# GM

CRYPTOGRAPHIC INDUSTRY STANDARD

OF THE PEOPLE'S REPUBLIC OF CHINA

ICS 35.030

CCS L 80

# GM/T 0005-2021

Replacing GM/T 0005-2012

# Randomness test specification

随机性检测规范

**Issued on: October 18, 2021** **Implemented on: May 01, 2022**

**Issued by: State Cryptography Administration**

# Table of Contents

# Randomness test specification

## 1 Scope

This document specifies the randomness test indicators and test methods applicable to binary sequence.

## 2 Normative references

There are no normative references in this document.

## 3 Terms and definitions

For the purposes of this document, the following terms and definitions apply.

### 3.1 binary sequence

A bit string consisting of "0" and "1".

**NOTE:** Unless otherwise specified, the sequences referred to in this document are all binary sequences.

### 3.2 randomness hypothesis

When testing the randomness of a binary sequence, it is first assumed that the sequence is random. This hypothesis is called the null hypothesis or the null hypothesis, denoted as $H_0$. The hypothesis contrary to the null hypothesis, that is, the sequence is not random, is called the alternative hypothesis, denoted as $H_\alpha$.

### 3.3 randomness test

A function or procedure for binary sequence test. It can be used to judge whether to accept the randomness null hypothesis.

### 3.4 significance level

The probability of incorrectly judging random sequences as non-random sequences in randomness test.

### 3.5 sample

Binary sequences for randomness test.

**3.6 sample group**

A collection of multiple samples.

**3.7 sample length**

The number of bits in the sample.

**3.8 sample size**

The number of samples in the sample group.

**3.9 test parameter**

Parameters that need to be set for randomness test.

# 4 Symbols

The following symbols apply to this document.

d: The number of bits for the logical left shift of the sequence in autocorrelation test.

$H_0$: Original hypothesis (null hypothesis).

$H_\alpha$: Alternative hypothesis.

K: The number of L-bit subsequences of the sequence to be tested in the general statistical test.

L: General statistical neutron sequence length.

$L_i$: The linear complexity of subsequence in the linear complexity test.

M: The number of rows of the matrix in the matrix rank test.

m: The bit length of the subsequence.

N: The number of m-bit subsequences in an n-bit sequence to be tested.

n: The bit length of the binary sequence to be tested.

Q: The number of columns of the matrix in the matrix rank test, or the number of L-bit subsequences of the initial sequence in the general statistical test.

V: Statistical value.

$X_i$: $2\varepsilon_i-1$.

α: Significance level for sample passing rate test.

$\alpha_T$: Significance level used for sample distribution uniformity test.

ε: Binary sequence to be tested.

ε': A new sequence generated according to certain rules on the basis of ε.

π: The proportion of 1 in the binary sequence to be tested.

Σ: Summation symbol.

*: Multiplication, sometimes omitted.

ln(x): The natural logarithm of x.

$\log_2(x)$: The base 2 logarithm of x.

$\lfloor x \rfloor$: The largest integer not greater than x.

max: Take the maximum value from several elements.

min: Take the minimum value from several elements.

Φ(x): Cumulative distribution function of standard normal distribution.

P_value: A metric to measure the randomness of a sample, which is used to determine the passing rate of the sample.

Q_value: A metric to measure the randomness of samples, which is used to determine the uniformity of sample distribution.

erfc: Complementary Error Function.

igamc: Incomplete Gamma Function.

$V_n(obs)$: The total number of runs in the binary sequence to be tested.

ApEn(m): Approximate entropy of the binary sequence to be tested.

modulus(x): The operation used to calculate the modulus value of the complex coefficient x.

$\nabla \Psi_m^2$: The first statistic in overlapping subsequence test.

