ICS 35.030
CCS L 80

GM

中华人民共和国密码行业标准

GM/T 0005—2021
代替 GM/T 0005—2012

随机性检测规范

Randomness test specification

2021-10-18 发布　　2022-05-01 实施

国家密码管理局　　发布

目　　次

前　言

本文件依据 GB/T 1.1—2020 给出的规则起草。

本文件代替 GM/T 0005—2012《随机性检测规范》，对随机性检测进行规范，为二元序列的随机性检测工作提供科学依据。与 GM/T 0005—2012 相比，除编辑性修改外主要技术变化如下：

a) 本文件适用范围由“适用于对随机数发生器产生的二元序列的随机性检测”改为“适用于对二元序列的随机性检测”(见第 1 章和 2012 年版的第 1 章)；

b) 删除了“随机数发生器”、“P 值”、“游程”的术语以及“单比特频数检测”等 15 个检测项的术语定义(见 2012 年版的第 2 章)，新增了术语“样本集”(见 3.6)；

c) 修改了符号 α、P_value 的说明(见第 4 章和 2012 年版的第 3 章)，增加了符号 α_T、Q_value 的说明(见第 4 章)；

d) 删除了“二元序列的检测”章节，新增“随机性检测方法”章节，分别从概述、检测步骤、结果判定对 15 项检测方法进行展开说明，其中每项检测方法的检测步骤中均增加 Q_value 的计算(见第 5 章和 2012 年版的第 4 章)；

e) 删除了“随机数发生器的检测”章节，新增“随机性检测判定”章节，分别从概述、样本通过率判定、样本分布均匀性判定、随机性检测结果判定进行说明，其中增加了对 Q_value 的样本分布均匀性判定要求[见第 6 章和 2012 年版的第 5 章)]；

f) 修改游程分布检测方法中的统计值构造方法(见 5.6.2 和 2012 年版的 4.4.7)；

g) 块内最大游程检测方法新增块内最大“0”游程检测模式(见 5.7)；

h) 累加和检测方法新增后向累加和检测模式(见 5.11)；

i) 删除“随机性检测参数设置表”(见 2012 年版的表 B.1)；

j) 新增三种样本长度及检测设置表(见附表 A.1、A.2、A.3)；

k) 删除“随机性检测结果分析表”(见 2012 年版的附录 C)；

l) 随机性检测原理调整为附录 B(见附录 B 及 2012 年版的附录 A)；

m) 修改块内最大游程的 π_i 取值(见附表 B.4 及 2012 年版的附表 A.3)；

n) 新增随机性检测结果示例(见附录 C)。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别这些专利的责任。

本文件的附录 A 是规范性附录。本标准的附录 B、附录 C 是资料性附录。

本文件由密码行业标准化技术委员会提出并归口。

本文件起草单位：国家密码管理局商用密码检测中心、中国科学院软件研究所、中国科学院信息工程研究所、北京宏思电子技术有限责任公司、浙江大学。

本文件主要起草人：罗鹏、毛颖颖、陈华、范丽敏、马原、李亚威、张文婧、沈海斌、陈美会、朱少峰、张贺、朱双怡。

本文件的历次版本发布情况为：

——GM/T 0005—2012。

随机性检测规范

1 范围

本文件规定了适用于二元序列的随机性检测指标和检测方法。

2 规范性引用文件

本文件没有规范性引用文件。

3 术语和定义

下列术语和定义适用于本文件。

3.1

二元序列 binary sequence

由“0”和“1”组成的比特串。

注：如无特别说明，本文件所指的序列均为二元序列。

3.2

随机性假设 randomness hypothesis

对二元序列做随机性检测时，首先假设该序列是随机的，这个假设称为原假设或零假设，记为 H_0。与原假设相反的假设，即这个序列是不随机的，称为备择假设，记为 H_α。

3.3

随机性检测 randomness test

用于二元序列检测的一个函数或过程，可以通过它来判断是否接受随机性原假设。

3.4

显著性水平 significance level

随机性检测中错误地判断随机序列为非随机序列的概率。

3.5

样本 sample

用于随机性检测的二元序列。

3.6

样本集 sample group

多个样本的集合。

3.7

样本长度 sample length

样本的比特个数。

3.8

样本数量 sample size

样本集中的样本个数。

3.9

检测参数　test parameter

随机性检测需要设定的参数。

4　符号

下列符号适用于本文件。

d：自相关检测中序列逻辑左移的位数。

H_0：原假设(零假设)。

H_α：备择假设。

K：通用统计检测中待检序列 L 位子序列个数。

L：通用统计中子序列长度。

L_i：线性复杂度检测中子序列的线性复杂度。

M：矩阵秩检测中矩阵的行数。

m：子序列的比特长度。

N：一个待检测的 n 比特序列中 m 位子序列的个数。

n：待检二元序列的比特长度。

Q：矩阵秩检测中矩阵的列数，或者是通用统计检测中初始序列 L 位子序列的个数。

V：统计值。

X_i：$2\varepsilon_i-1$。

α：用于样本通过率检测的显著性水平。

α_T：用于样本分布均匀性检测的显著性水平。

ε：待检二元序列。

ε'：在 ε 的基础上按照一定的规则产生出的新序列。

π：待检二元序列中 1 的比例。

$\sum$：求和符号。

$*$：乘法，有时省略。

$\ln(x)$：x 的自然对数。

$\log_2(x)$：以 2 为底的 x 的对数。

$\lfloor x \rfloor$：不大于 x 的最大整数。

max：从若干个元素中取最大值。

min：从若干个元素中取最小值。

$\Phi(x)$：标准正态分布的累积分布函数。

P_value：一种衡量样本随机性好坏的度量指标，用于样本通过率判定。

Q_value：一种衡量样本随机性好坏的度量指标，用于样本分布均匀性判定。

$erfc$：余误差函数(Complementary Error Function)。

$igamc$：不完全伽玛函数(Incomplete Gamma Function)。

$V_n(obs)$：待检二元序列中游程的总数。

$ApEn(m)$：待检二元序列的近似熵。

$modulus(x)$：用来计算复系数 x 的模值的运算。

$\nabla\Psi_m^2$：重叠子序列检测中的第一个统计值。

$\nabla^2\Psi_m^2$：重叠子序列检测中的第二个统计值。

5 随机性检测方法

5.1 单比特频数检测方法

5.1.1 概述

单比特频数检测是最基本的检测，用来检测一个二元序列中 0 和 1 的个数是否相近。随机序列应具有较好的 0、1 平衡性。

5.1.2 检测步骤

单比特频数检测步骤如下。

第一步：该检测将待检序列 ε 中的 0 和 1 分别转换成 −1 和 1，$X_i=2\varepsilon_i-1(1\leqslant i\leqslant n)$。

第二步：累加求和计算得到 $S_n=\sum_{i=1}^{n}X_i$。

第三步：计算统计值 $V=\frac{S_n}{\sqrt{n}}$。

第四步：计算 $P\_value=erfc\left(\frac{|V|}{\sqrt{2}}\right)$。

第五步：计算 $Q\_value=\frac{1}{2}erfc\left(\frac{V}{\sqrt{2}}\right)$。

检测设置按附录 A 要求，检测原理见附录 B 的 B.1。

5.1.3 结果判定

将 5.1.2 中计算得出的 P_value 结果与 α 进行比较，如果 $P\_value\geqslant\alpha$，则认为待检序列通过单比特频数检测，否则未通过单比特频数检测。

