IOSR Journal of Mathematics (IOSR-JM)
e-ISSN: 2278-5728, p-ISSN: 2319-765X. Volume 12, Issue 1 Ver. V (Jan. - Feb. 2016), PP 17-23
www.iosrjournals.org

Applied Poker Test for General Digital Sequences

Sahar A. Mohammed

Abstract: *The Poker test is considered one of the important statistical randomness tests. This test, for independence, is based on the frequency in which certain digits are repeated in a series of numbers.*
In this paper we will attempt to generalize the binary poker test to be suitable to applied on not only binary sequences, in another word, the generalized poker test could be applied on digital (m-) sequences (for m≥2). The generalized poker test called Grouped poker test.

Keywords: *Statistical Randomness Tests, Poker Test, Hypothesis Test, Chi-Square Test, cryptography.*

I. Introduction

The problem of testing randomness is motivated by the need to evaluate of the quality of different random number generators used by many practical applications including computer simulations, cryptography and communications industry [1].

The **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 ([2]).

Definition(1) [3]: A Pseudo Random Bit Generator (PRBG) is a deterministic algorithm which, given a truly random binary sequence of length k, outputs a binary sequence of length L, k which "appears" to be random. The input to the PRBG is called the seed, while the output of the PRBG is called a pseudorandom bit sequence.

Remark (1) [3]: The χ^2 (chi-square) distribution can be used to compare the goodness-of-fit of the observed frequencies of events to their expected frequencies under a hypothesized distribution. The χ^2 distribution with υ degrees of freedom arises in practice when the squares of υ independent random variables having standard normal distributions are summed.

In 1982, the five statistical tests for local randomness are being found to be applied on binary key stream sequences [4].

In 1994, a new package of randomness introduce by CRYPT-X [5]. This package is a microcomputer package that is intended to be used to test either large binary strings that are to be used as key stream in stream ciphers or block cipher algorithms. One of the tests of this package is subblock test which is a similar to poker test, its partitions the stream into F hands of length m bits. For a stream of size n, where $F/2^m \geq 5$, the total number of hands is $\lfloor n/m \rfloor$, where $\lfloor\ \rfloor$ denotes the integer value. The aim of this test is to show that there is an equal number of each of 2^m possible hands. If f_i denotes the frequency of hand pattern s_i, then the test statistic used is [6]:

$$\text{T}=\left(\frac{2^m}{F}\right)\sum_{i=0}^{2^m-1} f_i^{\,2} - F \qquad \ldots(1)$$

This compared with chi-squared distribution with degree of freedom equal to 2^m-1.

In 2009, Ali et al. [3] introduce a paper that include the generalization of three basic and important tests of the standard statistical basic randomness tests, these tests are frequency, Run and Autocorrelation tests.

In 2014, Ibraheem [7], extends the 2-tuple to d-tuple (d≥3) to apply binary serial test for binary sequences. Second, she generalized the 2-tuple binary serial test to 2-tuple digital serial test for digital (m-) sequences (m≥3) generated from digital generators. Lastly, she extends the 2-tuple to d-tuple digital serial test for digital sequences.

The results of randomness tests of applying the binary and digital poker tests in the binary and digital (m-) sequences were introduced in typical tables using hypothesis test using Chi-square test.

The binary and digital poker test results are obtained by program using version 10.0 of MATLAB Language with no time mentioned to obtain the randomness results.

II. Hypothesis Test for Randomness Using Chi-Square Test

Suppose we have an experiment with $n \geq 2$ possible outcomes, with unknown probabilities $p_1,p_2,...,p_n$, now we attempt to decide which of the two hypotheses H_0 or H_A is true, where H_0: $p_i = p_{0i}$ for all i=1,2,...,n and H_A:$p_i \neq p_{0i}$ for some i=1,2,...,n, where $p_{01},p_{02},...,p_{0n}$ are known.

