.494
5	0.669	0.565	0.510	0.474	0.446
6	0.618	0.521	0.470	0.436	0.410
7	0.577	0.486	0.438	0.405	0.381
8	0.543	0.457	0.411	0.381	0.358
9	0.514	0.432	0.388	0.360	0.339
10	0.490	0.410	0.368	0.342	0.322
11	0.468	0.391	0.352	0.326	0.307
12	0.450	0.375	0.338	0.313	0.295
13	0.433	0.361	0.325	0.302	0.284
14	0.418	0.349	0.314	0.292	0.274
15	0.404	0.338	0.304	0.283	0.266
16	0.392	0.328	0.295	0.274	0.258
17	0.381	0.318	0.286	0.266	0.250
18	0.371	0.309	0.278	0.259	0.244
19	0.363	0.301	0.272	0.252	0.237
20	0.356	0.294	0.264	0.246	0.231
25	0.320	0.270	0.240	0.220	0.210
30	0.290	0.240	0.220	0.200	0.190
35	0.270	0.230	0.210	0.190	0.180
40	0.250	0.210	0.190	0.180	0.170
45	0.240	0.200	0.180	0.170	0.160
50	0.230	0.190	0.170	0.160	0.150
OVER 50	$\frac{1.63}{\sqrt{n}}$	$\frac{1.36}{\sqrt{n}}$	$\frac{1.22}{\sqrt{n}}$	$\frac{1.14}{\sqrt{n}}$	$\frac{1.07}{\sqrt{n}}$

Numerical- Gap Test

So the calculated value of D is

$$D = \max |\ F(x) - S_N(x)\ |$$
$$= 0.0224$$

The critical value of D for $\alpha = 0.05$ is $D_\alpha = \frac{1.36}{\sqrt{N}} = \frac{1.36}{\sqrt{100}}$

$$= 0.136$$

Since $D < D_\alpha$, the null the null hypothesis is not rejected. So the numbers are independent.

Gap Test Example For Exam

Explain the algorithm for gap test with an example.

Let us assume 110 random numbers between 0 to 9 with varying gap length. Let the maximum gap length be 34.

Let H_0 represents null hypothesis.

H_0 : The numbers are independent

Here digits are from 0 to 9. So total number of distinct digits = 10

So number of gaps(N) = Number of data values – Number of distinct digits

= 110 -10 = 100

Numerical- Gap Test

Gap Length	Frequency	Relative Frequency	Cumulative Relative frequency	$F(x) = 1 - 0.9^{x+1}$	$\| F(x) - S_N(x) \|$
0 – 5	45	0.45	0.45	0.4685	0.0185
6 – 11	15	0.15	0.6	0.7175	0.1175
12 – 17	12	0.12	0.72	0.8499	0.1299
18 – 23	8	0.08	0.8	0.920	0.12
24 – 29	13	0.13	0.93	0.9576	0.0276
30 – 35	7	0.07	1	0.9774	0.0226

Numerical- Gap Test

- So the calculated value of D is

$$D = \max| F(x) - S_N(x) |$$
$$= 0.1299$$

The critical value of D for $\alpha = 0.05$ is $D_\alpha = \frac{1.36}{\sqrt{N}} = \frac{1.36}{\sqrt{100}}$

$$= 0.136$$

Since $D < D_\alpha$, the null the null hypothesis is not rejected. So the numbers are independent.

Poker Test

- This test gets its name from a game of cards called poker.
- Poker test for independence is based on the frequency with which certain digits are repeated.
- Poker test not only tests for randomness of the sequence of numbers, but also the digits comprising of each number.
- Poker test treats numbers grouped together as a poker hand. Then the hands obtained are compared to what is expected using the chi-square test.

Poker test for 3 digit random number

 Possibilities for 3 digit number:

a. Three different digits
b. Three like digits
c. Exactly one pair

Calculating Probabilities of each possibility

1. Probability(Three different Digits) $= \frac{10}{10} * \frac{9}{10} * \frac{8}{10}$

$= 0.72$

2. Probability(Three like digits) $= \frac{10}{10} * \frac{1}{10} * \frac{1}{10}$

$= 0.01$

3. Probability(Exactly One Pair) $= 3C2 * \frac{10}{10} * \frac{1}{10} * \frac{9}{10}$

$= \frac{3!}{(3-2)! * 2!} * \frac{10}{10} * \frac{1}{10} * \frac{9}{10}$

$= 0.27$ ‘

OR last probability value can be calculated by subtracting other probabilities from 1

$= 1 - 0.72 - 0.01$

$= 0.27$

Numerical

A sequence of 1000 three-digit numbers has been generated and an analysis indicates that 680 have three different digits, 289 contain exactly one pair of like digits, and 31 contain three like digits. Based on the poker test, are these numbers independent? Take $\alpha = 0.05$.

soln: Let H_0 be null hypothesis.