$\nabla^2 \Psi_m^2$: The second statistic in overlapping subsequence test.

$$P\_value = 1 - \sum_{i=(-(n/z)+1)/4}^{((n/z)-1)/4} \left[ \Phi\left(\frac{(4i+1)z}{\sqrt{n}}\right) - \Phi\left(\frac{(4i-1)z}{\sqrt{n}}\right) \right] +$$

$$\sum_{i=(-(n/z)-3)/4}^{((n/z)-1)/4} \left[ \Phi\left(\frac{(4i+3)z}{\sqrt{n}}\right) - \Phi\left(\frac{(4i+1)z}{\sqrt{n}}\right) \right] .$$

Step 5: Calculate $Q\_value = P\_value$.

Test settings are as required in Annex A. See B.11 for the test principle.

### 5.11.3 Result determination

Compare the result of P_value calculated in 5.11.2 with α. If P_value≥α, it is considered that the sequence to be tested passes the test of the cumulative sum, otherwise it fails to pass the test of the cumulative sum.

## 5.12 Approximate entropy test method

### 5.12.1 Overview

Approximate entropy test evaluates its randomness by comparing the frequency of m-bit overlapping subsequence modes with the frequency of m+1-bit overlapping subsequence modes. Calculate the frequency difference between the m-bit overlapping subsequence mode and the m+1-bit overlapping subsequence mode. A small difference value indicates that the sequence to be tested has regularity and continuity. A large difference value indicates that the sequence to be tested has irregularity and discontinuity. For any m, the approximate entropy of the random sequence shall be approximately equal to ln2.

### 5.12.2 Test steps

The approximate entropy test steps are as follows.

Step 1: Construct a new sequence ε' from the sequence ε to be tested. The construction method is as follows: Add the m-1 bits of data at the beginning of the sequence ε to the end of the sequence ε to get ε'. The length of the new sequence ε' is n'=n+m-1.

Step 2: Count the frequency of occurrence of the sequence mode of m bits for all $2^m$ in n. Denote the frequency of occurrence of the m-bit mode $i_1i_2\ldots i_m$ as $v_{i1i2\ldots im}$.

Step 3: For all j ($0 \le j \le 2^m-1$), calculate $C_j^m = \frac{v_{i_1 i_2 \cdots i_m}}{n}$, of which j is the decimal value corresponding to the m-bit mode $i_1i_2\ldots i_m$.

### 5.15.3 Result determination

Compare the result of P_value calculated in 5.15.2 with α. If P_value≥α, it is considered that the sequence to be tested passes the discrete Fourier test, otherwise it fails the discrete Fourier test.

# 6 Determination of randomness test

## 6.1 Overview

It shall use the 15 randomness test methods specified in Chapter 5 and the test settings specified in Annex A to test the randomness of the binary sequence sample group. A randomness test method corresponds to at least one randomness test item. If a randomness test method adopts different test parameter settings (see Annex A for details) or has different test modes (such as the in-block maximum run test method, cumulative sum test method), or has multiple statistical values (such as overlapping subsections) sequence test method), it shall be tested as a separate random test item. The passing rate and distribution uniformity of each test item in the binary sequence sample group shall be respectively subjected to determination of conformity. For example, the test method of cumulative sum includes two modes: forward cumulative sum and backward cumulative sum. The forward cumulative sum and the backward cumulative sum shall be tested as 2 independent test items. The passing rate and distribution uniformity of the forward cumulative sum and backward cumulative sum of the binary sequence sample group are respectively judged for conformity.

This document determines the sample size in the binary sequence sample group to be 1000.

## 6.2 Determination of sample passing rate

For each randomness test item, count the number of samples whose P _value is greater than or equal to α in the binary sequence sample group. The significance level determined by this document for sample passing rate test is α=0.01.

Let the sample size be s. When the number of samples passing a test item is greater than or equal to $s\left(1-\alpha-3\sqrt{\frac{\alpha(1-\alpha)}{s}}\right)$, then the sample group shall be considered to pass this test, otherwise it fails this test. For example, if the sample size is 1000, the number of samples that pass the test item shall be greater than or equal to 981.

## 6.3 Determination of sample distribution uniformity

For each randomness test item, the Q_value value of each sample in the binary sequence

Maurer general statistics (referred to as general statistics) test mainly tests whether the sequence to be tested can be compressed losslessly. If the sequence to be tested can be significantly compressed, the sequence is considered non-random, because random sequences cannot be significantly compressed.