5.2 块内频数检测方法

5.2.1 概述

块内频数检测用来检测待检序列的 m 位子序列中 1 的个数是否接近$\frac{m}{2}$。对随机序列来说，其任意长度的 m 位子序列中 1 的个数都应该接近$\frac{m}{2}$。

5.2.2 检测步骤

块内频数检测步骤如下。

第一步：将待检序列 ε 分成 $N=\left\lfloor\frac{n}{m}\right\rfloor$个长度为 m 的非重叠子序列，将多余的比特舍弃。

第二步：计算每个子序列中 1 所占的比例 $\pi_i=\frac{\sum_{j=1}^{m}\varepsilon_{(i-1)m+j}}{m}$，$1\leqslant i\leqslant N$。

第三步：计算统计量 $V=4m\sum_{i=1}^{N}\left(\pi_i-\frac{1}{2}\right)^2$。

第四步：计算 $P\_value=igamc\left(\frac{N}{2},\frac{V}{2}\right)$。

第五步：计算 $Q\_value=P\_value$。

检测设置按附录 A 要求，检测原理见 B.2。

5.2.3 结果判定

将 5.2.2 中计算得出的 P_value 结果与 α 进行比较，如果 $P\_value\geqslant\alpha$，则认为待检序列通过块内频数检测，否则未通过块内频数检测。

5.3 扑克检测方法

5.3.1 概述

扑克检测用来检测长度为 m 的 2^m 类子序列的个数是否接近。对于随机的序列，2^m 类子序列的个数应该接近。

5.3.2 检测步骤

扑克检测步骤如下。

第一步：将待检序列 ε 划分成 $N=\left\lfloor\frac{n}{m}\right\rfloor$ 个长度为 m 的非重叠子序列，将多余的比特舍弃。统计第 i 类子序列模式出现的频数，用 $n_i(1\leqslant i\leqslant 2^m)$ 表示。

第二步：计算统计值 $V=\frac{2^m}{N}\sum_{i=1}^{2^m}n_i^2-N$。

第三步：计算 $P\_value=igamc\left(\frac{2^m-1}{2},\frac{V}{2}\right)$。

第四步：计算 $Q\_value=P\_value$。

检测设置按附录 A 要求，检测原理见 B.3。

5.3.3 结果判定

将 5.3.2 中计算得出的 P_value 结果与 α 进行比较，如果 $P\_value\geqslant\alpha$，则认为待检序列通过扑克检测，否则未通过扑克检测。

5.4 重叠子序列检测方法

5.4.1 概述

对任意的正整数 m，长度为 m 的二元序列有 2^m 类。重叠子序列检测将长度为 n 的待检序列划分成 n 个可叠加的 m 位子序列。对随机二元序列来说，由于其具有均匀性，故 m 位可叠加子序列的每一类模式出现的概率应该接近。

5.4.2 检测步骤

重叠子序列检测步骤如下。

第一步：由待检序列 ε 构造一个新的序列 ε'，构造方法如下：将序列 ε 最开始的 $m-1$ 位数据添加到序列 ε 的结尾即可得到新序列 ε'，新序列 ε' 的长度为 $n'=n+m-1$。

第二步：计算 ε' 中每一类 m 位子序列模式（共有 2^m 类）出现的频数，记 m 位子序列模式 $i_1i_2\cdots i_m$ 的

出现频数为 $v_{i_1 i_2 \cdots i_m}$。计算每一类 $m-1$ 位子序列模式(共有 2^{m-1} 类)出现的频数,记 $m-1$ 位子序列模式 $i_1 i_2 \cdots i_{m-1}$ 的出现频数为 $v_{i_1 i_2 \cdots i_{m-1}}$。计算每一个 $m-2$ 位子序列模式(共有 2^{m-2} 类)出现的频数,记 $m-2$ 位子序列模式 $i_1 i_2 \cdots i_{m-2}$ 的出现频数为 $v_{i_1 i_2 \cdots i_{m-2}}$。

第三步:计算

$$\Psi_m^2 = \frac{2^m}{n} \sum_{i_1 i_2 \cdots i_m} v_{i_1 i_2 \cdots i_m}^2 - n,$$

$$\Psi_{m-1}^2 = \frac{2^{m-1}}{n} \sum_{i_1 i_2 \cdots i_{m-1}} v_{i_1 i_2 \cdots i_{m-1}}^2 - n,$$

$$\Psi_{m-2}^2 = \frac{2^{m-2}}{n} \sum_{i_1 i_2 \cdots i_{m-2}} v_{i_1 i_2 \cdots i_{m-2}}^2 - n。$$

第四步:计算

$$\nabla \Psi_m^2 = \Psi_m^2 - \Psi_{m-1}^2,$$

$$\nabla^2 \Psi_m^2 = \Psi_m^2 - 2\Psi_{m-1}^2 + \Psi_{m-2}^2。$$

第五步:计算 $P\_value1 = igamc\left(2^{m-2}, \frac{\nabla \Psi_m^2}{2}\right)$,$P\_value2 = igamc\left(2^{m-3}, \frac{\nabla^2 \Psi_m^2}{2}\right)$。

第六步:计算 $Q\_value1 = P\_value1$,$Q\_value2 = P\_value2$。

检测设置按附录 A 要求,检测原理见 B.4。

5.4.3 结果判定

将 5.4.2 中计算得出的两个结果 P_value1 和 P_value2 分别与 α 进行比较。如果 $P\_value1 \geqslant \alpha$,则认为待检序列通过 P_value1 的重叠子序列检测,否则未通过 P_value1 的重叠子序列检测;如 $P\_value2 \geqslant \alpha$,则认为待检序列通过 P_value2 的重叠子序列检测,否则未通过 P_value2 的重叠子序列检测。

5.5 游程总数检测方法

5.5.1 概述

游程是指序列中由连续的“0”或者“1”组成的子序列,并且该子序列的前导与后继元素都与其本身的元素不同。游程总数检测主要检测待检序列中游程的总数是否服从随机性要求。

5.5.2 检测步骤

游程总数检测步骤如下。

第一步:对长度为 n 的待检序列 $\varepsilon_1 \varepsilon_2 \cdots \varepsilon_n$,计算 $V_n(obs) = \sum_{i=1}^{n-1} r(i) + 1$。其中,当 $\varepsilon_i = \varepsilon_{i+1}$ 时,$r(i) = 0$;否则,$r(i) = 1$。

第二步:计算序列中 1 的比例 $\pi = \frac{\sum_{i=1}^{n} \varepsilon_i}{n}$。

第三步:计算统计值 $V = \frac{V_n(obs) - 2n\pi(1-\pi)}{2\sqrt{n}\pi(1-\pi)}$。

第四步:计算 $P\_value = erfc\left(\frac{|V|}{\sqrt{2}}\right)$。

第五步：计算 $Q\_value=\frac{1}{2}erfc\left(\frac{V}{\sqrt{2}}\right)$。

检测设置按附录 A 要求，检测原理见 B.5。

5.5.3 结果判定

将 5.5.2 中计算得出的 P_value 结果与 α 进行比较。如果 $P\_value\geqslant\alpha$，则认为待检序列通过游程总数检测，否则未通过游程总数检测。

