NEW STATISTICAL BINARY RANDOM TEST

Dr. Imad F. Elshaikhli **Luay Flaieh Hasan**

Abstract

Binary streams[1] can be tested, using standard tests[2]. The efficiency of some tests is not remarkable big, because they do not examine some stream features, which are important to be verified [for example, mono-bit test satisfies every stream that consists of approximate one half zeros (i.e. ones), no matter if all of them are consecutive or alternate), and so several tests are to be combined. In our proposed test, distribution of frequencies of blocks[3] and gaps[4] are determined and for given number of the blocks and the gaps expected distribution of frequencies of the blocks and the gaps is expected. χ^2 is calculated, and it is checked whether the probability of calculated χ^2 is above the threshold θ of significance. In the same way both distributions (blocks and gaps) are evaluated.

[1] streams with two elements: 0 and 1.
[2] see standard tests P-2 for details.
[3] number of consecutive ones preceded and followed by zero.
[4] number of consecutive zeros preceded and followed by ones.

Introduction:

If a binary sequence are to be used as enciphering sequences in a stream cipher system, it must resemble a random sequence[1,6,7]. Thus it is important to apply statistical tests to our sequence and to check that they appear to be random. In this paper we will present a new test for checking the randomness of a binary sequence which can used to substitute the standard tests.

Standard Random Tests:[2,3,4,7]

The randomness of a binary sequence can be tested using several statistical tests:

1. Frequency Test

For a random sequence of length n the number n_0 of zeros and the number n_1 of ones should be about the same.

$$T = \frac{(n_0 - n_1)^2}{n}$$

Is distributed according to the χ2-distribution with 1 degree of freedom.

2. Serial Test

In this test the bit pair are considered

◊ sequence of pairs of length n/2

n_{00}= # 00-pairs, n01= # 01-pairs

n10= # 10-pairs, n11= # 11-pairs

n00+ n01+ n10+ n11=n/2

Expected value for nij=n/8

$$T = \sum \frac{\left(n_{ij} - \frac{n}{8}\right)^2}{\frac{n}{8}}$$

χ2-distribution with 3 degrees of freedom.

3. Poker Test

In this test the m-bit groups are considered.

◊sequence of m-bit groups of length n/m

n_i = # groups with i ones.

Expected value for $n_i = \binom{m}{i}\frac{1}{2^m}\frac{n}{m}$

$$T = \sum_{i=0}^{m} \frac{\left(n_i - \binom{m}{i} \frac{1}{2^m} \frac{n}{m} \right)^2}{\binom{m}{i} \frac{1}{2^m} \frac{n}{m}}$$

χ2-distribution with m degrees of freedom.

4.Run Test

Gap of length i is called 0-run of length i.
Block of length i is called 1-run of length i.
n_{0i} = # 0-run of length i.
n_{1i} = # 1-run of length i.
The expected number of runs of length i (both 0-runs and 1-runs) is $\frac{n-i+3}{2^{i+2}} \approx \frac{n}{2^i}$

$$T_0 = \sum_{i=1}^{k} \frac{\left(n_{0i} - \frac{n}{2^{i+2}} \right)^2}{\frac{n}{2^{i+2}}} \qquad T_1 = \sum_{i=1}^{k} \frac{\left(n_{1i} - \frac{n}{2^{i+2}} \right)^2}{\frac{n}{2^{i+2}}}$$

Have approximately χ2-distribution with k-1 degrees of freedom.

5.Autocorrelation Test

A given sequence $a_1,a_2,\ldots,a_{n-1},$an is compared with a shift of itself
$n_0(\partial) = \#\{n : a_{n+\partial} = a_n\}$
$n_1(\partial) = \#\{n : a_{n+\partial} \neq a_n\}$

$n_0(\partial)+n_1(\partial)=n-\partial$

the expected value for n0(∂) and n1(∂) is $\frac{n-\partial}{2}$

$$T = \frac{(n_0(\partial) - n_1(\partial))^2}{n-\partial}$$

has χ2-distribution with 1 degree of freedom.