H_0 : The numbers are independent

Total number of three-digit numbers(N) = 1000

Numerical

- Here,

1. Probability(Three different Digits) $= \frac{10}{10} * \frac{9}{10} * \frac{8}{10}$
   $= 0.72$

2. Probability(Three like digits) $= \frac{10}{10} * \frac{1}{10} * \frac{1}{10}$
   $= 0.01$

3. Probability(Exactly One Pair) $= 3C2 * \frac{10}{10} * \frac{1}{10} * \frac{9}{10}$
   $= \frac{3!}{(3-2)! * 2!} * \frac{10}{10} * \frac{1}{10} * \frac{9}{10}$
   $= 0.27$

Combination (i)	Expected Frequency (E_i) = Probability(i)*N	Observed Frequency (O_i)	
Three Different Digits	0.72*1000 = 720	680	2.22
Three like digits	0.01*1000 = 10	31	44.10
Exactly one pair	0.27*1000 = 270	289	1.33
	1000	1000	

-

$So\ \chi^2_{calculated} = 47.65$

Here degree of freedom = n-1 = 3-1 = 2

So $\chi^2_{standard} = 5.99$

Since $\chi^2_{calculated} > \chi^2_{standard}$ the null hypothesis is rejected. So given numbers are not independent.

Numerical

A sequence of 1000 three-digit numbers has been generated and an analysis indicates that 695 have three different digits, 293 contain exactly one pair of like digits, and 12 contain three like digits. Based on the poker test, are these numbers independent? Take $\alpha = 0.05$.

Poker test for 4 digit random number

- Possibilities for 4 digit number:

a. Four different digits
b. Exactly one pair
c. Two pairs
d. Three of a kind
e. All four like digits

Calculating Probabilities of each possibility

1. Probability(Four different Digits) = $\frac{10}{10} * \frac{9}{10} * \frac{8}{10} * \frac{7}{10}$
$= 0.504$

2. Probability(Exactly one pair) =4C2* $\frac{10}{10} * \frac{1}{10} * \frac{9}{10} * \frac{8}{10}$
$= 0.432$

3. Probability(Two Pairs) = 4C2 * $\frac{10}{10} * \frac{1}{10} * \frac{2C2}{2!} * \frac{9}{10} * \frac{1}{10}$
$= 0.027$

4. Probability(Three of a kind) = 4C3 * $\frac{10}{10} * \frac{1}{10} * \frac{1}{10} * \frac{9}{10}$
$= 0.036$

Calculating Probabilities of each possibility

5. Probability(All four like digits) $= \frac{10}{10} * \frac{1}{10} * \frac{1}{10} * \frac{1}{10}$
$= 0.001$

Numerical

A sequence of 1000 four-digit numbers has been generated and an analysis indicates:

Combinations	Observed Frequency
Four Different Digits	540
One pair	320
Two pairs	70
Three like digits	50
Four like digits	20
	1000

Based on poker test, test these numbers are independent for $\alpha = 0.05$

soln: Let H_0 be null hypothesis.

H_0 : The numbers are independent

Total number of four-digit numbers(N) = 1000

Here,

Probability(Four different Digits) = $\frac{10}{10} * \frac{9}{10} * \frac{8}{10} * \frac{7}{10}$

= 0.504

Probability(Exactly one pair) =4C2* $\frac{10}{10} * \frac{1}{10} * \frac{9}{10} * \frac{8}{10}$

= 0.432

Probability(Two Pairs) = 4C2 * $\frac{10}{10} * \frac{1}{10} * \frac{2C2}{2!} * \frac{9}{10} * \frac{1}{10}$

= 0.027

Probability(Three like digits) = 4C3 * $\frac{10}{10} * \frac{1}{10} * \frac{1}{10} * \frac{9}{10}$

= 0.036

Probability(Four like digits) = $\frac{10}{10} * \frac{1}{10} * \frac{1}{10} * \frac{1}{10}$

= 0.001

Combination (i)	Expected Frequency (E_i) = Probability(i)*N	Observed Frequency (O_i)	
Four different Digits	0.504*1000 = 504	540	2.5714
Exactly one pair	0.432*1000 = 432	320	29.037
Two Pairs	0.027*1000 = 27	70	68.4814
Three like digits	0.036*1000= 36	50	5.444
Four like digits	0.001*1000 = 1	20	361
	1000	1000	

So $\chi^2_{calculated} = 466.5298$

Here degree of freedom = n-1 = 5-1 = 4

So $\chi^2_{standard} = 9.49$

Since $\chi^2_{calculated} > \chi^2_{standard}$ the null hypothesis is rejected. So given numbers are not independent.

Numerical

A set of 10,000 4-digit random values have been generated. An observation shows than 5065 values have all different digits, 2000 have 2 of a kind digits, 760 have 3 of a kind, 1500 have 2 pairs and 675 have all same digits. Test these values for randomness using Poker test (Use α :0.05).