We consider n independent repetitions of the experiment with the random variables N_i denoting the number of times the i^{th} outcome occurs, for i=1,2,…,n where $\sum_{i=1}^{n} N_i = L$. We use the test statistic [7]:

$$\overset{\frown}{\chi}^2 = \sum_{i=1}^{n} \frac{(N_i - Lp_{0i})^2}{Lp_{0i}} \qquad \ldots(2)$$

we obtain the P-value of the test approximately for large L, using the chi-square table for $P=P(\chi^2 \geq \overset{\frown}{\chi}^2)$, where $\overset{\frown}{\chi}^2$ is the observed value of χ^2 . In particular, a small value of $\overset{\frown}{\chi}^2 \leq \overset{\frown}{\chi}_0^2$, where $\overset{\frown}{\chi}_0^2$ was obtained from Chi-square table at freedom degree ($\upsilon = n$-1) and significant value (α=0.01 or α=0.05) that leads to a large P-value is strong evidence in favor of H_0 (provided that the data are really generated from a random sample).

In order to test the hypothesis, we applied the following steps [8]:
□ State the null and alternative hypotheses.
□ Select the distribution to use.
□ Determine the rejection and non-rejection regions.
□ Calculate the value of the test statistic.
□ Make a decision.

III. Binary Poker Test

Let d be a positive integer such that d≥3, and let k=d. Divide the sequences into k non-overlapping parts each of length d, and let n_i be the observed number of occurrences of the i^{th} type of sequence of length d, 0≤i≤d. The poker test determines whether the sequences of length d each appear approximately the same number of times in S, as would be expected for a random sequence. The expected value of the string which consists of i (1’s) with no consideration to arrangement of (1’s) is [9]:

$$E_i = C_i^d \cdot \frac{1}{2^d} \cdot \frac{L}{d}$$

The statistic used is:

$$\overset{\frown}{\chi}^2 = \sum_{i=0}^{d} \frac{(n_i - C_i^d \cdot \frac{1}{2^d} \cdot \frac{L}{d})^2}{C_i^d \cdot \frac{1}{2^d} \cdot \frac{L}{d}} \qquad \ldots(3)$$

which approximately follows a χ^2 distribution with υ =d degrees of freedom. Note that the poker test is a generalization of the frequency test: setting d=1 in the poker test yields the frequency test.

Example (1): lets have the following sequence with length L=30, m=2.
Let S=100 111 101 110 011 001 101 001 010 000. For d=3, C^d_i=1,3,3,1, then E_i=1.25,3.75,3.75,1.25 respectively. From the sequence we have the following frequencies using relation (3):

n_i	Samples	Freq.
n_0	000	1
n_1	001,010,001	4
n_2	011,101,110	4
n_3	111	1

$$\overset{\frown}{\chi}^2 = \frac{(1-1.25)^2}{1.25} + \frac{(4-3.75)^2}{3.75} + \frac{(4-3.75)^2}{3.75} + \frac{(1-1.25)^2}{1.25}$$

=0.1333, note $\overset{\frown}{\chi}^2$ =0.1333 < $\overset{\frown}{\chi}_0^2$ =7.81, this mean we accept H_0 and reject H_A, this sequence pass the binary poker test.

Example (2): We test two sequences, S_1 which is generated from random generator (random function of the computer) while the sequence S_2 generated from non-random generator (two shift registers with AND boolean function) using binary or classical poker test (CPT) for m=2 and d=3,4,5. Table (1) includes the information of the length of the two sequences with the expected (E_i) and the frequencies (n_i) values. While table (2) shows the

$\widehat{\chi}^{2}$ values and the accepted or rejection decision for the two sequence S_1 and S_2, the information of applying the relation (3) obtained from table (1). $\widehat{\chi}_{0}^{2}$ =7.815, 9.488 and 11.071for υ=d=3,4 and 5 respectively.

Table (1): Information of the two sequences S_1 and S_2.