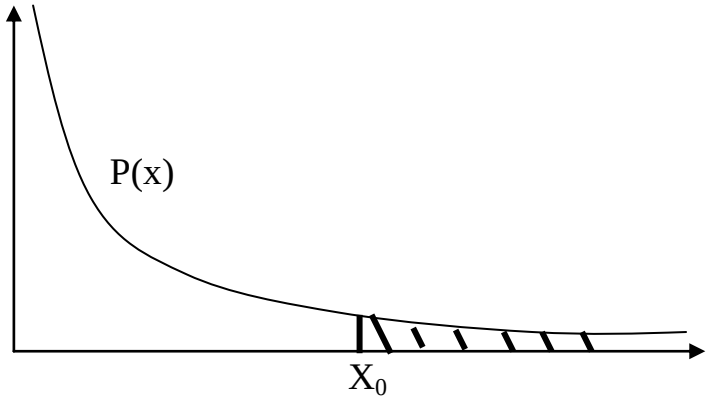
Chi Square Method (χ^2-test)[5,8]

Assume that the outcome of a random experiment falls into one of k categories and assume by hypothesis that p_i is the probability that the

$$\sum_{i=1}^{k} n_i = n \qquad T = \sum_{i=1}^{k} \frac{(n_i - np_i)^2}{np_i}$$

outcome falls into category i. Assume that n independent observations are made and let ni be the number of observations falling into category i.
In order to test the hypothesis the quantity is computed. If the hypothesis is true, the value T is distributed according to χ2-distribution with k-1 degree of freedom.

Significance Level[5,8]



The hypothesis is rejected if n_0 and n_1 are too different from the expected value n/2, i.e. if is too big. That means we set some pass mark x_0 and reject the hypothesis if χ^2 is greater than x_0. Probability for a wrong rejection of the hypothesis $= \int_{x_0}^{\infty} P(X)dX = \alpha$

α is called the significance level of the test.

Decision

Suppose our pass mark is 95%. This means that a given sequence passes the tests if its value lies in the range in which we would expect to find 95% of all truly random sequences.For example, the passmark for the 5% significance level with one degree of freedom is 3.84.

Proposed Test

Throughout the following discussions we divide the binary random of our sequence into number of samples, a sample of n bits contains N_{0i} of consecutive zeros of length i and N_{1i} of consecutive ones of length i. The algorithm can be described now (See APPENDEX A for details):

x2: If no more sample goto x1;
New sample test
Calculate N_{0i} and N_{1i} (i=1..50)
Find the value of H_0 represent χ^2 for N_{0i} with d degrees of freedom.
Find the value of H_1 represent χ^2 for N_{1i} with d degrees of freedom.
Calculate the significance level $TST_0 = \left(\chi^2_{H_0,d}\right)^{-1}$ **and** $TST_1 = \left(\chi^2_{H_1,d}\right)^{-1}$
If $TST_0 < \theta$ (threshold), increment np_0
If TST1 < θ (threshold), increment np_1
$e_{1i} = e_{1i} + \sum_i N_{1i}$ **and** $e_{0i} = e_{0i} + \sum_i N_{0i}$ **(i=1..50)**
k=11-(1+10TST_0) increment B_{0k}
k=11-(1+10TST_1) increment B_{1k}
goto x2
x1: Find the value of H_0 represent χ^2 for B_{0i} with 9 degrees of freedom.
Find the value of H_1 represent χ^2 for B_{1i} with 9 degrees of freedom.
Calculate the significance level $T_0 = \left(\chi^2_{H_0,9}\right)^{-1}$ **and** $T_1 = \left(\chi^2_{H1,9}\right)^{-1}$
Find the value of H_0 represent χ^2 for e_{0i} with d degrees of freedom.
Find the value of H_1 represent χ^2 for e_{1i} with d degrees of freedom.
Calculate the significance level $SEG_0 = \left(\chi^2_{H_0,d}\right)^{-1}$ **and** $SEG_1 = \left(\chi^2_{H_1,d}\right)^{-1}$

The sequence has good properties if np_0, np_1 are equals to zero and T_0, T_1, SEG_0 and SEG_1 are all greater than or equals to the threshold.