5.6 游程分布检测方法

5.6.1 概述

游程分布检测用于检测序列中相同长度游程分布是否均匀，随机的序列中，相同长度的游程数目应该接近一致，且游程长度每增加一比特，游程数目应接近减半。

5.6.2 检测步骤

游程分布检测步骤如下。

第一步：计算 $e_i=\frac{n-i+3}{2^{i+2}}$，$1\leqslant i\leqslant n$，令 k 为满足 $e_i\geqslant 5$ 的最大正整数 i。

第二步：统计待检序列 ε 中每一个游程的长度。变量 b_i，g_i 分别记录一个二元序列中长度为 i 的“1”游程和“0”游程的数目。其中，长度超过 k 的游程分别计入 b_k 和 g_k。

第三步：计算 $T=\sum_{i=1}^{k}(b_i+g_i)$。

第四步：计算 $\begin{cases}e'_i=T/2^{i+1}, 1\leqslant i\leqslant k-1\\ e'_i=T/2^{i}, i=k\end{cases}$

第五步：计算 $V=\sum_{i=1}^{k}\frac{(b_i-e'_i)^2}{e'_i}+\sum_{i=1}^{k}\frac{(g_i-e'_i)^2}{e'_i}$。

第六步：计算 $P\_value=igamc\left(k-1,\frac{V}{2}\right)$。

第七步：计算 $Q\_value=P\_value$。

检测设置按附录 A 要求，检测原理见 B.6。

5.6.3 结果判定

将 5.6.2 中计算得出的 P_value 结果与 α 进行比较。如果 $P\_value\geqslant\alpha$，则认为待检序列通过游程分布检测，否则未通过游程分布检测。

5.7 块内最大游程检测方法

5.7.1 概述

块内最大游程检测方法分别对块内最大“1”游程和块内最大“0”游程两种模式进行检测。将待检序列划分成 N 个长度为 m 的子序列，此时 $n=N*m$，统计各个子序列中的最长“1”游程长度和最长“0”游程长度，根据各个子序列中最大“1”游程、最大“0”游程的分布来评价待检序列的随机性。

5.7.2 检测步骤

块内最大“1”/“0”游程的检测步骤如下。

第一步：将待检序列 ε 划分成 $N=\left\lfloor\frac{n}{m}\right\rfloor$ 个长度为 m 的非重叠子序列，舍弃多余的位不用。

第二步：设置 $K+1$ 个集合（K 的取值见表 B.2），计算每一个子序列中最大“1”/“0”游程的长度，参照表 B.3 的规则将其归入相应的集合中。

第三步：计算统计值 $V=\sum_{i=0}^{K}\frac{(v_i-N\pi_i)^2}{N\pi_i}$，其中 v_i 表示第 i 个集合中的元素个数，π_i 的定义见本文件附表 B.4。

第四步：计算 $P\_value=igamc\left(\frac{K}{2},\frac{V}{2}\right)$。

第五步：计算 $Q\_value=P\_value$。

检测设置按附录 A 要求，检测原理见 B.7。

5.7.3 结果判定

将 5.7.2 中计算得出的 P_value 结果与 α 进行比较。如果 $P\_value\geqslant\alpha$，则认为待检序列通过块内最大游程检测，否则未通过块内最大游程检测。

5.8 二元推导检测方法

5.8.1 概述

二元推导检测的目的是判定第 k 次二元推导序列中 0 和 1 的个数是否接近一致。对于长度为 n 的二元初始序列，依次将初始序列中两个相邻比特做异或操作，即可得到该序列的一次二元推导序列，长度为 $n-1$。依次执行上述操作 k 次，即可得到该初始序列的 k 次二元推导序列，长度为 $n-k$。对于一个随机的序列，无论进行多少次推导，其 0、1 的个数都应该接近一致。

5.8.2 检测步骤

二元推导检测步骤如下。

第一步：对待检序列 ε，依次将初始序列中相邻两个比特做异或操作得到新序列 ε′，即 $\varepsilon'_i=\varepsilon_i\oplus\varepsilon_{i+1}$。

第二步：重复第一步，操作 k 次。

第三步：将新序列 ε′中的 0 和 1 分别转换成 -1 和 1，然后对其累加求和得 $S_{n-k}=\sum_{i=1}^{n-k}(2\varepsilon'_i-1)$。

第四步：计算统计值 $V=\frac{S_{n-k}}{\sqrt{n-k}}$。

第五步：计算 $P\_value=erfc\left(\frac{|V|}{\sqrt{2}}\right)$。

第六步：计算 $Q\_value=\frac{1}{2}erfc\left(\frac{V}{\sqrt{2}}\right)$。

检测设置按附录 A 要求，检测原理见 B.8。

5.8.3 结果判定

将 5.8.2 中计算得出的 P_value 结果与 α 进行比较。如果 $P\_value\geqslant\alpha$，则认为待检序列通过二元推导检测，否则未通过二元推导检测。

5.9 自相关检测方法

5.9.1 概述

自相关检测用来检测待检序列与将其左移(逻辑左移)d 位后所得新序列的关联程度。一个随机序列应该和将其左移任意位所得的新序列都是独立的,故其关联程度也应该很低,即初始序列与将其左移 d 位后所得新序列进行异或操作形成的新序列中,0、1 的个数应该接近一致。

5.9.2 检测步骤

自相关检测步骤如下。

第一步:计算 $A(d)=\sum_{i=0}^{n-d-1}(\varepsilon_i \oplus \varepsilon_{i+d})$。

第二步:计算统计值 $V=\dfrac{2(A(d)-((n-d)/2))}{\sqrt{n-d}}$。

第三步:计算 $P\_value=erfc\left(\dfrac{|V|}{\sqrt{2}}\right)$。

第四步:计算 $Q\_value=\dfrac{1}{2}erfc\left(\dfrac{V}{\sqrt{2}}\right)$。

检测设置按附录 A 要求,检测原理见 B.9。

5.9.3 结果判定

将 5.9.2 中计算得出的 P_value 结果与 α 进行比较。如果 $P\_value \geqslant \alpha$,则认为待检序列通过自相关检测,否则未通过自相关检测。

5.10 矩阵秩检测方法

5.10.1 概述

矩阵秩检测用来检测待检序列中给定长度的子序列之间的线性独立性。由待检序列构造矩阵,然后检测矩阵的行或列之间的线性独立性,矩阵秩的偏移程度可以给出关于线性独立性的量的认识,从而影响对二元序列随机性好坏的评价。

5.10.2 检测步骤

矩阵秩检测步骤如下。

第一步:将待检序列 ε 分成大小为 32 * 32 的子序列,共有 $N=\left\lfloor\dfrac{n}{32*32}\right\rfloor$ 个,舍弃多余的位不用。将每一个 32 * 32 的子序列组装成一个 32×32 的矩阵,此矩阵有 32 行 32 列,每一行则由序列 ε 中连续的 32 位填充。

第二步:计算每一个矩阵的秩 $R_i(i=1,2,\cdots,N)$。

第三步:令 F_M 为秩为 32 的矩阵的个数,令 F_{M-1} 为秩为 31 的矩阵的个数,则 $N-F_M-F_{M-1}$ 为秩小于 31 的矩阵的个数。