Numerical

Write an algorithm for gap test. Formulate 4-digit poker test with suitable data with example.

2071 Bhadra

Poker test for 5 digit random number

 Possibilities for 5 digit number:

a. All different digits
b. Exactly one pair
c. Two pairs
d. Three of a kind
e. Full House/Three of a kind + Two of a kind
f. Four of a kind
g. Five of a kind

Calculating Probabilities of each possibility

1. Probability(All different Digits) $= \frac{10}{10} * \frac{9}{10} * \frac{8}{10} * \frac{7}{10} * \frac{6}{10}$
   $= 0.3024$
2. Probability(Exactly one pair) $= 5C2 * \frac{10}{10} * \frac{1}{10} * \frac{9}{10} * \frac{8}{10} * \frac{7}{10}$
   $= 0.504$
3. Probability(Two Pairs) $= 5C2 * \frac{10}{10} * \frac{1}{10} * \frac{3C2}{2!} * \frac{9}{10} * \frac{1}{10} * \frac{8}{10}$
   $= 0.108$
4. Probability(Three of a kind) $= 5C3 * \frac{10}{10} * \frac{1}{10} * \frac{1}{10} * \frac{9}{10} * \frac{8}{10}$
   $= 0.072$

Calculating Probabilities of each possibility

5. Probability(Full House/Three of a kind + Two of a kind)

$$= 5C3*\frac{10}{10}*\frac{1}{10}*\frac{1}{10}*2C2*\frac{9}{10}*\frac{1}{10}$$

$$= 0.009$$

6. Probability(Four of a kind) $= 5C4*\frac{10}{10}*\frac{1}{10}*\frac{1}{10}*\frac{1}{10}*\frac{9}{10}$

$$= 0.0045$$

7. Probability(Five of a kind) $= \frac{10}{10}*\frac{1}{10}*\frac{1}{10}*\frac{1}{10}*\frac{1}{10}$

$$=0.0001$$

Numerical

Write an algorithm for Gap Test. A sequence of 10,000 five digit numbers has been generated and analysis indicates the following combinations and frequencies. Based on Poker Test check whether the number are independent. Use $\chi^2 0.056,6 = 12.592$ [4+6]

Combinations	Observed Frequencies
All Different	3044
One pair	5020
Two pair	1090
Three of a kind	700
Full house	95
Four of a kind	40
Five of a kind	11
Total	**10,000**

soln: Let H_0 be null hypothesis.

H_0 : The numbers are independent

Total number of five-digit numbers(N) = 10000

Here, Probability(All different Digits) = $\frac{10}{10} * \frac{9}{10} * \frac{8}{10} * \frac{7}{10} * \frac{6}{10}$

= 0.3024

Probability(one pair) =5C2* $\frac{10}{10} * \frac{1}{10} * \frac{9}{10} * \frac{8}{10} * \frac{7}{10}$

= 0.504

Probability(Two Pairs) = 5C2 * $\frac{10}{10} * \frac{1}{10} * \frac{3C2}{2!} * \frac{9}{10} * \frac{1}{10} * \frac{8}{10}$

= 0.108

Probability(Three of a kind) = 5C3 * $\frac{10}{10} * \frac{1}{10} * \frac{1}{10} * \frac{9}{10} * \frac{8}{10}$

= 0.072

- Probability(Full House/Three of a kind + Two of a kind)

$$= 5C3*\frac{10}{10} * \frac{1}{10} * \frac{1}{10} * 2C2 * \frac{9}{10} * \frac{1}{10}$$

$$= 0.009$$

Probability(Four of a kind) $= 5C4 * \frac{10}{10} * \frac{1}{10} * \frac{1}{10} * \frac{1}{10} * \frac{9}{10}$

$$= 0.0045$$

Probability(Five of a kind) $= \frac{10}{10}*\frac{1}{10} * \frac{1}{10} * \frac{1}{10} * \frac{1}{10}$

$$=0.0001$$

Combination (i)	Expected Frequency (E_i) = Probability(i)*N	Observed Frequency (O_i)	
All different Digits	3024	3044	0.1322
one pair	5040	5020	0.0793
Two Pairs	1080	1090	0.0925
Three of a kind	720	700	0.5556
Full House	90	95	0.2778
Four of a kind	45	40	0.5556
Five of a kind	1	11	100
	10000	10000	

So $\chi^2_{\text{calculated}} = 101.693$

Here degree of freedom = n-1 = 7-1 = 6

So $\chi^2_{\text{standard}} = 12.59$

Since $\chi^2_{\text{calculated}} > \chi^2_{\text{standard}}$ the null hypothesis is rejected. So given numbers are not independent.

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

Numerical

A sequence of 10,000 random numbers has been generated and an analysis shows following combinations and frequencies. For $\alpha = 0.05$ check whether generated numbers are independent or not.