Our aim is to show that the new test can:
1.Decrease the amount of computation.
2.Examine some stream features, not examined in standard tests.

Experimental Results

In this part of research we test three different sequences are generated by three different systems (see APPENDIX B for systems details, and APPENDIX C for testing results of output sequences generated by the systems), the same sequences are tests in standard and proposal tests in the same time.

Tests of SYS1 Results

We test the sequence generated by this system by both the proposed and the standard tests. The sequence pass through standard tests and two samples of nine were failed using the proposal test. This proved that some features observed by the proposed test but not by the standard tests.

Tests of SYS2 Results

The sequence generated by this system by the proposed and the standard test pass through both tests, that mean all features of the standard test can be observed by the proposal test.

Tests of SYS3 Results

The sequence generated by this system by the proposed and the standard test fail through both tests, that mean again all features of the standard test can be observed by the proposal test.

Conclusions and Future Work

A new statistical test has been proposed for testing the randomness of binary sequences. We showed that this test can reduce the amount of calculations needed for applying the standard tests, at the same time we showed that it can examine the stream features that standard test cannot. This test can be used as a cryptanalysis tool, this is an ongoing work.

References:

1. Bruce S., "Applied Cryptography, Second Edition: Protocols, Algorithms And Source Coding In C", John Wiley & Sons, Inc., 1996.
2. Beker H. and Piper F. "Cipher Systems, The Protection Of Communications", Northwood Publications, U. K., 1982.
3. Carter G., "Statistical Tests For Randomness", EISS, Karlsruhe, England, 1989.
4. Carter G. & Hooper M. H., "Randomness Properties Of Binary Sequences", Racal Comsec Ltd., England, 1989.
5. John E., "Modern Elementary Statistics: Seventh Edition", Prentice Hall International, Inc., 1988.
6. Massey J. L., "Cryptography Fundamentals And Applications", ", Advanced Technology Seminars, Zurich, 1989.
7. Massey J. L., "Coding And Cryptography", Advanced Technology Seminars, September 4-6, 1984.
8. Roy R. & Harry S., "Statistics: A Beginning", John Wiley & Sons, Inc.

APPENDECIES

APPENDIX A (Details of our proposed test)

1. **If S_n>sample_no., goto 21**
2. **Increment S_n (Test new sample)**
3. **compute number of consecutive ones N_{1i} and zeros N_{0i} (i=1..50)**
4. $Z = \sum_{i=1}^{50} i(N_{0i} + N_{1i})$
5. **Z=Z/4**
6. **Store $Z/2^i$ into M_i (i=1..50)**
7. $S = \sum_{i=50.1}^{S\geq 10} M_i$
8. **d=i-1 (degree of freedom)**
9. **M_i=S**
10. $P_0 = \sum_{j=50}^{i} N_{0j}$ **and** $P_1 = \sum_{j=50}^{i} N_{1j}$
11. **$G_{0i}=P_0$ and $G_{1i}=P_1$**
12. **store N_{0j} into G_{0j} and N_{1j} into G_{1j} (j=1..d)**
13. $H_0 = \sum_{j=1}^{i} \left(\frac{(G_{0j} - M_j)^2}{M_j} \right)$ **and** $H_1 = \sum_{j=1}^{i} \left(\frac{(G_{1j} - M_j)^2}{M_j} \right)$
14. $TST_0 = \left(\chi^2_{H_0,d}\right)^{-1}$ **and** $TST_1 = \left(\chi^2_{H_1,d}\right)^{-1}$ **(significance level)**
15. **if TST_0 < Threshold, increment np_0**
16. **if TST_1 < Threshold, increment np_1**
17. $e_{0i} = e_{0i} + \sum_i N_{0i}$ **and** $e_{1i} = e_{1i} + \sum_i N_{1i}$ **(i=1..50)**
18. **k=11-(1+10TST_0) increment B_{0k}**
19. **k=11-(1+10TST_1) increment B_{1k}**
20. **increment S_n, goto 1**
21. **OC = Sample_no./10**
22. $H_0 = \sum_{i=1}^{10} \left(\frac{(OC - B_{0i})^2}{OC} \right)$ **and** $H_1 = \sum_{i=1}^{10} \left(\frac{(OC - B_{1i})^2}{OC} \right)$
23. $T_0 = \left(\chi^2_{H_0,9}\right)^{-1}$ **and** $T_1 = \left(\chi^2_{H1,9}\right)^{-1}$