第四步:计算统计值

$$V=\frac{(F_M-0.288\ 8N)^2}{0.288\ 8N}+\frac{(F_{M-1}-0.577\ 6N)^2}{0.577\ 6N}+\frac{(N-F_M-F_{M-1}-0.133\ 6N)^2}{0.133\ 6N}。$$

第五步：计算 $P\_value=igamc\left(1,\frac{V}{2}\right)$。

第六步：计算 $Q\_value=P\_value$。

检测设置按附录 A 要求，检测原理见 B.10。

5.10.3 结果判定

将 5.10.2 中计算得出的 P_value 结果与 α 进行比较。如果 $P\_value\geqslant\alpha$，则认为待检序列通过矩阵秩检测，否则未通过矩阵秩检测。

5.11 累加和检测方法

5.11.1 概述

累加和检测方法分别对前向累加和、后向累加和两种模式进行检测。前向累加和检测从待检序列第 1 比特开始，逐比特向后计算，后向累加和检测从待检序列最后 1 比特开始，逐比特向前计算，通过判断待检序列的各个子序列中最大的偏移（与 0 之间），也就是最大累加和与一个随机序列应具有的最大偏移相比较，以判断待检序列的随机性。

5.11.2 检测步骤

累加和检测步骤如下。

第一步：将待检序列 ε 中的 0 和 1 分别转换为 -1 和 1，$X_i=2\varepsilon_i-1(1\leqslant i\leqslant n)$。

第二步：计算序列累加和 $S_i(1\leqslant i\leqslant n)$。

第三步：计算 $z=max_{1\leqslant i\leqslant n}|S_i|$。

第四步：计算

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

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

第五步：计算 $Q\_value=P\_value$。

检测设置按附录 A 要求，检测原理见 B.11。

5.11.3 结果判定

将 5.11.2 中计算得出的 P_value 结果与 α 进行比较。如果 $P\_value\geqslant\alpha$，则认为待检序列通过累加和检测，否则未通过累加和检测。

5.12 近似熵检测方法

5.12.1 概述

近似熵检测通过比较 m 位可重叠子序列模式的频数和 $m+1$ 位可重叠子序列模式的频数来评价其随机性。计算 m 位可重叠子序列模式和 $m+1$ 位可重叠子序列模式之间的频数差异，差异值较小则表明待检序列具有规则性和连续性；差异值较大则表明待检序列具有不规则性和不连续性。对任意一个 m 来说，随机序列的近似熵应该近似等于 ln2。

5.12.2 检测步骤

近似熵检测步骤如下。

第一步：由待检序列 ε 构造一个新的序列 ε'，构造方法如下：将序列 ε 最开始的 $m-1$ 位数据添加到序列 ε 的结尾即可得到 ε'，新序列 ε' 的长度为 $n'=n+m-1$。

第二步：计算 ε' 中所有的 2^m 个 m 位子序列模式的出现频数，记 m 位模式 $i_1 i_2 \cdots i_m$ 出现的频数为 $v_{i_1 i_2 \cdots i_m}$。

第三步：对于所有的 $j(0 \leqslant j \leqslant 2^m-1)$，计算 $C_j^m=\frac{v_{i_1 i_2 \cdots i_m}}{n}$，其中 j 为 m 位模式 $i_1 i_2 \cdots i_m$ 对应的十进制数值。

第四步：计算 $\varphi^{(m)}=\sum_{i=0}^{2^m-1} C_i^m \ln C_i^m$，如果 $C_i^m=0$，则定义 $C_i^m \ln C_i^m=0$。

第五步：用 $m+1$ 代替 m，重复操作第一步至第四步，计算得到 $\varphi^{(m+1)}$。

第六步：计算 $ApEn(m)=\varphi^{(m)}-\varphi^{(m+1)}$，计算统计值 $V=2n[\ln 2-ApEn(m)]$。

第七步：计算 $P\_value=igamc(2^{m-1},\frac{V}{2})$。

第八步：计算 $Q\_value=P\_value$。

检测设置按附录 A 要求，检测原理见 B.12。

5.12.3 结果判定

将 5.12.2 中计算得出的 P_value 结果与 α 进行比较。如果 $P\_value \geqslant \alpha$，则认为待检序列通过近似熵检测，否则未通过近似熵检测。

5.13 线性复杂度检测方法

5.13.1 概述

线性复杂度检测用于检测各等长子序列的线性复杂度分布是否符合随机性的要求。将待检序列划分成 N 个长度为 m 的子序列，此时 $n=N*m$，然后利用 Berlekamp-Massey 算法计算每个子序列的线性复杂度 L_i，根据 L_i 的分布情况判断待检二元序列的随机性。

5.13.2 检测步骤

线性复杂度检测步骤如下：

第一步：将待检序列 ε 划分为 $N=\left\lfloor \frac{n}{m} \right\rfloor$ 个长度为 m 的非重叠子序列，将多余的比特舍弃。

第二步：计算每一个子序列的线性复杂度 $L_i(1 \leqslant i \leqslant N)$。

第三步：计算 $\mu=\frac{m}{2}+\frac{9+(-1)^{m+1}}{36}-\frac{1}{2^m}\left(\frac{m}{3}+\frac{2}{9}\right)$。

第四步：对每一个子序列，计算 $T_i=(-1)^m(L_i-\mu)+\frac{2}{9}$。

第五步：设置 7 个正整数 $v_0, v_1, \cdots, v_6$，将这 7 个正整数的初值都设为 0。对所有的 $1 \leqslant i \leqslant N$ 有：

$T_i \leqslant -2.5$，v_0 加 1；

$-2.5 < T_i \leqslant -1.5$，$v_1$ 加 1；

$-1.5<T_i\leqslant-0.5$，v_2加 1；

$-0.5<T_i\leqslant0.5$，v_3加 1；

$0.5<T_i\leqslant1.5$，v_4加 1；

$1.5<T_i\leqslant2.5$，v_5加 1；

$T_i>2.5$，v_6加 1。

第六步：计算统计值 $V=\sum_{i=0}^{6}\frac{(v_i-N\pi_i)^2}{N\pi_i}$。其中，$\pi_i$值为：

$\pi_0=0.010\ 417$，$\pi_1=0.031\ 250$，$\pi_2=0.125$，$\pi_3=0.500$，$\pi_4=0.250$，$\pi_5=0.062\ 500$，$\pi_6=0.020\ 833$。

第七步：计算 $P\_value=igamc\left(3,\frac{V}{2}\right)$。

第八步：计算 $Q\_value=P\_value$。

检测设置按附录 A 要求，检测原理见 B.13。

5.13.3 结果判定

将 5.13.2 中计算得出的 P_value 结果与 α 进行比较。如果 $P\_value\geqslant\alpha$，则认为待检序列通过线性复杂度检测，否则未通过线性复杂度检测。

5.14 Maurer 通用统计检测方法

5.14.1 概述

Maurer 通用统计检测用于检测待检序列能否被无损压缩。因为随机序列是不能被显著压缩，因此如果待检序列能被显著地压缩，则认为该序列不随机。

5.14.2 检测步骤

Maurer 通用统计检测步骤如下：

第一步：将待检序列 ε 分成两部分：初始序列和测试序列。初始序列包括 Q 个 L 位的非重叠的子序列，测试序列包括 K 个 L 位的非重叠的子序列，$K=\left\lfloor\frac{n}{L}\right\rfloor-Q$，末尾不够组成一个完整 L 位子序列的多余位舍弃不用。