Combination	Observed Frequency
All different Digits	3054
one pair	5020
Two Pairs	1073
Three of a kind	710
Full House	95
Four of a kind	44
Five of a kind	4

Runs Test

- The runs test can be used to decide if a dataset is from a random process.
- A run is defined as a series of increasing values or a series of decreasing values.
- A run is a succession of occurrences of certain type preceded and followed by occurrences of the alternate type or by no occurrences at all.
- If N is the total number of numbers in a sequence, the maximum number of runs is N-1 and minimum number of runs is 1.
- Test statistics is

$$Z_0 = \frac{x_r - \mu_r}{\sigma_r}$$

where x_r = observed number of runs

μ_r = expected number of runs

σ_r = standard deviate of number of runs

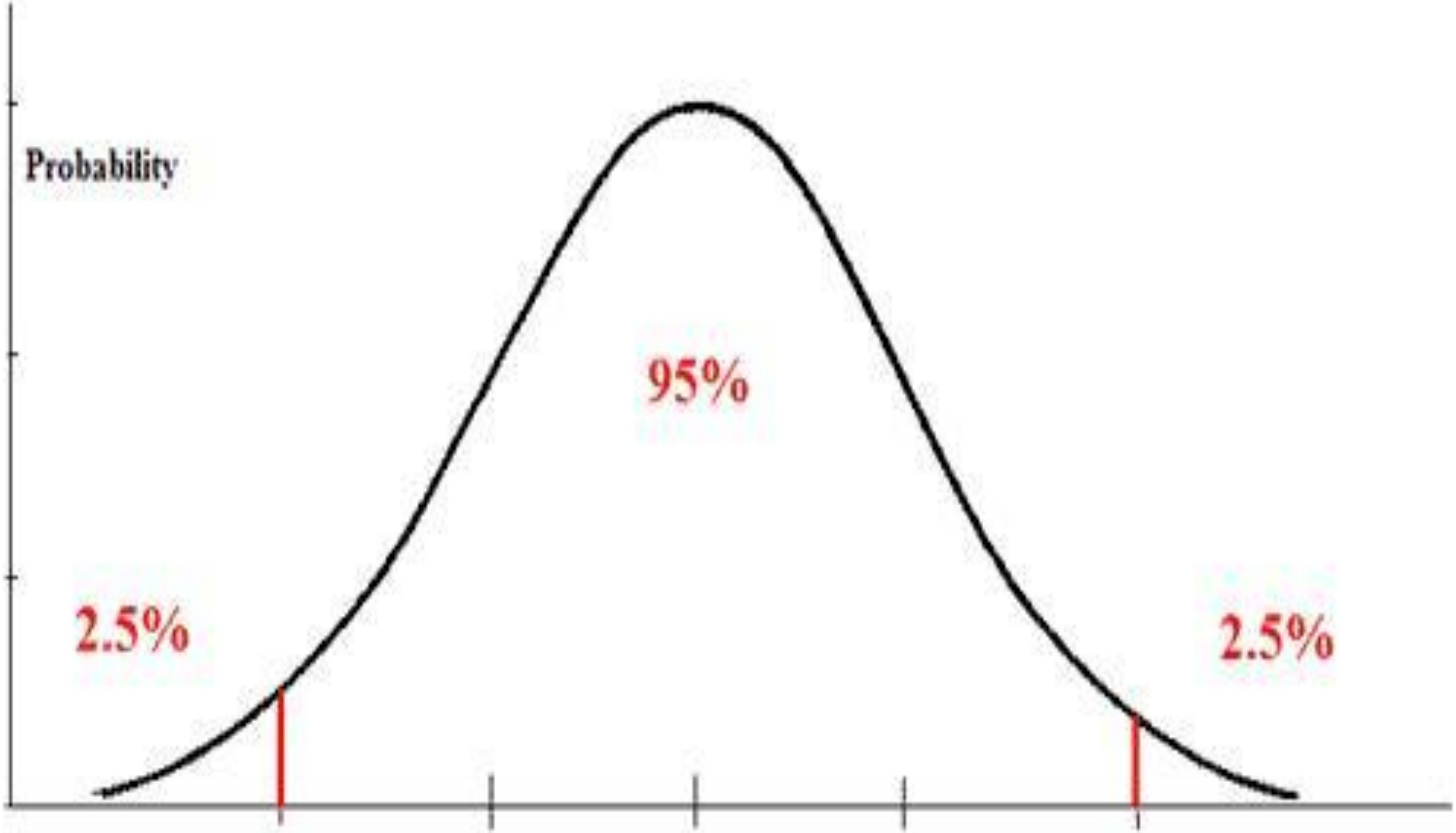
Runs Test

- Here $\mu_r = \dfrac{2n_1n_2}{n_1+n_2} + 1$

and $\sigma_r = \sqrt{\dfrac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1+n_2)^2\,(n_1+n_2-1)}}$

where n1 and n2 represents number of occurrence of a type and number of occurrence of alternate type respectively