24. $X_0 = \sum_{i=1}^{50} ie_{0i}$ **and** $X_1 = \sum_{i=1}^{50} ie_{1i}$

25. **total=X_0+X_1**

26. **U=X_0/total**

27. $from = 0.5 - 3\sqrt{\frac{1}{4} total}$ **and** $to = 0.5 + 3\sqrt{\frac{1}{4} total}$

28. **Z=total/4**

29. **store Z/2^i into M_i (i=1..50)**

30. $S = \sum_{i=50.1}^{S \geq 10} M_i$

31. **d=i-1 (degree of freedom)**

32. **M_i=S**

33. $P_0 = \sum_{j=50}^{i} e_{0j}$ **and** $P_1 = \sum_{j=50}^{i} e_{1j}$

34. **G_{0i}=P_0 and G_{1i}=P_1**

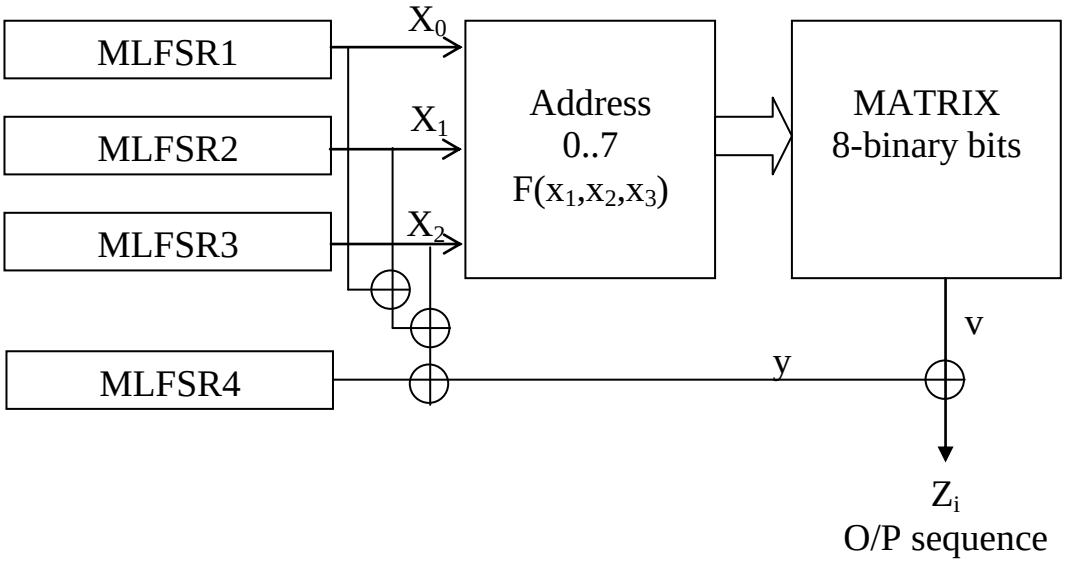
35. **store e_{0j} into G_{0j} and e_{1j} into G_{1j} (j=1..d)**

36. $H_0 = \sum_{j=1}^{i} \left(\frac{(G_{0j} - M_j)^2}{M_j} \right)$ **and** $H_1 = \sum_{j=1}^{i} \left(\frac{(G_{1j} - M_j)^2}{M_j} \right)$

