Cryptography And Cryptanalysis

Ph. D. Course/ 2019-2020

Introduced By

Dr. Faez Hassan Ali

Lecture Two-1

Randomness

Background

Definition: A **random bit generator** is a device or algorithm which outputs a sequence of statistically independent and unbiased binary digits.

Remark: A random bit generator can be used to generate (uniformly distributed) random numbers. For example, a random integer in the interval [0,n] can be obtained by generating a random bit sequence of length $\log_2 |n+1|$ bits, and converting it to an integer.

Definition: A **Pseudo Random Bit Generator** (PRBG) is a deterministic algorithm which, given a truly random binary sequence 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**.

A minimum security requirement for a pseudorandom bit generator is that the length k of the random seed should be sufficiently large so that a search over 2^k elements (the total number of possible seeds) is infeasible for the adversary.

Random and Pseudorandom Bit Generation

Linear congruence: The following algorithm which is called **Linear Congruential Generation** (LCG) generates a sequence S of random numbers $S=\{x_1,x_2,\ldots,x_k\}$.

INPUT : x_0,a,b,m,k
PROCESS : For j := 1 to k
$x_j := a.x_{j-1} + b \pmod m$
EndFor {j}
OUPUT : the sequence S
END.

Example :

Let $x_0=5$, a=11, b=73, m=1399 and k=5, then
$x_1=11(5)+73 \pmod{1399} = 128$
$x_2=11(128)+73 \pmod{1399} = 82$
$x_3=11(82)+73 \pmod{1399} = 975$
$x_4=11(975)+73 \pmod{1399} = 1005$
$x_5=11(1005)+73 \pmod{1399} = 1335$
S = {128,82,975,1005,1335}.

HW: Use the **Quadratic Congruential Generation (QCG)** to generate S with length L.
$x_j := a.x^2_{j-1} + b.\ x_{j-2} + c \pmod m$

Random and Pseudorandom Bit Generation

Requirements for Random Number Generators

Ideally a pseudo-random number generator would produce a stream of numbers that:

- Are uniformly distributed.
- Are uncorrelated.
- Never repeats itself.
- Satisfy any statistical test for randomness.
- Are reproduceable (for debugging purposes).
- Are portable (the same on any computer).
- Can be changed by adjusting an initial “seed” value.
- Can easily be split into many independent subsequences.
- Can be generated rapidly using limited computer memory.

In practice it is impossible to satisfy all these requirements exactly.

Statistical Tests

The χ^2 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.

Table gives some percentiles of the χ^2 distribution for various degrees of freedom degree.

	α					
v	0.100	0.050	0.025	0.010	0.005	0.001
1	2.7055	3.8415	5.0239	6.6349	7.8794	10.8276
2	4.6052	5.9915	7.3778	9.2103	10.5966	13.8155
3	6.2514	7.8147	9.3484	11.3449	12.8382	16.2662
4	7.7794	9.4877	11.1433	13.2767	14.8603	18.4668
5	9.2364	11.0705	12.8325	15.0863	16.7496	20.5150
6	10.6446	12.5916	14.4494	16.8119	18.5476	22.4577
7	12.0170	14.0671	16.0128	18.4753	20.2777	24.3219
8	13.3616	15.5073	17.5345	20.0902	21.9550	26.1245
9	14.6837	16.9190	19.0228	21.6660	23.5894	27.8772
10	15.9872	18.3070	20.4832	23.2093	25.1882	29.5883
11	17.2750	19.6751	21.9200	24.7250	26.7568	31.2641
12	18.5493	21.0261	23.3367	26.2170	28.2995	32.9095
13	19.8119	22.3620	24.7356	27.6882	29.8195	34.5282
14	21.0641	23.6848	26.1189	29.1412	31.3193	36.1233
15	22.3071	24.9958	27.4884	30.5779	32.8013	37.6973
16	23.5418	26.2962	28.8454	31.9999	34.2672	39.2524
17	24.7690	27.5871	30.1910	33.4087	35.7185	40.7902
18	25.9894	28.8693	31.5264	34.8053	37.1565	42.3124
19	27.2036	30.1435	32.8523	36.1909	38.5823	43.8202
20	28.4120	31.4104	34.1696	37.5662	39.9968	45.3147
21	29.6151	32.6706	35.4789	38.9322	41.4011	46.7970
22	30.8133	33.9244	36.7807	40.2894	42.7957	48.2679
23	32.0069	35.1725	38.0756	41.6384	44.1813	49.7282
24	33.1962	36.4150	39.3641	42.9798	45.5585	51.1786
25	34.3816	37.6525	40.6465	44.3141	46.9279	52.6197
26	35.5632	38.8851	41.9232	45.6417	48.2899	54.0520
27	36.7412	40.1133	43.1945	46.9629	49.6449	55.4760
28	37.9159	41.3371	44.4608	48.2782	50.9934	56.8923
29	39.0875	42.5570	45.7223	49.5879	52.3356	58.3012
30	40.2560	43.7730	46.9792	50.8922	53.6720	59.7031
31	41.4217	44.9853	48.2319	52.1914	55.0027	61.0983
63	77.7454	82.5287	86.8296	92.0100	95.6493	103.4424
127	147.8048	154.3015	160.0858	166.9874	171.7961	181.9930
255	284.3359	293.2478	301.1250	310.4574	316.9194	330.5197
511	552.3739	564.6961	575.5298	588.2978	597.0978	615.5149
1023	1081.3794	1098.5208	1113.5334	1131.1587	1143.2653	1168.4973