Acceptance region for acceptance of hypothesis is $-Z_{\alpha/2} \leq Z_0 \leq Z_{\alpha/2}$


Probability
95%
2.5%
2.5%

z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-0	.50000	.49601	.49202	.48803	.48405	.48006	.47608	.47210	.46812	.46414
-0.1	.46017	.45620	.45224	.44828	.44433	.44034	.43640	.43251	.42858	.42465
-0.2	.42074	.41683	.41294	.40905	.40517	.40129	.39743	.39358	.38974	.38591
-0.3	.38209	.37828	.37448	.37070	.36693	.36317	.35942	.35569	.35197	.34827
-0.4	.34458	.34090	.33724	.33360	.32997	.32636	.32276	.31918	.31561	.31207
-0.5	.30854	.30503	.30153	.29806	.29460	.29116	.28774	.28434	.28096	.27760
-0.6	.27425	.27093	.26763	.26435	.26109	.25785	.25463	.25143	.24825	.24510
-0.7	.24196	.23885	.23576	.23270	.22965	.22663	.22363	.22065	.21770	.21476
-0.8	.21186	.20897	.20611	.20327	.20045	.19766	.19489	.19215	.18943	.18673
-0.9	.18406	.18141	.17879	.17619	.17361	.17106	.16853	.16602	.16354	.16109
-1	.15866	.15625	.15386	.15151	.14917	.14686	.14457	.14231	.14007	.13786
-1.1	.13567	.13350	.13136	.12924	.12714	.12507	.12302	.12100	.11900	.11702
-1.2	.11507	.11314	.11123	.10935	.10749	.10565	.10383	.10204	.10027	.09853
-1.3	.09680	.09510	.09342	.09176	.09012	.08851	.08692	.08534	.08379	.08226
-1.4	.08076	.07927	.07780	.07636	.07493	.07353	.07215	.07078	.06944	.06811
-1.5	.06681	.06552	.06426	.06301	.06178	.06057	.05938	.05821	.05705	.05592
-1.6	.05480	.05370	.05262	.05155	.05050	.04947	.04846	.04746	.04648	.04551
-1.7	.04457	.04363	.04272	.04182	.04093	.04006	.03920	.03836	.03754	.03673
-1.8	.03593	.03515	.03438	.03362	.03288	.03216	.03144	.03074	.03005	.02938
-1.9	.02872	.02807	.02743	.02680	.02619	.02559	.02500	.02442	.02385	.02330
-2	.02275	.02222	.02169	.02118	.02068	.02018	.01970	.01923	.01876	.01831
-2.1	.01786	.01743	.01700	.01659	.01618	.01578	.01539	.01500	.01463	.01426
-2.2	.01390	.01355	.01321	.01287	.01255	.01222	.01191	.01160	.01130	.01101
-2.3	.01072	.01044	.01017	.00990	.00964	.00939	.00914	.00889	.00866	.00842
-2.4	.00820	.00798	.00776	.00755	.00734	.00714	.00695	.00676	.00657	.00639
-2.5	.00621	.00604	.00587	.00570	.00554	.00539	.00523	.00508	.00494	.00480
-2.6	.00466	.00453	.00440	.00427	.00415	.00402	.00391	.00379	.00368	.00357
-2.7	.00347	.00336	.00326	.00317	.00307	.00298	.00289	.00280	.00272	.00264
-2.8	.00256	.00248	.00240	.00233	.00226	.00219	.00212	.00205	.00199	.00193
-2.9	.00187	.00181	.00175	.00169	.00164	.00159	.00154	.00149	.00144	.00139
-3	.00135	.00131	.00126	.00122	.00118	.00114	.00111	.00107	.00104	.00100
-3.1	.00097	.00094	.00090	.00087	.00084	.00082	.00079	.00076	.00074	.00071
-3.2	.00069	.00066	.00064	.00062	.00060	.00058	.00056	.00054	.00052	.00050
-3.3	.00048	.00047	.00045	.00043	.00042	.00040	.00039	.00038	.00036	.00035
-3.4	.00034	.00032	.00031	.00030	.00029	.00028	.00027	.00026	.00025	.00024
-3.5	.00023	.00022	.00022	.00021	.00020	.00019	.00019	.00018	.00017	.00017
-3.6	.00016	.00015	.00015	.00014	.00014	.00013	.00013	.00012	.00012	.00011
-3.7	.00011	.00010	.00010	.00010	.00009	.00009	.00008	.00008	.00008	.00008
-3.8	.00007	.00007	.00007	.00006	.00006	.00006	.00006	.00005	.00005	.00005
-3.9	.00005	.00005	.00004	.00004	.00004	.00004	.00004	.00004	.00003	.00003
-4	.00003	.00003	.00003	.00003	.00003	.00003	.00002	.00002	.00002	.00002

Numerical

Consider the following series representing 44 computer chips which may be either Defective(D) or Acceptable(A). Based on the runs up and down, determine the hypothesis of independence for $\alpha = 0.05$.