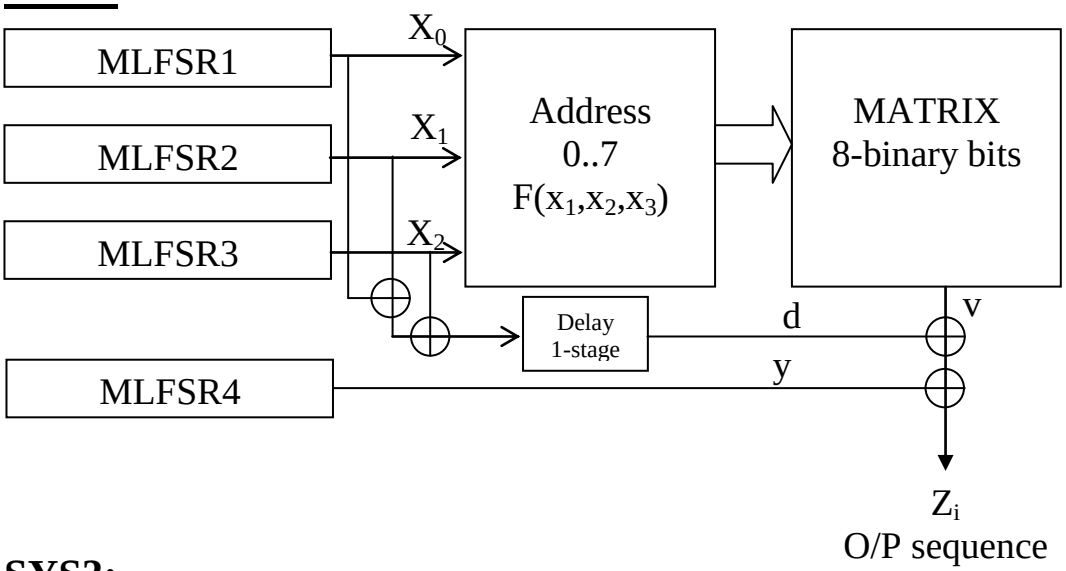
37. $SEG_0 = \left(\chi^2_{H_0,d}\right)^{-1}$ **and** $SEG_1 = \left(\chi^2_{H_1,d}\right)^{-1}$ **(significance level)**

APPENDIX B (Details of systems)

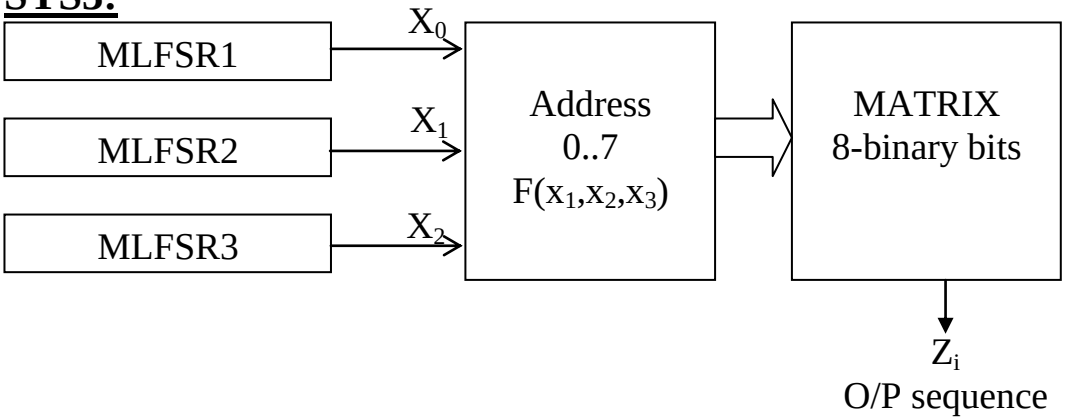
SYS1:



SYS2:



SYS3:



MLFSR: MAXIMAL LINEAR FEEDBACK SHIFT REGISTER.