***L*=1000**																
d		**3**				**4**					**5**					
***E**_i_*		41.67	125	125	41.67	15.63	62.5	93.75	62.5	15.63	6.25	31.25	62.5	62.5	31.25	6.25
***n**_i_*	***S*₁**	42	120	125	46	11	69	92	56	22	7	29	61	64	29	10
	***S*₂**	161	104	57	11	111	56	59	18	6	71	52	48	16	9	4
***L*=5000**																
d		**3**				**4**					**5**					
***E**_i_*		208.3	625	625	208.3	78.13	312.5	468.8	312.5	78.13	31.3	156.3	312.5	312.5	156.3	31.3
***n**_i_*	***S*₁**	203	655	606	202	68	356	447	291	88	39	153	320	303	152	33
	***S*₂**	832	558	225	51	525	391	252	61	21	355	311	193	109	22	10
***L*=10000**																
D		**3**				**4**					**5**					
***E**_i_*		416.7	1250	1250	416.7	156.3	625	937.5	625	156.3	62.5	312.5	625	625	312.5	62.5
***n**_i_*	***S*₁**	394	1238	1290	411	138	615	967	618	162	45	303	651	620	319	62
	***S*₂**	1493	1172	561	22	920	798	567	176	39	597	626	489	225	68	13

Table (2): classical Poker test (CPT) results for two sequences.

Seq.	CPT Values	*L*								
		1000			5000			10000		
	d	3	4	5	3	4	5	3	4	5
S_1	$\widehat{\chi}^{2}$	0.6533	5.3547	2.736	2.3467	11.104	2.672	2.7053	3.51	6.45
	Decision	Accept H_0	Accept H_0	Accept H_0	Accept H_0	Accept H_0	Accept H_0	Accept H_0	Accept H_0	Accept H_0
S_2	$\widehat{\chi}^{2}$	404.86	633.34	739.2	2249.1	2920.3	3815.3	3395.2	4338.1	5098.9
	Decision	Reject H_0	Reject H_0	Reject H_0	Reject H_0	Reject H_0	Reject H_0	Reject H_0	Reject H_0	Reject H_0

Remark (2): From example (1), if we re-calculate the frequency and expected the values for the samples of binary sequence:

n_{ijk}	n_{000}	n_{001}	n_{010}	n_{011}	n_{100}	n_{101}	n_{110}	n_{111}
Samples	000	001	010	011	100	101	110	111
Freq.	1	2	1	1	1	2	1	1

These samples have the same expected value s.t. E_i=E=(1/2^3)*(L/d) =(1/2^3)*(30/3)=1.25.
So relation (3) can be extended to following relation for d=3:

$$\widehat{\chi}^{2} = \sum_{i=0}^{1}\sum_{j=0}^{1}\sum_{k=0}^{1}\frac{(n_{ijk} - E)^{2}}{E} \qquad \ldots(4)$$

$$\widehat{\chi}^{2} = \frac{(1-1.25)^{2}}{1.25} + \frac{(2-1.25)^{2}}{1.25} + \cdots + \frac{(1-1.25)^{2}}{1.25} = 1.2,$$

Notes the closed results for example (1) in relations (3) and (4).

In the next proposition we will prove that if the binary sequence S passes the poker test, $\widehat{\chi}^{2} \le \widehat{\chi}_{0}^{2}$, in relation (3), then it will satisfies poker test for relation (4) and vice versa.

Proposition (1): The sequence S passes binary poker test for relation (3) if and only if it passes binary poker test for relation (4).

Proof:

1 →2: this mean $n_i \approx E_i = C_i p$, where p=(1/2d)(L/d). For simplicity let's choose d=2, then:
n_0=n_{00}, n_1=n_{01}+n_{10}, n_2=n_{11} and C_0=C_{00}=1, C_1=C_{01}+C_{10}=1+1=2, C_2=C_{11}=1.

$$\hat{\chi}^2 = \sum_{i=0}^{2} \frac{(n_i - C_i \cdot p)^2}{C_i \cdot p} = \frac{(n_0 - C_0 \cdot p)^2}{C_0 \cdot p} + \frac{(n_1 - C_1 \cdot p)^2}{C_1 \cdot p} + \frac{(n_2 - C_2 \cdot p)^2}{C_2 \cdot p}$$

notice

$$\frac{(n_0 - C_0 \cdot p)^2}{C_0 \cdot p} = \frac{(n_{00} - C_{00} \cdot p)^2}{C_{00} \cdot p} = \frac{(n_{00} - p)^2}{p} \qquad \ldots(a)$$

and

$$\frac{(n_2 - C_2 \cdot p)^2}{C_2 \cdot p} = \frac{(n_{11} - C_{11} \cdot p)^2}{C_{11} \cdot p} = \frac{(n_{11} - p)^2}{p} \qquad \ldots(b)$$