D A A A A A A A D D D D A A A A A A A A D D A A A A A A A A

D D D D A A A A A A A A A A

soln: Here observed number of runs $x_r = 8$

Let n_1 represents number of Defective(D) chips and n_2 represents Acceptable(A) chips

So $n_1 = 11$ and $n_2 = 33$

Expected Number of runs $\mu_r = \frac{2n_1n_2}{n_1+n_2} + 1$

$= 17.6$

Similarly $\sigma_r = \sqrt{\frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1+n_2)^2\,(n_1+n_2-1)}} = 2.4387$

So $Z_0 = \frac{X_r - \mu_r}{\sigma_r} = -3.895$

Confidence Level	Alpha	Alpha/2	z alpha/2
90%	10%	5.0%	1.645
95%	5%	2.5%	1.96
98%	2%	1.0%	2.326
99%	1%	0.5%	2.576

Here $-Z_{\alpha/2} = -1.96$

Since $Z0 < -Z_{\alpha/2}$ (i.e. $-Z_{\alpha/2} \le Z_0 \le Z_{\alpha/2}$ not valid) the hypothesis for independence is rejected.

Auto-Correlation Test

- Autocorrelation is a statistical test that determines whether a random number generator is producing independent random numbers in a sequence.
- The tests for auto-correlation are concerned with the dependence between numbers in a sequence.
- The test computes the autocorrelation between every m numbers (m is also known as the lag) starting with the i^{th} number (i is also known as the index).
- Important variables to remember:
  1. m - is the lag, the space between the numbers being tested
  2. i - is the index, or the number in the sequence that you start with
  3. N - the number of numbers generated in a sequence
  4. M – is the largest integer such that $i + (M + 1)m \leq N$

Auto-Correlation Test Algorithm

1. Define the hypothesis.
2. Find the value of ‘i’ and lag value ‘m’
3. Using the value of ‘i’, ‘m’ and ‘N’ calculate the value of M as $i + (M + 1)m \leq N$ where
    a. m - is the lag, the space between the numbers being tested
    b. i - is the index, or the number in the sequence that we start with
    c. N - the number of numbers generated in a sequence
    d. M – is the largest integer such that $i + (M + 1)m \leq N$

4. Compute the test statistics as:

$$Z_0 = \frac{\rho_{in}}{\sigma_{\rho in}}$$

where $\rho_{in} = \frac{1}{M+1}\left[\sum_{k=0}^{M}\left(R_{(i\ +\ km)} * R_{(i\ +\ (k+1)m)}\right)\right] - 0.25$

$$\sigma_{\rho in} = \frac{\sqrt{13M+7}}{12(M+1)}$$

5. If $-Z_{\alpha/2} \leq Z_0 \leq Z_{\alpha/2}$ then the null hypothesis is not rejected.

Numerical

Consider a sequence of 30 numbers generated by a random number generator. Test whether the 3rd, 8th and 13th numbers in the sequence are auto-correlated with $\alpha = 0.05$ and $Z_{0.025} = 1.96$.

0.12, 0.01, 0.23, 0.28, 0.89, 0.31, 0.64, 0.28, 0.83, 0.93, 0.99, 0.15, 0.33, 0.35, 0.91

0.41, 0.60, 0.27, 0.75, 0.88, 0.68, 0.49, 0.05, 0.43, 0.95, 0.58, 0.19, 0.36, 0.69, 0.87

Soln: Let H_0 represent null hypothesis where H_0 : Numbers in sequence are auto-correlated.

Here m = 5

We have, $i + (M + 1)m \leq N$

or, $3 + (M + 1)*5 \leq 30$

or, $M \leq 4.4 \sim 4$

We have, $\rho_{in} = \frac{1}{M+1}\left[\sum_{k=0}^{M}\left(R_{(i+km)} * R_{(i+(k+1)m)}\right)\right] - 0.25$

$$= \frac{1}{4+1}\left[\sum_{k=0}^{4}\left(R_{(i+km)} * R_{(i+(k+1)m)}\right)\right] - 0.25$$

$$= \frac{1}{5}\left[R_3 * R_8 + R_8 * R_{13} + R_{13} * R_{18} + R_{18} * R_{23} + R_{23} * R_{28}\right] - 0.25$$

$$= -0.1945$$

$$\sigma_{\rho in} = \frac{\sqrt{13M+7}}{12(M+1)} = \frac{\sqrt{13*4+7}}{12(4+1)} = 0.1280$$

Now $Z_0 = \frac{\rho_{in}}{\sigma_{\rho in}} = -1.581$

For $\alpha = 0.05$, $Z_{\alpha/2} = 1.96$

Since $-Z_{\alpha/2} \leq Z_0 \leq Z_{\alpha/2}$ then the null hypothesis is not rejected.
Hence the generated numbers are auto-correlated.

