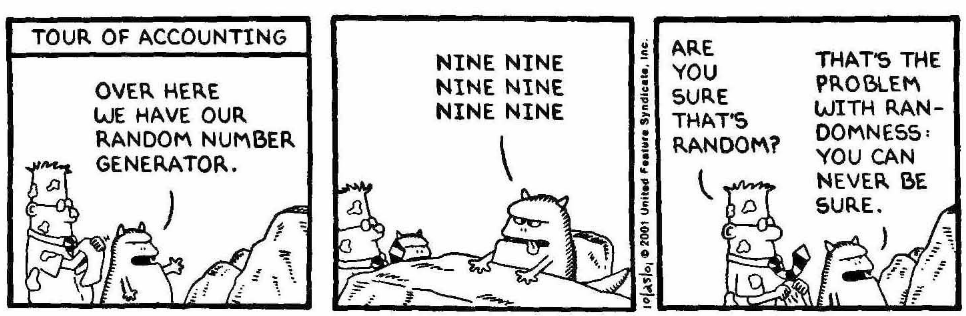
Mobile Systems Security

Randomness tests

Introduction (1)

[Definition] **Random**, adj: lacking a pattern
(Longman concise English dictionary)



Introduction (2)

- What are the requirements for a RNG-Random Number Generator?
  1. There is no sequence of random numbers showing statistical weakness.
  2. The knowledge of a random subsequence does not allow to practically compute predecessors or sucessors.
  3. It is not practically feasible to compute preceding random numbers from the generator internal state.
  4. It is not practically feasible to compute future random numbers from the generator internal state.

Introduction (3)

- Every PRNG has its regularities.
- It is impossible to give a mathematica proof that a generator is indeed a random bit generator.
- Quality of a PRNG may be measured by taking a sample output sequence and subjecting the sample to several statistic tests. Test results are
  - Fail
  - Not rejected (Note: it is not “accepted”!)
- Random numbers [0,n] can be generated from random bits of length $\lfloor \lg n \rfloor + 1$.

Introduction (4)

- In this course we divide randomness identification in 3 parts:
  A. **Golomb** randomness postulates are one of the first attempts to establish necessary conditions for a periodic pseudorandom sequence to look random.
  B. **FIP140-2** specifies four statistical tests for randomness.
  C. Test suites provide the degree of quality for a RNG. The most-widely used test suite are
     - **NIIST statistical test suite**
     - **Diehard**
- Analysis of number sequences work mainly around two statistical distributions, normal and χ^2.

Statistics background (1)

[Definition] A **statistical hypothesis** is an assertion about a distribution of variables.

[Definition] The **null hypothesis**, H_0, is a hypothesis set up to be nullified or refuted in order to support an alternative hypothesis.

In this course we use H_0 as "a sequence is produced by a random generator".

Statistics background (2)

[Definition] A **statistical hypothesis test** is a method of making statistical decisions from and about experimental data.

[Definition] The **significance level** of a statistical hypothesis test, α, is the probalility of rejecting the statistical hypothesis when it is true.

For H_0 “a sequence is produced by a random generator”:

– If α is too high, the test may reject sequences, which were produced by a random generator.

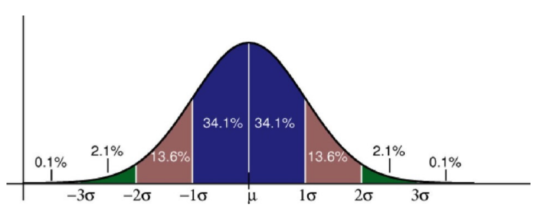
– If α is too low, the test may accept sequences, which were not produced by a random generator.