Now

$$\frac{(n_1 - C_1 \cdot p)^2}{C_1 \cdot p} = \frac{(n_{01} + n_{10} - (C_{01} + C_{10}) \cdot p)^2}{(C_{01} + C_{10}) \cdot p} = \frac{(n_{01} + n_{10} - 2p)^2}{2p}$$

$$= \frac{(n_{01} - p)^2 + (n_{10} - p)^2}{2p} + \left[\frac{1}{p} \cdot (n_{01} - p) \cdot (n_{10} - p) \right] \qquad \ldots(c)$$

Since $n_1 \approx E_1 = C_1 p$, then

$$\left| n_{01} + n_{10} \right| \approx \left| (C_{01} + C_{10}) p \right| = \left| p + p \right|$$

$$\left| n_{01} - p \right| \approx \left| n_{10} - p \right|$$

$$(n_{01} - p) \cdot (n_{10} - p) \approx \frac{1}{2} \left[(n_{01} - p)^2 + (n_{10} - p)^2 \right] \qquad \ldots(d)$$

Substitute (d) in (c), obtain:

$$\frac{(n_1 - C_1 \cdot p)^2}{C_1 \cdot p} \approx \frac{(n_{01} - p)^2 + (n_{10} - p)^2}{p} \qquad \ldots(e)$$

From (a), (b) and (e) we obtain:

$$\hat{\chi}^2 = \sum_{i=0}^{d} \frac{(n_i - C_i^d \cdot \frac{1}{2^d} \cdot \frac{L}{d})^2}{C_i^d \cdot \frac{1}{2^d} \cdot \frac{L}{d}} \approx \sum_{i=0}^{1} \sum_{j=0}^{1} \sum_{k=0}^{1} \frac{(n_{ijk} - E)^2}{E}$$

In the same way we can prove the other side. □

Proposition (1) can be proved for any *d*, and now its no matter if we applied relation (3) or (4) to decide whether that *S* will passes binary poker test.

IV. Generalizing the Binary *d*-Tuple Digital Poker Test

In this section we will generalize the binary *d*-tuple sequences to calculate the digital poker test for digital (*m*-) sequences.

Let *S* be the digital *m*-sequence generated from digital generator with length *L*, where $s_i \in S$, and $0 \le s_i \le m-1$, $i=0,1,2,\ldots,L-1$.

Now we will propose a new poker test, we called it Grouped poker test (GPT), s.t. we will divide the samples of d-tuple subsequence to m disjoint sets of samples. Let $A=\{0,1,\ldots,m^d-1\}$ be the set of all different integer sum of the all probable *d*-tuple samples of the tested digital sequence *S*. we partitioned the set *A* into equal order subsets $A_0=\{x_0=0,1,\ldots,x_1-1\}$, $A_1=\{x_1,x_1+1,\ldots,x_2-1\}$,…, $A_{m-1}=\{x_{m-1},x_{m-1}+1,\ldots,x_m=m^d-1\}$, s.t. $x_i=i(m^{d-1})=ih$, where $i=0,1,\ldots,m-1$. Notice that:

- $A = \bigcup_{i=0}^{m-1} A_i$.
- $A_k \cap A_j$, where $k \neq j$, $\forall\ 0 \leq k\ , j \leq m$-1 (any two different subsets of A are disjoint).
- $|A|=m^d$ and $|A_i|=h=m^d$-1, $\forall\ 0 \leq i \leq m$-1.