Methods of generating non-uniform Variables: Generating discrete distributions

- A discrete distribution describes the probability of occurrence of each value of a discrete random variable.
- A discrete random variable is a random variable that has countable values such as list of non-negative integers.
- When the discrete distribution is uniform, the requirement is to pick one of N alternatives with equal probability given to each.
- Given a random number $U (0 \leq U < 1)$, the process of multiplying by N and taking the integral portion of the product, which is denoted mathematically by the expression $[UN]$, gives N different outputs. The output are the numbers $0,1,2,\ldots,(N-1)$.
- The result can be changed to the range of values C to $N+C-1$ by adding C.

Methods of generating non-uniform Variables: Generating discrete distributions

- A discrete distribution describes the probability of occurrence of each value of a discrete random variable.
- A discrete random variable is a random variable that has countable values such as list of non-negative integers.
- When the discrete distribution is uniform, the requirement is to pick one of N alternatives with equal probability given to each.
- Given a random number $U (0 \leq U < 1)$, the process of multiplying by N and taking the integral portion of the product, which is denoted mathematically by the expression $[UN]$, gives N different outputs. The output are the numbers $0,1,2,\ldots,(N-1)$.

Methods of generating non-uniform Variables: Generating discrete distributions

- The result can be changed to the range of values C to N+C-1 by adding C.
- Generally, the requirement is for a discrete distribution that is not uniform, so that a different probability is associated with each output.

Number of Items X_i	Number of Customers N_i	Probability Distribution $P(X_i)$	Cumulative Probability Distribution
1	25	0.10	0.10
2	128	0.51	0.61
3	47	0.19	0.8
4	38	0.15	0.95
5	12	0.05	1

Generating discrete distributions

- Suppose, for example , it is necessary to generate a random variable representing the number of items bought by a customer at store , where the probability function is the discrete distribution given in previous table.
- A table is formed to list the number of items X , and the cumulative probability Y, as shown below:

Number of Items X	Probability P(X)	Cumulative Probability(Y)
1	0.10	0.10
2	0.51	0.61
3	0.19	0.8
4	0.15	0.95
5	0.05	1

Generating discrete distributions

- Taking the output of a uniform random number generator, U, the value is compared with the values of Y.
- If the value falls in an interval $Y_i < U \le Y_{i+1}$ (i=0,1,…….,4), the corresponding value of X_{i+1} is taken as desired output.
- It is not necessary that the intervals be in any particular order.
- A computer routine will usually search the table from the first entry .
- The amount of searching can be <u>minimized by selecting the intervals in decreasing order of probability</u>

Generating discrete distributions

- For computer routine above data can be arranged as:

Probability	Cumulative Probability	Number of Items
0.51	0.51	2
0.19	0.70	3
0.15	0.85	4
0.10	0.95	1
0.05	1	5

- With this arrangement , 51% of the searches will only need to go to the first entry, 70% to the first or second and so on.
- With the original ordering, only 10% are satisfied with the first entry and only 61% with the first two

Inversion, Rejection and Composition - Inversion

- In the simplest case of inversion, we have a continuous random variable X with a strictly increasing distribution function F.
- Then F has an inverse F^{-1} defined on the open interval (0,1): for $0<u<1$, $F^{-1}(u)$ is the unique real number x such that $F(x)=u$ i.e.

$$F(F^{-1}(u))=u, \text{ and } F^{-1}(F(x))=x$$

$$P(F^{-1}(u)\leq x) = P(u\leq F(x)) = F(x)$$

- Let u~unif(0,1) denote a uniform random variable on (0,1) Then <u>$F^{-1}(u)$ has distribution function F.</u>

Inversion, Rejection and Composition - Inversion

- To extend this result to a general distribution function F, the generalized inverse of F is:

$$F^{-}_{0<u<1}(u) = \inf\{x: F(x) \geq u\}$$

Where inf represents the Infimum value(greatest lower bound value)