APPENDIX C (Result of proposed and standard tests)[1]

Proposed Tests of SYS1

Distribution of 0-Groups

1	2	3	4	5	6	7	8	9	10	11	12
2320	1143	572	280	156	72	33	23	5	2	4	1

CHi^2 is: 5.0703 NO#. Degrees Of Freedom is: 8
Corresponding Probability is: 0.75004
NUMBER OF SAMPLES UNDER SIGNIFICANCE THRESHOLD OF 0.001 IS: 0

Distribution of 1-Groups

1	2	3	4	5	6	7	8	9	10	11	12	13
2308	1139	589	293	129	80	38	20	7	3	1	3	2

CHi^2 is: 3.5408 NO#. Degrees Of Freedom is: 8
Corresponding Probability is: 0.89600
NUMBER OF SAMPLES UNDER SIGNIFICANCE THRESHOLD OF 0.001 IS: 2

DISTRIBUTION CHi^2 1-GROUPS :

1	2	3	4	5	6	7	8	9	10
0	0	1	2	1	0	1	0	1	3

CHi^2 is: 9.8889 and 9 d.o.f corresponding prob. is: 0.3596

DISTRIBTION CHIi^2 0-GROUPS :

1	2	3	4	5	6	7	8	9	10
1	1	0	0	1	1	2	0	1	2

CHi^2 is: 5.4444 and 9 d.o.f corresponding prob. is: 0.7940

P(1)= 0.50141 3 SIGMA INTERVAL IS: 0.48895 - 0.51105
TESTED TOTAL 18432 BITS AND 9 SAMPLES

Standard Tests of SYS1

FREQUENCY TEST

PASS VALUE 0.003 WITH FREEDOM DEGREE " 1 " MUST BE <= 3.84

RUN TEST

PASS VALUE T0 = 5.180 WITH FREEDOM DEGREE "11" MUST BE <= 19.391
PASS VALUE T1 = 11.774 WITH FREEDOM DEGREE "12" MUST BE <= 20.742

POKER TEST

PASS VALUE 2.765 WITH FREEDOM DEGREE " 5 " MUST BE <= 11.1

SERIAL TEST

PASS VALUE 2.001 WITH FREEDOM DEGREE " 3 " MUST BE <= 7.81

AUTO_CORRELATION TEST

SHIFT NO. 1 >--> PASS VALUE 0.014
SHIFT NO. 2 >--> PASS VALUE 0.016
SHIFT NO. 3 >--> PASS VALUE 0.092
SHIFT NO. 4 >--> PASS VALUE 1.066
WITH FREEDOM DEGREE " 1 " MUST BE <= 3.84

Proposed Tests of SYS2

Distribution of 0-Groups

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
3509	1823	869	461	228	116	74	20	14	6	5	1	0	0	1

CHi^2 is: 11.6537 NO#. Degrees Of Freedom is: 9
Corresponding Probability is: 0.23354
NUMBER OF SAMPLES UNDER SIGNIFICANCE THRESHOLD OF 0.001 IS: 0

Distribution of 1-Groups

1	2	3	4	5	6	7	8	9	10	11
3464	1870	909	444	222	101	58	32	12	8	1

CHi^2 is: 11.4498 NO#. Degrees Of Freedom is: 9
Corresponding Probability is: 0.24613
NUMBER OF SAMPLES UNDER SIGNIFICANCE THRESHOLD OF 0.001 IS: 0

DISTRIBUTION CHi^2 1-GROUPS :

1	2	3	4	5	6	7	8	9	10
1	1	0	0	1	5	0	1	1	4

CHi^2 is: 18.8571 and 9 d.o.f corresponding prob. is: 0.0264

DISTRIBTION CHIi^2 0-GROUPS :