第二步：针对初始序列，创建一个表，它以 L 位值作为表中的索引值，$T_j(0\leqslant j<2^L)$表示表中第 j 个元素的值，计算 $T_j=i(1\leqslant i\leqslant Q)$，其中 j 是初始序列中第 i 个 L 位子序列的十进制表示。

第三步：计算 $sum=\sum_{i=Q+1}^{Q+K}\log_2(i-T_j)$，其中，$j$ 是待检序列中第 i 个 L 位子序列的十进制表示，T_j 表示当前表中第 j 个元素的值，在遍历第 i 个$(Q+1\leqslant i\leqslant Q+K)$$L$ 位子序列后，应更新 $T_j=i$。

第四步：计算 $V=\frac{\frac{sum}{K}-E(L)}{\sigma}$，$E(L)$和 σ 的计算见 B.14。

第五步：计算 $P\_value=erfc\left(\frac{|V|}{\sqrt{2}}\right)$。

第六步：计算 $Q\_value=\frac{1}{2}erfc\left(\frac{V}{\sqrt{2}}\right)$。

检测设置按附录 A 要求，检测原理见 B.14。

5.14.3 结果判定

将5.14.2中计算得出的P_value结果与α进行比较。如果$P_value \geqslant \alpha$，则认为待检序列通过Maurer通用统计检测，否则未通过Maurer通用统计检测。

5.15 离散傅立叶检测方法

5.15.1 概述

离散傅立叶检测使用频谱的方法来检测序列的随机性。对待检序列进行傅立叶变换后可以得到尖峰高度，根据随机性的假设，这个尖峰高度不能超过某个门限值(与序列长度n有关)，否则将其归入不正常的范围；如果不正常的尖峰个数超过了允许值，即可认为待检序列是不随机的。

5.15.2 检测步骤

离散傅立叶检测步骤如下：

第一步：将待检序列ε中的0和1分别转换成-1和1，得到新序列$X_1, X_2, \cdots, X_n$，其中$X_i = 2\varepsilon_i - 1$。

第二步：对新序列进行傅立叶变换，得到一系列的复数$f_0, f_1, \cdots, f_{n-1}$。

第三步：对每一个f_j，计算其系数$mod_j = modulus(f_j) = |f_j|$，这里$j \in [0, n/2-1]$。

第四步：计算门限值$T = \sqrt{2.995\ 732\ 274 * n}$。

第五步：计算$N_0 = 0.95 * \frac{n}{2}$。

第六步：计算系数mod_j小于门限值T的复数个数，记作N_1。

第七步：计算统计值$V = \frac{N_1 - N_0}{\sqrt{0.95 * 0.05 * \frac{n}{3.8}}}$。

第八步：计算$P_value = erfc\left(\frac{|V|}{\sqrt{2}}\right)$。

第九步：计算$Q_value = \frac{1}{2} erfc\left(\frac{V}{\sqrt{2}}\right)$。

检测设置按附录A要求，检测原理见B.15。

5.15.3 结果判定

将5.15.2中计算得出的P_value结果与α进行比较。如果$P_value \geqslant \alpha$，则认为待检序列通过离散傅立叶检测，否则未通过离散傅立叶检测。

6 随机性检测判定

6.1 概述

应采用第5章规定的15种随机性检测方法和附录A规定的检测设置对二元序列样本集进行随机性检测。一种随机性检测方法对应至少一个随机性检测项目，其中如某一项随机性检测方法采用不同检测参数设置(详见附录A)，或具有不同检测模式(如块内最大游程检测方法、累加和检测方法)，或具

有多个统计值(如重叠子序列检测方法),应作为单独的随机性检测项目进行检测,并分别对二元序列样本集的每个检测项目的样本通过率、分布均匀性进行合格判定。比如累加和检测方法包括前向累加和、后向累加和两种模式,前向累加和、后向累加和应作为2个独立的检测项目进行检测,并分别对二元序列样本集中前向累加和、后向累加和的样本通过率、分布均匀性进行合格判定。

本文件确定二元序列样本集中的样本数量为1 000。

6.2 样本通过率判定

对于每一个随机性检测项目,统计二元序列样本集中 P_value 值大于或等于 α 的样本个数,本文件确定的用于样本通过率检测的显著性水平 $\alpha=0.01$。

记样本数量为 s,当通过某检测项目的样本个数大于或等于 $s\left(1-\alpha-3\sqrt{\frac{\alpha(1-\alpha)}{s}}\right)$ 时,认为该样本集通过该项检测,否则未通过此项检测。例如,样本数量为1 000个,则通过该检测项目的样本个数应大于或等于981。

6.3 样本分布均匀性判定

对于每一个随机性检测项目,二元序列样本集中各样本的 Q_value 值应该在区间[0,1]均匀分布。记样本数量为 s,将区间[0,1]分为 k 个均匀的子区间,统计二元序列样本集中各样本的 Q_value 在 k 个子区间的实际数量 F_i,并利用 χ^2 分布与理论值 s/k 进行比较,计算统计值 $V=\sum_{i=1}^{k}\frac{(F_i-s/k)^2}{s/k}$,并计算 $P_T=igamc\left(\frac{k-1}{2},\frac{V}{2}\right)$。当 P_T 大于或等于 α_T 时,认为该二元序列样本集通过此项目检测;否则,未通过此项目检测。本文件确定的用于样本分布均匀性检测的显著性水平 $\alpha_T=0.0001$,子区间数量 $k=10$。

6.4 随机性检测结果判定

当所有随机性检测项目的检测结果同时满足6.2规定的样本通过率判定和6.3规定的样本分布均匀性判定时,该二元序列样本集通过本文件的随机性检测,否则未通过随机性检测。

附 录 A
（规范性）
样本长度及检测设置

A.1 20 000 比特样本使用的检测方法与检测参数

长度为 20 000 比特的样本按照表 A.1 的随机性检测方法和检测参数进行随机性统计检测。

表 A.1 20 000 比特样本检测设置

序号	随机性检测方法	检测参数
1	单比特频数检测	—
2	块内频数检测	$m=1\ 000$
3	扑克检测	$m=4,8$
4	重叠子序列检测	$m=3,5$
5	游程总数检测	—
6	游程分布检测	—
7	块内最大游程检测	$m=128$
8	二元推导检测	$k=3,7$
9	自相关检测	$d=2,8,16$
10	累加和检测	—
11	近似熵检测	$m=2,5$
12	离散傅立叶检测	——

A.2 1 000 000 比特样本适用的检测方法与检测参数

长度为 1 000 000 比特的样本按照表 A.2 的随机性检测方法和检测参数进行随机性统计检测。

表 A.2 1 000 000 比特样本检测设置

序号	随机性检测方法	检测参数
1	单比特频数检测	—
2	块内频数检测	$m=10\ 000$
3	扑克检测	$m=4,8$
4	重叠子序列检测	$m=3,5$
5	游程总数检测	—
6	游程分布检测	—
7	块内最大游程检测	$m=10\ 000$

