Lecture Two

Randomness

6. Standard Statistical Randomness Tests

Let $S=s_0,s_1,s_2\ldots,s_{n-1}$ 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.

It is important to mention that the frequency, run and auto correlation test are called the **Main Binary Standard Randomness Tests** (MBSRT).

Before we shed light on the five basic tests, we have to construct the law of Chi-square which we really used is:

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 outcome falls into category i, assume that L independent observation is made, and let Q_i be the number of observation falling into category i, in order to test the hypothesis the quantity T is compared:

$$T=\sum_{i=1}^{k}\frac{(Q_i-Lp_i)^2}{Lp_i} \qquad \ldots(6.1)$$

If the hypothesis is true, the value T is distribute according to the χ^2 distribution with υ=k-1 degree of freedom, the hypothesis is rejected if Q_i and Lp_i are too different, i.e. if T is too big, that means we set some pass mark x_0 and reject the hypothesis if T greater than x_0, α will be the significance level of the test, of course $E_i=Lp_i$ s.t. E_i is the expected value of occurrence of outcome i.

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 n_0, n_1 denote the observed number of 0′s and 1′s in S, respectively. The expected value is n/2.

The statistic used is:

$$X_1=\sum_{i=0}^{1}\frac{(n_i-n/2)^2}{n/2}=\frac{(n_0-n_1)^2}{n} \qquad \ldots(6.2)$$

which approximately follows a χ^2 distribution with 1 degree of freedom.

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 n_0, n_1 denote the number of 0′s and 1′s in s, respectively, and let n_{00}, n_{01}, n_{10}, n_{11} denote the observed number of occurrences of 00,01,10,11 in s, respectively. Note that $n_{00}+n_{01}+n_{10}+n_{11}=n-1$ since the subsequences are allowed to overlap. The expected value is (n-1)/4.

The statistic used is:

$$X_2=\sum_{i=0}^{1}\sum_{j=0}^{1}\frac{(n_{ij}-(n-1)/4)^2}{(n-1)/4} \qquad \ldots(6.3)$$

which approximately follows a χ^2 distribution with 3 degrees of freedom.

III. Poker test

Let m be a positive integer such that m≥3, and let k=m. Divide the sequences into k non-overlapping parts each of length m, and let n_i be the observed number of occurrences of the i^{th} type of sequence of length m, $0 \le i \le m$. 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 expected value of the string which consists of i (1's) with no consideration to arrangement of (1's) is:

$$E_i = C_i^m \cdot \frac{1}{2^m} \cdot \frac{n}{m}$$

The statistic used is:

$$X_3 = \sum_{i=0}^{m} \frac{(n_i - C_i^m \cdot \frac{1}{2^m} \cdot \frac{n}{m})^2}{C_i^m \cdot \frac{1}{2^m} \cdot \frac{n}{m}} \qquad \ldots(6.4)$$

which approximately follows a χ^2 distribution with $\upsilon = m$ 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.

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:

$$E_i = \frac{n - i + 3}{2^{i+2}}$$

Let k be equal to the largest gap (block). Let B_i, G_i, be the observed number of blocks and gaps, respectively, of length i in S for each i, $1 \le i \le k$. The statistic used is:

$$X_4 = \sum_{i=1}^{k} \frac{(G_i - E_i)^2}{E_i} + \frac{(B_i - E_i)^2}{E_i} \qquad \ldots(6.5)$$

which approximately follows a χ^2 distribution with 2k-2 as a degrees of freedom.

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 τ be a fixed integer, $1\leq\tau\leq n/2$. The expect value $E=(n-\tau)/2$. The number of bits in S not equal to their τ-shifts is:

$S^{\tau}=\left\{s_i^{\tau}=s_i\oplus s_{i+\tau}\right\}_{i=1}^{n-\tau}$,

where $\oplus$ denotes the XOR operator.

Let $n_0(\tau)$ and $n_1(\tau)$ denote the observed number of 0′s and 1′s in $A(\tau)$, respectively. The statistic used is:

$$X_5=\frac{\left(n_0(\tau)-\frac{n-\tau}{2}\right)^2}{\frac{n-\tau}{2}}+\frac{\left(n_1(\tau)-\frac{n-\tau}{2}\right)^2}{\frac{n-\tau}{2}}=\frac{(n_0(\tau)-n_1(\tau))^2}{n-\tau} \qquad \ldots(6.6)$$

which approximately follows a χ^2 distribution with $\upsilon=1$ degrees of freedom.

<u>Example(6.2):</u> (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:** $n_0=84$, $n_1=76$, and the value of the statistic X_1 is 0.4.

II. **Serial test:** $n_{00}=44$, $n_{01}=40$, $n_{10}=40$, $n_{11}=35$, expected value is E=39.75, and the value of the statistic X_2 is 1.025.

III. **Poker test:** Here m=3. The blocks #“000”=5, #(“001”+“010”+ “001”)=28, #(“011”+“110”+“101”)=12, #“111”=7, expected values are $E_0=6.667$, $E_1=20.001$, $E_2=20.001$, $E_3=6.667$ and the value of the statistic X_3 is 6.834.

IV. **Runs test:** Here E_1=20.25, E_2=10.0625, E_3=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 τ=3, $n_0(3)$=80 and $n_1(3)$=77. The value of the statistic X_5 is 0.115.

For a significance level of α=0.05, the threshold values for X_1, X_2, X_3, X_4, and X_5 are 3.8415, 7.8415, 7.8415, 31.787, and 0.115, respectively (see Tables 4.1 and 4.2). Hence, the given sequence S passes the frequency, serial, poker and autocorrelation tests, but fails the runs test.