General statistics test can be used to test various characteristics of the sequence to be tested. But this does not mean that general statistics test is an assembly of the previous tests. General statistics test takes a completely different approach from other tests. Certain statistical defects of the sequence to be tested can be tested. A sequence can be tested by general statistics if and only if the sequence is incompressible.

General statistics test requires a large amount of data. It divides the sequence into subsequences of length L. Then the sequence to be tested is divided into two parts: the initial sequence and the test sequence. The initial sequence includes Q subsequences. Q shall be greater than or equal to $10*2^L$. The test sequence includes K subsequences. K shall be greater than or equal to $1000*2^L$. Therefore, the sequence length n shall be greater than or equal to $10*2^L*L+1000*2^L*L$. The value range of L shall be $1 \le L \le 16$. The value of L shall not be less than 6. Obviously, when L=6, n is at least 387840. When the sequence length n is constant, take $K=\left\lfloor \frac{n}{L} \right\rfloor -Q$.

First, traverse the initial sequence (unit is blocks) from the beginning. Find the last occurrence of each L-bit pattern in the initial sequence (block number). If an L-bit mode does not appear in the initial sequence, set its position to 0. After that, the test sequence is traversed from the beginning. Obtain an L-bit subsequence each time. Calculate the difference between the position of this subsequence and the position of the last occurrence before it. That is the block number subtraction. Record the subtraction result as the distance len. Then take the base 2 logarithm of the distance len. Finally, add all the logarithm results. In this way, the statistical value shall be obtained:

$$f_n = \frac{1}{K} \sum_{i=Q+1}^{Q+K} \log_2(len)$$

Calculate the expected value:

$$\mu = E(f_n) = 2^{-L} \sum_{i=1}^{+\infty} (1-2^{-L})^{i-1} \log_2 i$$

In fact, the expected value of fn is the expected value of the random variable $\log_2 G$. Where, $G=G_L$ is the geometric distribution with parameters $1-2^{-L}$. The geometric distribution is defined as: Let the probability of a successful Bernoulli experiment be p, and take the random variable X as the number of independent Bernoulli experiments performed before the success, then:

**This is an excerpt of the PDF (Some pages are marked off intentionally)**

Full-copy PDF can be purchased from 1 of 3 websites:

1. https://www.ChineseStandard.us

- SEARCH the standard ID, such as GB 4943.1-2022.
- Select your country (currency), for example: USA (USD); Germany (Euro).
- Full-copy of PDF (text-editable, true-PDF) can be downloaded in 9 seconds.
- Tax invoice can be downloaded in 9 seconds.
- Receiving emails in 9 seconds (with download links).

2. https://www.ChineseStandard.net

- SEARCH the standard ID, such as GB 4943.1-2022.
- Add to cart. Only accept USD (other currencies - https://www.ChineseStandard.us).
- Full-copy of PDF (text-editable, true-PDF) can be downloaded in 9 seconds.
- Receiving emails in 9 seconds (with PDFs attached, invoice and download links).

3. https://www.google.com/search?tbm=bks&q=ChineseStandard.net

- SEARCH the standard ID, such as GB 4943.1-2022.
- Google Books -- Select your currency.
- Processed by Google (delivery, tax invoice etc.). Delivered in 9 seconds by Google.
- Tips: Download an unprotected True-PDF (text-editable) from Google-Books:
  1. https://play.google.com/books → 2. Sign in → Google account
  3. Find the BOOK you bought → 4. Click "3-dots" → Export
  5. Save as "*.pdf" (Save True-PDF to your local computer for offline reading/printing)

Translated by: Field Test Asia Pte. Ltd. (Incorporated & taxed in Singapore. Tax ID: 201302277C)

Accountable person and shareholder: Wayne Zheng

About Us (Goodwill, Policies, Fair Trading…): https://www.chinesestandard.net/AboutUs.aspx

Contact: Wayne Zheng, Sales@ChineseStandard.net

Linkin: https://www.linkedin.com/in/waynezhengwenrui/

**------ The End ------**