表 A.2　1 000 000 比特样本检测设置（续）

序号	随机性检测方法	检测参数
8	二元推导检测	$k=3,7$
9	自相关检测	$d=1,2,8,16$
10	矩阵秩检测	—
11	累加和检测	—
12	近似熵检测	$m=2,5$
13	线性复杂度检测	$m=500,1\ 000$
14	通用统计检测	$L=7,Q=1\ 280$
15	离散傅立叶检测	—

A.3　100 000 000 比特样本检测建议

长度为 100 000 000 比特的样本按照表 A.3 的随机性检测方法和检测参数进行随机性统计检测。

表 A.3　100 000 000 比特样本检测设置

序号	随机性检测方法	检测参数
1	单比特频数检测	—
2	块内频数检测	$m=100\ 000$
3	扑克检测	$m=4,8$
4	重叠子序列检测	$m=3,5,7$
5	游程总数检测	—
6	游程分布检测	—
7	块内最大游程检测	$m=10\ 000$
8	二元推导检测	$k=3,7,15$
9	自相关检测	$d=1,2,8,16,32$
10	矩阵秩检测	—
11	累加和检测	—
12	近似熵检测	$m=5,7$
13	线性复杂度检测	$m=5\ 000$
14	通用统计检测	$L=7,Q=1\ 280$
15	离散傅立叶检测	—

附 录 B
（资料性）
随机性检测原理

B.1 单比特频数检测

单比特频数检测是最基本的检测，用来检测一个二元序列中 0 和 1 的个数是否相近。已知一个长度为 n 的二元序列，令 n_0，n_1 分别表示该序列中 0 和 1 的数目。对一个随机序列，当其长度充分大时，其统计值 V 应服从标准正态分布：

$$V=2\sqrt{n}\left(\frac{n_1}{n}-\frac{1}{2}\right)。$$

B.2 块内频数检测

块内频数检测用来检测待检序列的 m 位子序列中 1 的个数是否接近 $\frac{m}{2}$。对随机序列来说，其任意长度的 m 位子序列中 1 的个数都应该接近 $\frac{m}{2}$。

块内频数检测将待检序列划分成 N 个子序列，每个子序列的长度为 m，有 $n=N*m$。当然，如果 n 不能被 m 整除，必然会有多余位，此时将多余的位舍弃。计算每一个子序列中 1 所占的比例，设为 $\pi_i=\frac{\sum_{j=1}^{m}\varepsilon_{(i-1)m+j}}{m}$，$1\leqslant i\leqslant N$。将所有 N 个子序列中 1 所占的比例的累加和作为统计值

$$V=4m\sum_{i=1}^{N}\left(\pi_i-\frac{1}{2}\right)^2,$$

该统计量应该服从自由度为 N 的 χ^2 分布。

B.3 扑克检测

对任意的正整数 m，长度为 m 的二元序列有 2^m 类。将待检序列划分成 $N=\left\lfloor\frac{n}{m}\right\rfloor$ 个长度为 m 的非重叠的子序列，用 $n_i(1\leqslant i\leqslant 2^m)$ 表示第 i 类子序列类型的个数。扑克检测用来检测这 2^m 类子序列类型的个数是否接近。

统计值 $V=\sum_{i=1}^{2^m}\frac{\left(n_i-\frac{N}{2^m}\right)^2}{\frac{N}{2^m}}=\frac{2^m}{N}\sum_{i=1}^{2^m}n_i^2-N$ 应服从自由度为 2^m-1 的 χ^2 分布。

B.4 重叠子序列检测

对任意的正整数 m，长度为 m 的二元序列有 2^m 类。重叠子序列检测将长度为 n 的待检序列划分成 n 个可重叠的 m 位子序列。对随机二元序列来说，由于其具有均匀性，故 m 位可重叠子序列的每一类模式出现的概率应接近。

在重叠子序列检测中，m 位子序列共有 2^m 类模式，记为 i_1，i_2，…，i_m。令 $v_{i_1i_2\cdots i_m}$ 表示模式为（i_1，i_2，…，i_m）的子序列出现的个数。则统计值

$$\Psi_m^2=\sum_{i_1 i_2\cdots i_m}\frac{\left(v_{i_1 i_2\cdots i_m}-\frac{n}{2^m}\right)^2}{\frac{n}{2^m}}=\frac{2^m}{n}\sum_{i_1 i_2\cdots i_m}\left(v_{i_1 i_2\cdots i_m}-\frac{n}{2^m}\right)^2=\frac{2^m}{n}\sum_{i_1 i_2\cdots i_m}v_{i_1 i_2\cdots i_m}^2-n$$

应该服从 χ^2 类型的分布，但是并不服从 χ^2 分布，因为各 $v_{i_1 i_2\cdots i_m}$ 之间并不独立。

令统计值 $\nabla\Psi_m^2$ 和 $\nabla^2\Psi_m^2$：

$$\nabla\Psi_m^2=\Psi_m^2-\Psi_{m-1}^2,$$

$$\nabla^2\Psi_m^2=\Psi_m^2-2\Psi_{m-1}^2+\Psi_{m-2}^2,$$

其中，$\Psi_0^2=\Psi_{-1}^2=0$，则统计值 $\nabla\Psi_m^2$ 和 $\nabla^2\Psi_m^2$ 应分别服从自由度为 2^{m-1} 和 2^{m-2} 的 χ^2 分布。

B.5 游程总数检测

游程是二元序列的一个子序列，由连续的 0 或者 1 组成，并且其前导和后继元素都与其本身的元素不同。

游程总数检测主要检测待检序列中游程的总数是否服从随机性要求。

令 $V_n(obs)$ 表示待检序列的游程总数，π 表示该序列中 1 所占的比例，则 $V=\dfrac{V_n(obs)-2n\pi(1-\pi)}{2\sqrt{n}\pi(1-\pi)}$ 应服从标准正态分布。

B.6 游程分布检测

连续 1(或 0)的一个游程称为一个块(或一个间断)。根据 Golomb 公设，如果待检二元序列是随机的，则相同长度游程的数目接近一致，且长度为 i 的游程个数约占游程总数的 2^{-i}。

对于 $e_i=\dfrac{n-i+3}{2^{i+2}}$，令 k 为满足 $e_i\geqslant 5$ 的最大整数 i。令 b_i、g_i 分别表示一个二元序列中长度为 i 的“1”游程和“0”游程的数目，$1\leqslant i\leqslant k$。其中，长度超过 k 的游程分别计入 b_k 和 g_k。

游程总个数 $T=\sum_{i=1}^{k}(b_i+g_i)$，长度为 i 的“1”游程和“0”游程的期望值为 $e'_i=T/2^{i+1}$，$1\leqslant i\leqslant k-1$；$e'_k=T/2^k$，构造统计值：

$$V=\sum_{i=1}^{k}\frac{(b_i-e'_i)^2}{e'_i}+\sum_{i=1}^{k}\frac{(g_i-e'_i)^2}{e'_i},$$