Example (3): For m=3 and d=3, we have the following samples:
(0,0,0),(0,0,1),(0,0,2),(0,1,0),(0,1,1),(0,1,2),(0,2,0),(0,2,1),(0,2,2),
(1,0,0),(1,0,1),(1,0,2),(1,1,0),(1,1,1),(1,1,2),(1,2,0),(1,2,1),(1,2,2),
(2,0,0),(2,0,1),(2,0,2),(2,1,0),(2,1,1),(2,1,2),(2,2,0),(2,2,1),(2,2,2),

This means we have $m^d=3^3$=27 samples. We have A={0,1,…,26}, A_0={0,1,…,8}, A_1={9,10,…,17} and A_2={18,19,…,26}. $|A|=3^3$=27, $|A_i|= 3^2=9=h$, x_i=0,9,18, i=0,1,2.

Now let N_i, i=0,1,…,m-1, be the total observed frequency of all elements of the set A_i occurred in digital sequence S. The probability P for each digit 0,...,2^m-1 in the set A is:

$P(s_k)= 1/2^m$, k=0,1,…,m^d-1. …(5)

Since the order of the set A_i, $|A_i|=h=m^d$-1, then the probability for each digit in set A_i is:

$P(s_t)=h(1/2^m)=1/m$, $t=i,i+1,\ldots,ih$. …(6)

Then the expected value (E) is:

$E=(1/m).(L/d)=L/(m.d)$ …(7)

which is equal for all digits s_t.

When applying chi-square test, $\hat{\chi}^2$ will compared with table value $\hat{\chi}_0^2$ with freedom degree $\upsilon = m$.

Its obvious that if the frequency N_i of each distinct digit s_t is approximate to other frequencies, then the digital sequence is satisfies the poker postulate, so must be:
$N_0 \approx N_1 \approx \ldots \approx N_{m-1}$
Then the GPT will be:

$$\hat{\chi}^2 = \sum_{i=0}^{m-1} \frac{(N_i - \frac{L}{m.d})^2}{\frac{L}{m.d}} \qquad \ldots(8)$$

Example (4): Let's call example (1), L=30, m=2, d=3. x_i=0, 4, A_0={0,1,2,3} and A_1={4,5,6,7}, E=0.5*(30/3)=5,

N_i	Samples	Freq.
N_0	000,001,010,100	5
N_1	100,101,110,111	5

$$\hat{\chi}^2 = \sum_{i=0}^{m-1} \frac{(N_i - \frac{L}{m.d})^2}{\frac{L}{m.d}} = \frac{(5-5)^2}{5} + \frac{(5-5)^2}{5} = 0$$

$\hat{\chi}^2$ =0 <7.815, then S passes GPT.

Example (5): let's call example (2), using GPT for m=2 and d=3,4,5. Table (3) includes the new information of the two sequences S_1 and S_2 with the new expected (E_i) and new frequencies (N_i) values. While table(4) shows the $\hat{\chi}^2$ values and the accepted or rejection decision for the sequence S_1 and S_2, the information of applying the relation (8) obtained from table (3). $\hat{\chi}_0^2$ =5.991for $\upsilon=m$=2.

Table (3): Information of the two sequences S_1 and S_2.

L	***d, E***							
1000	***d, E***		***d*=3, *E*=166.667**		***d*=4, *E*=125**		***d*=5, *E*=100**	
	N_i	S_1	169	164	137	113	98	102
		S_2	252	81	179	71	144	56
5000	***d, E***		***d*=3, *E*=833.33**		***d*=4, *E*=625**		***d*=5, *E*=500**	
	N_i	S_1	818	848	640	610	517	483
		S_2	1275	391	973	227	772	228
10000	***d, E***		***d*=3, *E*=1666.7**		***d*=4, *E*=1250**		***d*=5, *E*=1000**	
	N_i	S_1	1627	1706	1234	1266	985	1015
		S_2	2470	863	1848	652	1470	530

Table (4): GPT results for two sequences for *m*=2.