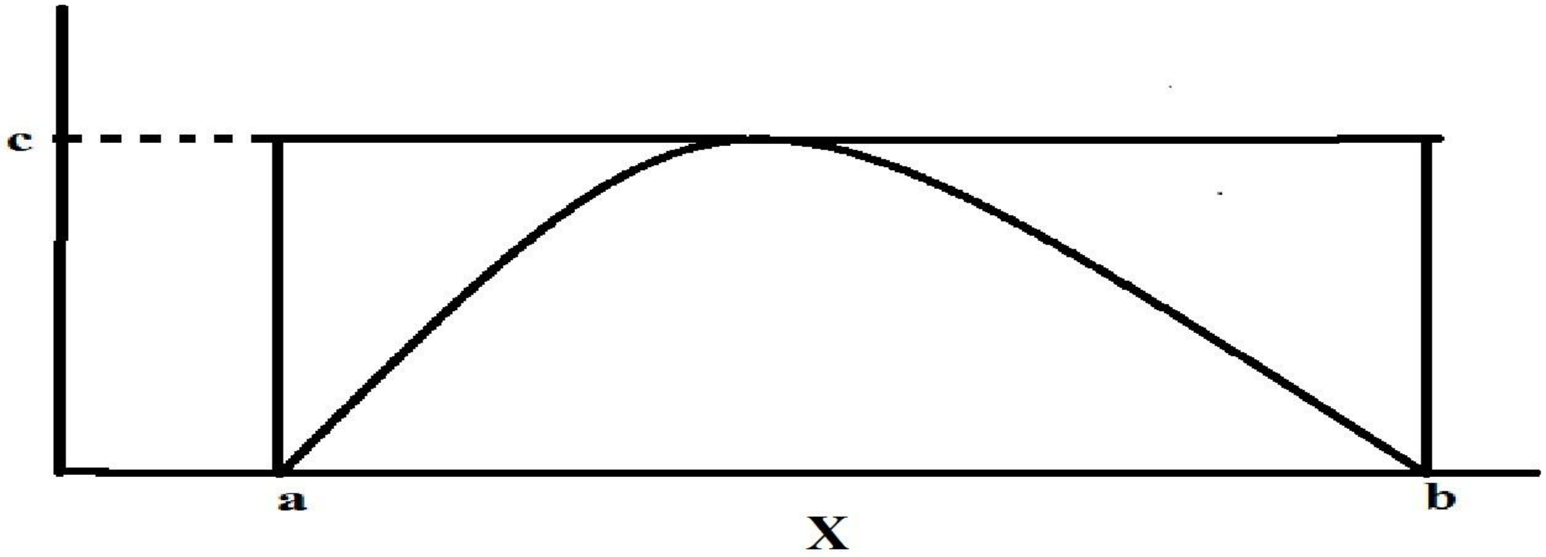
And $F^{-}(u)$ represents u- quartile

Inversion, Rejection and Composition - Rejection

- The rejection method is applied when the probability density function f(x), has a lower and upper limit to its range, lower bound a and b and an upper bound c respectively.
- The method can be specified as follows:
  - Compute the values of two independent uniformly distributed variates(a quantity having a numerical value for each member of group) U1 and U2.
  - Compute $X_0=a+U_1(b-a)$.
  - Compute $Y_0=cU_2$
  - Either accept X0 as the desired output otherwise repeat the process with two new uniform variates.

Inversion, Rejection and Composition - Rejection

- This method is closely related to the process of evaluating an integral using Monte-Carlo technique. The probability density function is enclosed in a rectangle with side lengths b-a and c.
- In the rejection method the curve is probability density function so that the area under curve must be 1 i.e. c(b-a)=1.



Inversion, Rejection and Composition - Rejection

- The probability of X being less than or equal to X_0 is by definition,

$$F(X_0) = \frac{\int_a^{X_0} f(x)dx}{c(X_0 - a)} * \frac{(X_0 - a)}{(b - a)}$$

- Since c(b-a)=1, it follows that

$$F(X_0) = \int_a^{X_0} f(x)dx$$

which shows that X0 has the desired distribution

Inversion, Rejection and Composition - Composition

- Sometimes the random variables X of interest involves the sum of n>1 independent random variables:

$$X=Y_1+Y_2+Y_3+\ldots+Y_n$$

- To generate a value for X, we can generate a value for each of the random variables Y_1, Y_2, Y_3, Y_n and add them together. This is called composition.
- Composition can also be used to generate random numbers that are approximately normally distributed.
- The normal distribution is one of the most important and frequently used continuo
- The notion $N(\mu , \sigma)$ refers to the normal distribution with mean μ and variance σ^2.

Inversion, Rejection and Composition - Composition

- The central limit theorem in probability says that if Y_1, Y_2, Y_3, ….Y_n are independent and identically distributed random variables with mean μ and positive variance σ^2, then random variable is

$$Z = \frac{\sqrt{n}(\frac{x}{n} - \mu)}{\sigma}$$

Convolution Method

- The probability distribution of a sum of two or more independent random variables is called a convolution of the distributions of the original variables.
- The convolution method thus refers to adding together two or more random variables to obtain a new random variable with the desired distribution.
- Technique can be used for all random variables X that can be expressed as the sum of n random variables

$$X = Y1 + Y2 + Y3 + . . . + Yn$$

- In this case, one can generate a random variate X by generating n random variates, one from each of the Yi, and summing them.

Numerical

Use Chi-Square test to test the uniformity of following random numbers for 95% confidence level And given critical value for degree of freedom = 8 is 15.51.

25	33	5	54	9	31	14	40	17	52	33
49	61	62	26	67	6	28	55	22	68	34
50	2	66	77	86	12	41	88	19	96	70
81	47	85	3	59	94	8	42	71	37	79
82	51	91	11	75	43	39	44	64	58	46