该统计值 V 近似地服从自由度为 $2k-2$ 的 χ^2 分布。

B.7 块内最大游程检测

将待检序列划分成 N 个等长的子序列，根据各个子序列中最大“1”/“0”游程的分布来评价待检序列的随机性。

将待检序列划分成 N 个长度为 m 的子序列，此时 $n=N*m$。根据 m 的大小，对应着 $K+1$ 个集合(K 的取值与 m 的大小有关，见表 B.2)，然后计算每个子序列的最大“1”/“0”游程的长度，并按照表 B.3 的规则将其归入相应的集合中。记这 $K+1$ 个集合中的元素个数分别为 v_i，显然有($v_0+v_1+v_2+\cdots+v_K=N$)，统计值 V 应该服从自由度为 K 的 χ^2 分布(其中 π_i 为最大游程长度落入第 i 个集合的概率)：

$$V=\sum_{i=0}^{K}\frac{(v_i-N\pi_i)^2}{N\pi_i}。$$

待检序列 ε 的长度和子序列 m 的长度取值参考表 B.1。K 和 π_i 的取值与 m 有关，表 B.2、表 B.3 和表 B.4 分别给出了当 m 取 8、128 和 10 000 时对应的 K 值大小、集合定义规则以及 π_i 的取值。

表 B.1　ε 和 m 的取值

n	m
≥128	8
≥6 272	128
≥750 000	10 000

表 B.2　K 的取值

m	8	128	10 000
K	3	5	6

表 B.3　集合定义规则

对应集合	$m=8$	$m=128$	$m=10\ 000$
集合 0	≤1	≤4	≤10
集合 1	2	5	11
集合 2	3	6	12
集合 3	≥4	7	13
集合 4		8	14
集合 5		≥9	15
集合 6			≥16

例如，当 $m=10\ 000$ 时，如某个子序列的最大“1”游程长度为 13，应将其归入集合 3，集合 3 中的元素个数 v_3 加 1。

表 B.4　π_i 的取值

π_i	$m=8$	$m=128$	$m=10\ 000$
π_0	0.214 8	0.117 4	0.086 632
π_1	0.367 2	0.243 0	0.208 201
π_2	0.230 5	0.249 4	0.248 419
π_3	0.187 5	0.175 2	0.193 913
π_4		0.102 7	0.121 458
π_5		0.112 4	0.068 011
π_6			0.073 366

B.8　二元推导检测

二元推导序列是由初始序列生成的一个新的序列。对于长度为 n 的二元初始序列，依次将初始序列中两个相邻比特做异或操作，即可得到该序列的一次二元推导序列，长度为 $n-1$。依次执行上述操作 k 次，即可得到该初始序列的 k 次二元推导序列，长度为 $n-k$。

二元推导检测的目的是判定第 k 次二元推导序列中 0 和 1 的个数是否接近一致。令 p_k 为第 k 次

二元推导序列中 1 的比例。统计值 $V=2\sqrt{n-k}\left(\frac{p_k}{n-k}-\frac{1}{2}\right)$ 应该服从标准正态分布。

B.9 自相关检测

自相关检测用来检测待检序列与将其左移(逻辑左移)d 位后所得新序列的关联程度。一个随机序列应该和将其左移任意位所得的新序列都是独立的,故其关联程度也应该很低。

令 $A(d)=\sum_{i=0}^{n-d-1}(\varepsilon_i \oplus \varepsilon_{i+d})$ 表示待检序列与将其左移 d 位后所得新序列之间不同元素的个数,称 d 为时延。

统计值 $V=\frac{2(A(d)-(n-d)/2)}{\sqrt{n-d}}$ 应服从标准正态分布。

B.10 矩阵秩检测

矩阵秩检测用来检测待检序列中给定长度的子序列之间的线性独立性。由待检序列构造矩阵,然后检测矩阵的行或列之间的线性独立性,矩阵秩的偏移程度可以给出关于线性独立性的量的认识,从而影响对序列随机性好坏的评价。

对于一个 $M\times Q$ 矩阵来说,其秩(用 R 表示)可以取 $r=0,1,2,\cdots,m[m=\min(M,Q)]$ 之间的数。对于一个由随机序列构造的 $M\times Q$ 矩阵来说,R 取 r 的概率 p_r 应为:

$$p_r=2^{r(Q+M-r)-MQ}\prod_{i=0}^{r-1}\frac{(1-2^{i-Q})(1-2^{i-M})}{1-2^{i-r}},$$

令 F_M、F_{M-1} 和 $N-F_M-F_{M-1}$ 分别表示秩为 M、$M-1$ 以及秩小于 $M-1$ 的矩阵个数,选取 $M=32$,$Q=32$,则统计值

$$V=\frac{(F_M-0.288\ 8N)^2}{0.288\ 8N}+\frac{(F_{M-1}-0.577\ 6N)^2}{0.577\ 6N}+\frac{(N-F_M-F_{M-1}-0.133\ 6N)^2}{0.133\ 6N},$$

应服从自由度为 2 的 χ^2 分布。其中,0.288 8、0.577 6 和 0.133 6 分别为秩为 32、31 以及小于 31 的矩阵概率,N 为由序列构造的矩阵的总个数。

B.11 累加和检测

累加和检测将待检序列的各个子序列中最大的偏移(与 0 之间),也就是最大累加和与一个随机序列应具有的最大偏移相比较,以判断待检序列的最大偏移是过大还是过小。实际上,随机序列的最大偏移应该接近 0,所以累加和不能过大,也不能过小(累加和可以是负数)。根据最大偏移值来判断待检序列的随机程度。

将待检序列 ε 中的 0 和 1 分别转换为 -1 和 1,构造随机变量 $X_i=2\varepsilon_i-1(1\leqslant i\leqslant n)$,

表 B.5 累加和的计算

前向累加和模式	后向累加和模式
$S_1=X_1$ $S_2=X_1+X_2$ ⋮ $S_k=X_1+X_2+\cdots+X_k$ ⋮ $S_n=X_1+X_2+\cdots+X_k+\cdots+X_n$	$S_1=X_n$ $S_2=X_n+X_{n-1}$ ⋮ $S_k=X_n+X_{n-1}+\cdots+X_{n-k+1}$ ⋮ $S_n=X_n+X_{n-1}+\cdots+X_{n-k+1}+\cdots+X_1$

累加和检测根据 S_i 绝对值的最大值 $\max_{1\leqslant i\leqslant n}|S_i|$ 来检测待检序列的随机性。

根据以下方法计算 P_value：

$$P(\max_{1\leqslant i\leqslant n}|S_i|\geqslant z)=1-\sum_{i=-\infty}^{+\infty}P((4i-1)z<S_n<(4i+1)z)+\sum_{i=-\infty}^{+\infty}P((4i+1)z<S_n<(4i+3)z)$$

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

B.12 近似熵检测

近似熵检测通过比较 m 位可重叠子序列模式的频数和 $m+1$ 位可重叠子序列模式的频数来评价其随机性。近似熵检测是对两个相邻长度的可重叠子序列模式出现频数的检测，设 $Y_i(m)=(\varepsilon_i,\varepsilon_{i+1},\varepsilon_{i+2},\cdots,\varepsilon_{i+m-1})$，对长度为 n 的序列进行检测，需将前 $m-1$ 比特填充到序列尾部，构成 n 个 m 位可重叠子序列，令

$$C_i^m=\frac{1}{n}\quad\{j:1\leqslant j\leqslant n,Y_j(m)=Y_i(m)\}=\pi_l$$

$$\varphi^{(m)}=\sum_{l=1}^{2^m}\pi_l\ln\pi_l$$

式中：C_i^m 表示模式 $Y_i(m)$ 在待检序列中出现的相对频数，π_l 表示模式 $l=(i_1,i_2,\cdots,i_m)$ 在待检序列中出现的相对频数，$-\varphi^{(m)}$ 表示所有 2^m 个 m 位子序列模式相对频数分布的熵。