Statistical Tests

Hypothesis Testing

- **Definition:** A **statistical hypothesis**, denoted H_0, is an assertion about a distribution of one or more random variables.
- A test of a statistical hypothesis is a procedure, based upon the observed values of the random variables, that leads to the acceptance or rejection of the hypothesis H_0.
- **Definition:** The significance level of the test of a statistical hypothesis H_0 is the probability of rejecting H_0 when it is true.
- H_0 : a given binary sequence was produced by a random bit generator.
- a significance level is α=0.05 might be employed in practice.
- Statistics are generally chosen so that they can be efficiently computed, and so that they (approximately) follow a χ^2 distribution. The value of the statistic for the sample output sequence is compared with the value expected for a random sequence.
- If the value X_s of the statistic for the sample output sequence satisfies $X_s > x_\alpha$, then the sequence fails the test; otherwise, it passes the test.

Golomb's Concept of Randomness

Definition: Let $S=s_0,s_1,s_2\ldots$ be an infinite sequence. The subsequence consisting of the first n terms of S is denoted by $S^n=s_0,s_1,s_2,\ldots,s_{n-1}$.

Definition: The sequence $S=s_0,s_1,s_2\ldots$ is said to be **n-periodic** if $s_i=s_{i+n}$ for all $i\geq 0$. The sequence s is **periodic** if it is n-periodic for some positive integer n. The period of a periodic sequence S is the smallest positive integer n for which s is n-periodic.

Definition: Let S be a sequence. A **run** of S is a subsequence of S consisting of consecutive 0's or consecutive 1's which is neither preceded nor succeeded by the same symbol. A run of 0's is called a **gap**, while a run of 1's is called a **block**.

Definition: Let $S= s_0,s_1,s_2\ldots$ be a periodic sequence of period n. The **autocorrelation function** of S is the integer-valued function C(t) defined as:

$n.C(\tau) = \sum_{i=0}^{n-1}(2s_i-1)\cdot(2s_{i+t}-1)$, $0\leq\tau\leq n-1$.

The autocorrelation function $C(\tau)$ measures the amount of similarity between the sequence S and a shift of S by τ positions.

Golomb's Concept of Randomness

Definition: Let S be a periodic sequence of period n. Golomb's randomness postulates are the following:

- **R1**: In the cycle S^n of S, the number of 1's differs from the number of 0's by at most 1.
- **R2**: In the cycle S^n at least half the runs have length 1, at least one-fourth have length 2, at least one-eighth have length 3, etc., as long as the number of runs so indicated exceeds 1. Moreover, for each of these lengths, there are (almost) equally many gaps and blocks.
- **R3**: The autocorrelation function $C(\tau)$ is two-valued. That is for some integer K:

$$n.C(\tau) = \sum_{i=0}^{n-1} (2s_i - 1) \cdot (2s_{i+t} - 1) = \begin{cases} n, & t = 0 \\ K, & 1 \le t \le n-1 \end{cases}$$

Definition: A binary sequence which satisfies Golomb's randomness postulates is called a **pseudo-noise sequence** or a **pn-sequence**.