Probabilistic distributions (1)

A. Normal distribution

[Definition] A continuous random variable X has a **normal** distribution with mean μ and variance σ^2, $N(\mu,\sigma^2)$, if its probability density function is equal to

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$



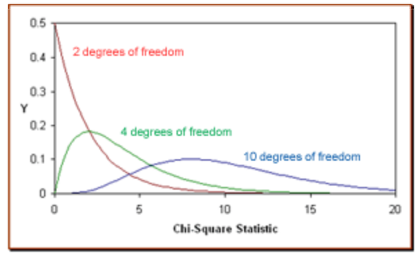
- Psychological and physical measurements follow a normal distribution, because many small and independent effects additively contribute to each observation.

Probabilistic distributions (2)

B. Chi-square distribution

[Definition] Let X_i, $1 \le i \le df$ be N(1,0) independent variables. Then, $X = \sum (X_i)^2$ has a χ^2 distribution and its probability density function is equal to

$$p(x) = \begin{cases} \dfrac{1}{\Gamma(\frac{df}{2}) 2^{\frac{df}{2}}} x^{(\frac{df}{2}-1)} e^{\frac{-x}{2}} & , x \ge 0 \\ 0 & , x < 0 \end{cases}$$

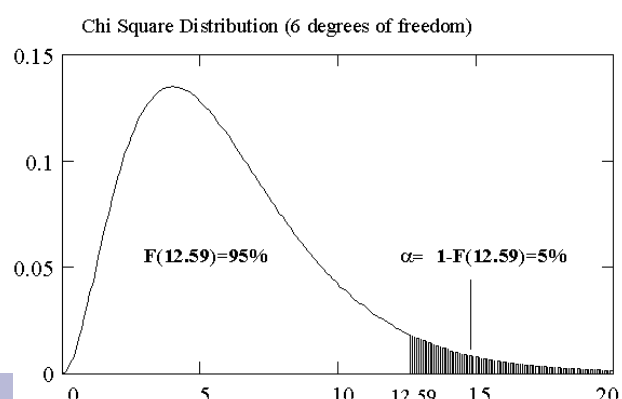


- Γ denotes gamma function (extention of factorial function to real numbers).

Note: The number of samples must be sufficiently large

Probabilistic distributions (3)

- Tables depict the critical value $\chi^2_{a,df}$,i.e, the value such as $P_{df}(X > \chi^2_{a,df}) = \alpha$



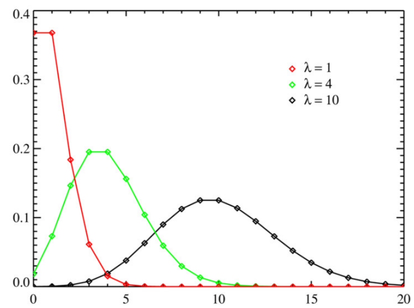
df	P = 0.05	P = 0.01	P = 0.001
1	3.84	6.64	10.83
2	5.99	9.21	13.82
3	7.82	11.35	16.27
4	9.49	13.28	18.47
5	11.07	15.09	20.52
6	12.59	16.81	22.46
7	14.07	18.48	24.32
8	15.51	20.09	26.13
9	16.92	21.67	27.88
10	18.31	23.21	29.59
11	19.68	24.73	31.26
12	21.03	26.22	32.91

Probabilistic distributions (4)

C. Poisson distribution

- Discrete probability to model the number of random occurrences of some phenomenon in a specified unit of space or time.
  The number of occurrences k is given by the probability function f(k; λ), λ is the expected number of occurrences that occur during the given interval

$$f(k;\lambda) = \frac{\lambda^k e^{-k}}{k!}$$



Goodness of fit (1)

[Definition] **Test goodnes of fit** establishes whether ,or not, observed and theoretical frequencies are different.

The most widely-used methods of testing goodness of fit are

- Person's Chi-square.
  Adopted for large number of samples.
- Kolgomorov-Smirnov
  Better for smaller number of samples.

[Definition] **p-value** is the probability related to the test statistic.

