جامعة المستقبل
AL MUSTAQBAL UNIVERSITY

College of science
Department of Cyber Security

Lecture 6:

Measure of Randomness

Stage :2

م.م مصطفى امير صبري

• **Definition:**

Run: sequence of identical bits (**0** or **1**).

EX. 01110000111

Runs are 0, 111, and 0000,111

Gap: run of zeroes. 1000011 contain of gap 0000

Block: run of ones. 1111001110 contain block 1111,111

• **Five Basic Tests**

Let s = s_0, s_1, s2, s_{n-1} be a binary sequence of length n. This subsection presents five statistical tests that are commonly used for determining whether the binary sequence **S** possesses some specific characteristics that a truly random sequence would be likely to exhibit. It is emphasized again that the outcome of each test is not definite, but rather probabilistic. If a sequence passes all five tests, there is no guarantee that it was indeed produced by a random bit generator.

(i) Frequency test (monobit test)

The purpose of this test is to determine whether the number of 0's and 1's in s are approximately the same, as would be expected for a random sequence. Let n0, n1 denote the number of 0's and 1's in s, respectively. The statistic used is

$$X_1 = (n0 - n1)_2 / n$$

Which approximately follows a χ_2 distribution with **1** degree of freedom if n ≥ 10.

(ii) Serial test (two-bit test)

The purpose of this test is to determine whether the number of occurrences of 00, 01, 10, and 11 as subsequences of s are approximately the same, as would be expected for a random sequence. Let n0, n1 denote the number of 0's and 1's in s, respectively, and let n00, n01, n10, n11 denote the number of occurrences of 00, 01, 10, 11 in s, respectively. Note that n00 + n01 + n10 + n11 = (n − 1) since the subsequences are allowed to overlap. The statistic used is

$$X_2 = \frac{4}{n-1}\left(n_{00}^2 + n_{01}^2 + n_{10}^2 + n_{11}^2\right) - \frac{2}{n}\left(n_0^2 + n_1^2\right) + 1$$

Which approximately follows a χ^2 distribution with **2** degrees of freedom if n ≥ 21.

Note: if block (000) is a (00) and (00)

(iii) Poker test

Let m be a positive integer such that [n /m] ≥ 5.(2^m), and let k = [n / m]. Divide the sequence **S** into k non-overlapping parts each of length m, and let n_i be the number of occurrences of the i^{th} type of sequence of length m, 1 ≤ i ≤ 2^m. The poker test determines whether the sequences of length m each appear approximately the same number of times in s, as would be expected for a random sequence. The statistic used is

$$X_3 = \frac{2^m}{k}\left(\sum_{i=1}^{2^m} n_i^2\right) - k$$

Which approximately follows a χ^2 distribution with **$2^m - 1$** degrees of freedom. Note that the poker test is a generalization of the frequency test:

setting m = 1 in the poker test yields the frequency test.

Note: must divides the s to blocks length each is (m)

(iv)Runs test

The purpose of the runs test is to determine whether the number of runs (of either zeros or ones) of various lengths in the sequence s is as expected for a random sequence. The expected number of gaps (or blocks) of length i in a random sequence of length n is $e_i = (n-i+3)/2^{i+2}$. **Let k** be equal to the largest integer i for which **$e_i \geq 5$**. Let B_i, G_i be the number of **blocks and gaps**, respectively, of length i in s for each i, $1 \leq i \leq k$. The statistic used is

$$\chi^2 = \frac{(B_i - e_i)^2}{e_i} + \frac{(G_i - e_i)^2}{e_i}$$

Which approximately follows a χ^2 distribution with **2k − 2** degrees of freedom

Note: (K) depend on the higher (i)

(v) Autocorrelation test

The purpose of this test is to check for correlations between the sequence s and (non-cyclic) shifted versions of it. Let d be a fixed integer, $1 \leq d \leq [n/2]$. The number of bits in s not equal to their d-shifts is

$$\sum_{i=1}^{n-d} s_i + s_{i+d}$$

A(d) = , where + denotes the XOR operator. The statistic used is

$$A_{\%} = \#\$\#\$!'' !\$ ''!! \; !\#''''!\sqrt{''!!}$$

Which approximately follows an N(0,1) distribution if n − d ≥ 10. Since small values of A(d) are as unexpected as large values of A(d), a two-sided test should be used.

Example :(*basic statistical tests*) Consider the (non-random) sequence s of length n =160 obtained by replicating the following sequence four times:

11100 01100 01000 10100 11101 11100 10010 01001

(i) (***Frequency test***) n0 = 84, n1 = 76 and the value of the statistic X_1 is 0.4.

(ii) (***Serial test***) n00 = 44, n01 = 40, n10 = 40, n11 = 35, and the value of the statistic X_2 is 0.6252.

(iii) (***poker test***) Here m = 3 and k = 53. The blocks 000, 001, 010, 011, 100, 101, 110,111 appear 5, 10, 6, 4, 12, 3, 6, and 7 times, respectively, and the value of the statistic X_3 is 9.6415.

(iv) (***runs test***) Here e1 = 20.25, e2 = 10.0625, e3 = 5, and k = 3.

There are 25, 4, 5 **blocks** of lengths 1, 2, 3, respectively, and 8, 20, 12 **gaps** of lengths 1, 2, 3, respectively .The value of the statistic X_4 is 31.7913.

(v) (***autocorrelation test***) If d = 8, then A(8) = 100. The value of the statistic X5 is 3.8933.

For a significance level of α = 0.05, the threshold values for X1, X2, X3, X4, and X5 are 3.8415, 5.9915, 14.0671, 9.4877, and 1.96,. Hence, the given sequence **s** passes the **frequency**, **serial**, and **poker** tests, but fails the **runs** and **autocorrelation** tests.