Example: (pn-sequence) Consider the periodic sequence S of period n=15 with cycle S^{15}=011001000111101

The following shows that the sequence S satisfies Golomb's randomness postulates.

R1: The number of 0's in s^{15} is 7, while the number of 1's is 8.

R2: S^{15} has 8 runs. There are 4 runs of length 1 (2 gaps and 2 blocks), 2 runs of length 2 (1 gap and 1 block), 1 run of length 3 (1 gap), and 1 run of length 4 (1 block).

R3: The autocorrelation function $C(\tau)$ takes on two values :C(0)=1 and $C(\tau)=1/15$ for $1 \le \tau \le 14$. Hence, S is a pn-sequence.

Standard Statistical Randomness Tests

Let $S=s_0,s_1,s_2\ldots,s_{n-1}$ be a binary sequence of length n. It is important to mention that the frequency, run and auto correlation test are called the **Main Binary Standard Randomness Tests** (MBSRT).

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:

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

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 T_0 and reject the hypothesis if T greater than T_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.

Standard Statistical Randomness Tests

Frequency test

this test is determine whether the number of 0's and 1's in S are approximately the same, 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:

$$X_1 = \sum_{i=0}^{1} \frac{(n_i - n/2)^2}{n/2} = \frac{(n_0 - n_1)^2}{n} \qquad \text{...(2) HW: Prove the equality?}$$

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

Serial test (two-bit test)

this test is determine whether the number of occurrences of 00, 01, 10, and 11 as subsequences of s are approximately the same. 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:

$$X_2 = \sum_{i=0}^{1}\sum_{j=0}^{1} \frac{(n_{ij} - (n-1)/4)^2}{(n-1)/4} \qquad \text{...(3)}$$

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

Poker test

Let m be a positive integer such that $m \geq 3$, 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, $0 \leq i \leq m$. This test determines whether the sequences of length m each appear approximately the same number of times in s. The expected value of the string which consists of i (1's) :

$E_i = C_i^m \cdot \frac{1}{2^m} \cdot \frac{n}{m}$ The statistic used is:

$$X_3 = \sum_{i=0}^{m} \frac{(n_i - C_i^m \cdot \frac{1}{2^m} \cdot \frac{n}{m})^2}{C_i^m \cdot \frac{1}{2^m} \cdot \frac{n}{m}} \qquad \text{...(4)}$$

which approximately follows a χ^2 distribution with υ=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.

Standard Statistical Randomness Tests

Runs test

This is 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:

$X_4 = \sum_{i=1}^{k} \frac{(G_i - E_i)^2}{E_i} + \frac{(B_i - E_i)^2}{E_i}$ …(5)

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

Autocorrelation test

this test is check for correlations between the sequence s and (non-cyclic) shifted versions of it. Let τ be a fixed integer, $1 \le \tau \le n/2$. 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:

$X_5 = \frac{\left(n_0(\tau) - \frac{n-\tau}{2}\right)^2}{\frac{n-\tau}{2}} + \frac{\left(n_1(\tau) - \frac{n-\tau}{2}\right)^2}{\frac{n-\tau}{2}} = \frac{(n_0(\tau) - n_1(\tau))^2}{n - \tau}$ …(6)

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

Standard Statistical Randomness Tests

Example: (**basic statistical tests**)

Consider the sequence S of length n = 160 obtained by replicating the sequence four times: 11100 01100 01000 10100 11101 11100 10010 01001.

- **Frequency test:** n_0=84, n_1=76, E=80 and X_1 is 0.4.
- **Serial test:** n_{00}=44, n_{01}=40, n_{10}=40, n_{11}=35, E=39.75, and X_2 is 1.025.
- **Poker test:** Here m=3. The blocks #"000"=5, #("001"+"010"+"001")=28, #("011"+"110"+"101")=12, #"111"=7, E_0=6.667, E_1=20.001, E_2=20.001, E_3=6.667 and X_3 is 6.834.
- **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. X_4 is 31.7913.
- **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. Hence, the sequence S passes the frequency, serial, poker and autocorrelation tests, but fails the runs test.