Goodness of fit (2)

A. Chi-square test

1. Evaluate chi-square statistic $\chi^2 = \sum_{i=1}^{N} \frac{(O_i - E_i)^2}{E_i}$
   N, number of possible outcomes
   O_i, observed value
   E_i, expected value asserted by H_0
2. Identify the degree of freedom, df.
3. Identify the critical value $\boldsymbol{\chi^2}_{a,df}$ (row df, column α equal to 0.05, 0.01 or 0.001).
4. Reject H_0 if evaluated χ^2 is greater than $\boldsymbol{\chi^2}_{a,df}$. Similarly, reject H_0 if α>p-value.

Notes

- Expected values must always be equal or greater than 5
- χ^2 calculator at http://www.stat.tamu.edu/~west/applets/chisqdemo.html

Goodness of fit (3)

Ex: Consider a coin is tossed 100 times, 47 heads and 53 tails were observed. H_0 is "the observed values are close to the predicted values of an uniform distribution".

- $X_H^2 = (47\text{-}50)^2/50 = 0.18$
  $X_T^2 = (53\text{-}50)^2/50 = 0.18$
  $\chi^2 = 0.18+0.18 = 0.36$
- df=1. For α=0.05, the critical value is $\boldsymbol{\chi^2}_{a,df}$=3.84 » 0.36
  Therefore, the null hypothesis is not rejected and the coin toss is considered fair.

p-value for $\chi^2 = 0.36$, df=1 is 0.549

Goodness of fit (4)

A. KS test

1. Identify cumulative distribution functions for samples, S(x), and hypothesized distribution, F(x).
2. The test statistic is $D = \max|S(x) - F(x)|$

   Identification of D can be done as:
   - Sort sample values $R_1,\ldots,R_N$
   - Identify D+ = max{(i/N)-R_i}, D- = max{R_i-(i-1)/N}
   - D is the greatest D+,D-.
3. Critical value is function of sample size N and significance level α:

$$CV(\alpha = 0.01) = \frac{1.63}{\sqrt{N}}$$
$$CV(\alpha = 0.05) = \frac{1.36}{\sqrt{N}}$$
$$CV(\alpha = 0.10) = \frac{1.22}{\sqrt{N}}$$

Goodness of fit (5)

- Let samples be {0.44, 0.81, 0.14, 0.05, 0.93} and hypothesize an uniform distribution with α=0.05

R_i	**0.05**	**0.14**	**0.44**	**0.81**	**0.93**
i/N	0.20	0.40	0.60	0.80	1.00
i/N-R_i	0.15	0.26	0.16	-	0.07
R_i-(i-1)/N	0.05	-	0.04	0.21	0.13

- D is max{0.26,0.21}=0.26
- For N=5, α=0.05 the critical value is 0.608: H_0 is accepted

Golomb postulates (1)

[Definition] A **block** (**gap**) is a continuous sequence of 1's (0's). A **run** is a block or a gap.

[Definition] A sequence **period** is the smallest τ, such as s(i+τ)=s(i) for all 0≤i<N-τ
Ex: sequence 010 010 01, with length 8, has period 3.

[Definition] **Autocorrelation** measures the amount of similarity between a sequence s and a shift of s by t positions.

$$C(t) = \frac{1}{N}\sum_{i=0}^{N-1}(2s_i - 1)(2s_{i+t} - 1), \quad 0 \le t \le N-1$$

Golomb postulates (2)

1. In a sequence, the number of 1's and 0's differ, at most, by one.
   Note: in sequence 1010101010101010101010101010101010, the number of 0's and 1's is equal. Yet, the sequenced is not random (that is detected by 2nd Golomb postulate).
2. At least half of the runs has length 1, at least one-fourth has length 2, at least one-eighth has length 3 and so while the number of runs is greater than one. For each of these runs, there are almost equally many gaps and blocks.
   Ex: let 00001100001110001111100001110001101110000100

Length	Blocks	Gaps	Runs
1	19	25	44 (46%) ← Fails!
2	12	17	29 (30%)
3	6	10	6 (17%)
4	2	4	6 (6%)
5	1	0	1 (1%)

Golomb postulates (3)

3. Autocorrelation function is two-valued

$$N.C(t) = \begin{cases} N & ,\text{if } t = 0 \\ K & ,\text{if } 1 \leq t \leq N \end{cases}$$