1	2	3	4	5	6	7	8	9	10
2	1	1	1	1	4	2	0	2	0

CHi^2 is: 8.8571 and 9 d.o.f corresponding prob. is: 0.4506

P(1)= 0.49819 3 SIGMA INTERVAL IS: 0.49114 - 0.50886
TESTED TOTAL 28672 BITS AND 14 SAMPLES

Standard Tests of SYS2

FREQUENCY TEST

PASS VALUE 0.941 WITH FREEDOM DEGREE " 1 " MUST BE <= 3.84

RUN TEST

PASS VALUE T0 = 16.679 WITH FREEDOM DEGREE " 14 " MUST BE <= 23.401
PASS VALUE T1 = 10.651 WITH FREEDOM DEGREE " 10 " MUST BE <= 18.023

POKER TEST

PASS VALUE 2.308 WITH FREEDOM DEGREE " 5 " MUST BE <= 11.1

SERIAL TEST

PASS VALUE 3.641 WITH FREEDOM DEGREE " 3 " MUST BE <= 7.81

AUTO_CORRELATION TEST

SHIFT NO. 1 >--> PASS VALUE 1.092
SHIFT NO. 2 >--> FAIL VALUE 6.454
SHIFT NO. 3 >--> PASS VALUE 0.952
SHIFT NO. 4 >--> PASS VALUE 3.000
WITH FREEDOM DEGREE " 1 " MUST BE <= 3.84

Proposed tests of SYS3:

Distribution of 0-Groups

1	2	3	4	5	6	7	8	9	10	11	12
4158	1598	586	207	82	30	8	5	2	0	0	1

CHi^2 is: 582.2779 NO#. Degrees Of Freedom is: 9
Corresponding Probability is: 0.00000
NUMBER OF SAMPLES UNDER SIGNIFICANCE THRESHOLD OF 0.001 IS: 14

Distribution of 1-Groups

2451 1541 1046 646 349 224 168 91 57 38 15 16 12 8 7 0 4 2 1 0 0 1

CHi^2 is: 1764.0968 NO#. Degrees Of Freedom is: 9
Corresponding Probability is: 0.00000
NUMBER OF SAMPLES UNDER SIGNIFICANCE THRESHOLD OF 0.001 IS: 14

DISTRIBUTION CHi^2 1-GROUPS :

1	2	3	4	5	6	7	8	9	10
0	0	0	0	0	0	0	0	0	14

CHi^2 is: 126.0000 and 9 d.o.f corresponding prob. is: 0.0000

DISTRIBTION CHIi^2 0-GROUPS :

```
-----------------------------------------------
  1   2   3   4   5   6   7   8   9  10
  0   0   0   0   0   0   0   0   0  14
```

CHi^2 is: 126.0000 and 9 d.o.f corresponding prob. is: 0.0000

```
-------------------------------------------------------------------------------
```

P(1)= 0.62835 3 SIGMA INTERVAL IS: 0.49114 - 0.50886
TESTED TOTAL 28672 BITS AND 14 SAMPLES

Standard tests of SYS3:

```
FREQUENCY TEST
------------------
FAIL VALUE 1940.563 WITH FREEDOM DEGREE " 1 " MUST BE <= 3.84

RUN TEST
------------
FAIL VALUE T0 = 594.993 WITH FREEDOM DEGREE "11" MUST BE <= 19.391
FAIL VALUE T1 = 3078.972 WITH FREEDOM DEGREE " 21 " MUST BE <= 32.386

POKER TEST
--------------
FAIL VALUE 2191.547 WITH FREEDOM DEGREE " 5 " MUST BE <= 11.1

SERIAL TEST
---------------
FAIL VALUE 2025.743 WITH FREEDOM DEGREE " 3 " MUST BE <= 7.81

AUTO_CORRELATION TEST
-------------------------
SHIFT NO.  1 >--> FAIL VALUE 140.225
SHIFT NO.  2 >--> FAIL VALUE 104.673
SHIFT NO.  3 >--> FAIL VALUE 145.479
SHIFT NO.  4 >--> FAIL VALUE 104.680
WITH FREEDOM DEGREE " 1 " MUST BE <= 3.84
```

[1]All Programs of Standard test are written by Faez H. Elazawi and Faris A. Elshibli