Seq.	GPT Values	*L*								
		1000			5000			10000		
	d	3	4	5	3	4	5	3	4	5
S_1	$\hat{\chi}^{2}$	0.0753	2.304	0.08	0.5403	0.72	1.156	1.8723	0.4096	0.45
	Decision	Accept H_0	Accept H_0	Accept H_0	Accept H_0	Accept H_0	Accept H_0	Accept H_0	Accept H_0	Accept H_0
S_2	$\hat{\chi}^{2}$	87.7233	46.656	38.72	468.87	387.53	295.94	774.73	572.17	441.8
	Decision	Reject H_0	Reject H_0	Reject H_0	Reject H_0	Reject H_0	Reject H_0	Reject H_0	Reject H_0	Reject H_0

Example (6): Now lets choose *m*=3,4 and 5, and *d*=3,4,5 for the two sequences S_1 and S_2 with length *L*=5000 digits. Table (5) illustrates the GPT results using chi-square test for the two chosen sequences. $\hat{\chi}^{2}_{0}$ =7.815, 9.488 and 11.071for υ=*m*=3,4 and 5 respectively.

Table (5): GPT results for two sequences for *m*=3,4,5 and *d*=3,4,5.

Seq.	**GPT Values**	***L*=5000**								
	m	3			4			5		
	d	3	4	5	3	4	5	3	4	5
S_1	$\hat{\chi}^{2}$	1.8303	4.5328	2.342	0.4299	9.3584	9.944	2.9607	0.664	4.37
	Decision	Accept H_0	Accept H_0	Accept H_0	Accept H_0	Accept H_0	Accept H_0	Accept H_0	Accept H_0	Accept H_0
S_2	$\hat{\chi}^{2}$	367.15	279.42	205.92	619.09	434.40	360.44	265.23	188.35	162.73
	Decision	Reject H_0	Reject H_0	Reject H_0	Reject H_0	Reject H_0	Reject H_0	Reject H_0	Reject H_0	Reject H_0

V. Conclusions

1. As known, the proof of randomness for any sequence is probabilistic, so it's preferable to choose *L* as long as possible and choose many examples of sequences to be tested by poker test to obtain accurate decision.
2. In order to obtain accurate randomness decision for any digital sequence, we have to choose different values for *d*.
3. From Ali et al. [3] Ibraheem [7] with this work we will obtain five digital randomness test for digital sequences so we suggest a digital statistical randomness package using five digital tests; frequency, run, autocorrelation, serial and poker tests.

References

[1]. Abdel-Rehim W. M. F., Ismail I. A. and Morsy E., "***Testing Randomness: Poker Test with Hands of Three Numbers***", Journal of Computer Science 8 (8): 1353-1357, ISSN 1549-3636, 2012.

[2]. Stewart, W. J., "***Probability, Markov Chains, Queues, and Simulation: The Mathematical Basis of Performance Modeling***", 1st Edn., Princeton University Press, Princeton, ISBN-10: 0691140626, pp: 758, 2009.

[3]. Ali F. H, Mohammed, S. A. and Shammran, M. A., "***Generalize the Randomness Tests to Test the Digital Sequences Produced from Digital Stream Cipher Systems***", Iraqi Journal for Science, Baghdad University, College of Science, 2009.

[4]. Beker, H. and Piper, F., "***Cipher Systems: The Protection of Communications***", John Wiley & Sons, New York, 1982.

[5]. Gustafson, H., Dawson, E., Nielsen, L. and Caelli, W., "***A Computer Package for Measuring the Strength of Encryption Algorithms***", Computers & Security, 13, 687–697, 1994.

[6]. Brassard, G., "***Modern Cryptology: A Tutorial***", LNCS 325, Springer-Verlag, New York, 1988.

[7]. Ibraheem S. K., "***Serial Test Extension and Generalization to Test the Digital Sequences***", Iraq Academic Scientific Journals, ISSN: 1814635X, Volume: 25 Issue:4 Pages: 83-96, 2014.

[8]. http://www.iiserpune.ac.in/.../20110404randomno02.pdf, 2011.

[9]. Al-Shammari, A. G., "***Mathematical Modeling and Analysis Technique of Stream Cipher Cryptosystems***", Ph. D. Thesis, University of Technology, Applied Sciences, 2009.