Ex: LFSR, defined by x^4+x^3+1, with sequence `011001000111101` satisfy Golomb's postulates

- Number of 0's is 7, number of 1's is 8
  - 15 runs of length 1 (7 gaps, 8 blocks)=57%.
  - 7 runs of length 2 (3 gap, 4 blocks)=27%.
  - 3 runs of length 3 (1 gap, 2 blocks)=12%.
  - 1 run of length 4 (0 gaps, 1 block).
- C(0)=1, C(t)=-1/15 for $1 \leq t \leq 14$

FIPS 140-2 (1)

- Federal Information Processing standard
  - Specifies the security requirements in 11 areas, to be satisfied by a cryptographic module.
  - Current version is FIPS 140-2, issued May 2001.
  - Certification provided by US federal agencies and industry entities.
- When the cryptographic module is powerup, collect from the generator a 20_000 bit string and submit it to 4 tests. If any of the tests fail, then the generator fails.

FIPS 140-2 (2)

A. **Monobit test**: determine if the number of 0's and 1's is similar.

Let n_0, n_1 be the number of 0's and 1's.

$$X_A = \frac{(n_0 - n_1)^2}{N}$$

follows a χ^2 distribution with 1 degree of freedom for a number of samples ≥10.

For $\alpha=10^{-6}$ and 20_000 samples, the number of 1's must be $9654<n_1<10346$.

FIPS 140-2 (3)

B. **Poker test**: identifies frequencies with which certain digits are repeated.

Divide the sequence into k non-overlapping parts, each of lenght m. There are 2^m types of parts, each one occurring n_i times in the sequence.

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

follows a χ^2 distribution with 2^m-1 degrees of freedom.

For 20_000 samples and m=4, parts are {0000, 0001, ..., 1111}. Test pass if $2.6<X_B<46.17$

FIPS 140-2 (4)

C. **Runs test**: determine if the number of is as expected for a random sequence.

In a random sequence of length k, the expected number of gaps (ou blocks) of length i is $e_i=(n-i+3)/2^{i+2}$. Let B_i,G_i be the number of block and gaps of length i and k the largest integer such as $e_i>5$

$$X_C = \sum_{i=1}^{k} \frac{(B_i - e_i)^2}{e_i} + \sum_{i=1}^{k} \frac{(G_i - e_i)^2}{e_i}$$

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

FIPS 140-2 (5)

For 20_000 samples, the number of runs must satisfy table

Length of Run	Required Interval
1	2,267-2733
2	1,079-1,421
3	502-748
4	223-402
5	90-223
6 +	90-223

D. **Long run test**

Passed if there are no runs with length 34 or more.

Approved security functions:

- Symmetric key: AES,Triple-DES
- Asymmetric key: DSA,ECDSA,RSA
- Hashing: SHA (1,224,256,384,512)

Diehard test suite (1)

- Battery of 15 statistical tests, maintained by G. Marsaglia.
- Available at http://www.stat.fsu.edu/pub /diehard
  - Input: specially formatted binary file containing 3 million 32-bit integers.
  - Output: a p-value on [0,1). Test fails if $p<0.025$ or $p>0.975$. Note: even for good PRNG's, such as Blum-Blum-Shub, occasional p-values fail.
  - Statistitical tests vary on 3 types:

Statistical test	Diehard tests
χ^2	Birthday spacing, Overlapping permutations, Rank of matrices (2), count the 1s (2), squeeze, craps
KS	Mimum distance, random spheres, overlapping sums, runs
N	Monkey (2), parking lot

Diehard test suite (2)

Student presentations

- Description on tests, selected by the instructor (one for each type).
- Diehard outputs for a set of 10 outputs generated by one PRNG, which must be implemented by the student.
  - LFSR defined by [32,7,6,2,0] primitive polynomial.
  - A5 cipher system.
  - Mersenne twister.
  - openSSL package.
  - Linux *rand* generator (C library).
  - π generator.