定义近似熵 $ApEn(m)$ 为：$ApEn(m)=\varphi^{(m)}-\varphi^{(m+1)}$。这里，$ApEn(0)=-\varphi^{(1)}$。

近似熵给出了当子序列长度 m 增加 1 时，m 位可重叠子序列模式和 $m+1$ 位可重叠子序列模式之间的频数之间的差异有多大。因此，小的 $ApEn(m)$ 值说明待检序列具有规则性和连续性；而大的 $ApEn(m)$ 值则表明待检序列具有不规则性和不连续性。

对任意一个 m 来说，可以得到随机序列（不规则序列）的近似熵 $ApEn(m)$ 应近似地等于 $\ln 2$。所以，统计值 $V=2n(\ln 2-ApEn(m))$ 应服从自由度为 2^m 的 χ^2 分布。

B.13 线性复杂度检测

将待检序列划分成 N 个长度为 m 的子序列，此时 $n=N*m$，然后利用 Berlekamp-Massey 算法[1)]计算每个子序列的线性复杂度 L_i，计算 $T_i=(-1)^m(L_i-\mu)+\frac{2}{9}$，其中 $\mu=\frac{m}{2}+\frac{9+(-1)^{m+1}}{36}-\frac{1}{2^m}\left(\frac{m}{3}+\frac{2}{9}\right)$。

选择 $K+1$ 个不相交的独立的集合，然后将各个子序列的 T_i 按集合分类，统计各个集合中出现的 T_i 个数，分别记作 $v_0,v_1,\cdots,v_K$，显然 $v_0+v_1+\cdots+v_K=N$。

统计值 $V=\sum_{i=0}^{K}\frac{(v_i-N\pi_i)^2}{N\pi_i}$ 应服从自由度为 K 的 χ^2 分布。

取 $K=6$，设置 7 个正整数 $v_0,v_1,\cdots,v_6$，将这 7 个正整数的初值都设置为 0，对所有的 $i\in[1,N]$：

如果：$T_i\leqslant-2.5$，v_0 加 1；

$-2.5<T_i\leqslant-1.5$，v_1 加 1；

$-1.5<T_i\leqslant-0.5$，v_2 加 1；

$-0.5<T_i\leqslant0.5$，v_3 加 1；

$0.5<T_i\leqslant1.5$，v_4 加 1；

1) 见 The Handbook of Applied Cryptography；A.Menezes，P.Van Oorschot and S.Vanstone；CRC Press，1997.

$1.5<T_i\leqslant 2.5$，v_5加 1；

$T_i>2.5$，v_6加 1。

其中，对应的 π_i值为：

$\pi_0=0.010\ 417$，$\pi_1=0.031\ 250$，$\pi_2=0.125$，$\pi_3=0.500$，$\pi_4=0.250$，$\pi_5=0.062\ 500$，$\pi_6=0.020\ 833$。

B.14 Maurer 通用统计检测

Maurer 通用统计（简称通用统计）检测主要检测待检序列能否被无损压缩。如果待检序列能被显著地压缩，则认为该序列是不随机的，因为随机序列是不能被显著压缩的。

通用统计检测可以用来检测待检序列多方面的特性，但这并不意味着通用统计检测是前面几个检测的拼装，通用统计检测完全采取了和其他检测所不同的方法，可以检测待检序列某些统计上的缺陷。一个序列可以通过通用统计检测当且仅当这个序列是不可压缩的。

通用统计检测需要的数据量很大，它将序列分成长度为 L 的子序列，然后将待检序列分成两部分：初始序列和检测序列。初始序列包括 Q 个子序列，Q 应该大于或等于 $10*2^L$；检测序列包括 K 个子序列，K 应该大于或等于 $1\ 000*2^L$。因此，序列长度 n 应大于或等于 $10*2^L*L+1\ 000*2^L*L$，而 L 的取值范围应为 $1\leqslant L\leqslant 16$，L 取值不宜小于 6。显然，当 $L=6$ 时，n 至少为 387 840。当序列长度 n 一定时，取 $K=\left\lfloor\frac{n}{L}\right\rfloor-Q$。

首先，从头开始遍历初始序列（以块为单位），找到每一个 L 位模式在初始序列中最后出现的位置（块号），如果一个 L 位模式在初始序列中没有出现，那么将其位置设置为 0；此后，从头开始遍历检测序列，每一次都会得到一个 L 位子序列，计算这个子序列所在的位置与其前面最后一次出现的位置差，也就是块号相减，将相减结果记为距离 len，再对距离 len 求以 2 为底的对数；最后，将所有的求对数的结果相加。这样，就可以得到统计值：

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

计算期望值：

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

实际上，f_n的期望值就是随机变量 $\log_2 G$ 的期望值，其中 $G=G_L$ 是参数为 $1-2^{-L}$ 的几何分布。几何分布的定义为，设一个贝努里实验成功的概率为 p，取随机变量 X 为成功以前进行的独立贝努里实验的次数，那么有：

$$P(X=1)=p$$

$$P(X=2)=(1-p)p$$

并且，对任意的 $x=1,2,\cdots$，有 $P(X=x)=(1-p)^{x-1}*p$。显然，对于几何分布有：

$$\sum_{x=1}^{\infty}P(X=x)=1$$

$$E(X)=\frac{1}{p}$$

标准差按如下计算：

$$\sigma=\sqrt{Var(f_n)}=c(L,K)\sqrt{\frac{Var(\log_2 G)}{K}}$$

这里 $c(L,K)$是一个影响因子，因为必须要考虑到各个模式之间的独立性。本文件采用如下的公式来估计 $c(L,K)$：

$$c(L,K)=0.7-\frac{0.8}{L}+\left(4+\frac{32}{L}\right)\frac{K^{-3/L}}{15}$$

统计值 $V=\dfrac{f_n-E(L)}{\sigma}$ 应服从标准正态分布。

B.15 离散傅立叶检测

离散傅立叶变换检测使用频谱的方法来检测序列的随机性。对待检序列进行傅立叶变换后可以得到尖峰高度，根据随机性的假设，这个尖峰高度不能超过某个门限值(与序列长度 n 有关)，否则将其归入不正常的范围；如果不正常的尖峰个数超过了允许值，即可认为待检序列是不随机的。

首先，将待检序列中的 0 和 1 分别转换成 −1 和 1，用 X 代表新序列，并用 x_k 代表新序列的第 k 位，令

$$f_j=\sum_{k=1}^{n}x_k\exp\left(\frac{2\pi i(k-1)j}{n}\right)=\sum_{k=1}^{n}x_k\left[\cos\left(\frac{2\pi(k-1)j}{n}\right)+i\sin\left(\frac{2\pi(k-1)j}{n}\right)\right]$$

式中：$j=0,1,\cdots,n-1$；$i\equiv\sqrt{-1}$。

基于实数到复数变换的对称性，只需考虑一半的傅立叶系数即可，即 $j=0,1,\cdots,n/2-1$ 的情况，这样可以显著地加快检测速度。用 mod_j 表示 f_j 的系数，根据随机性假设，可以设置一个范围(例如 95%)，即至少应该有 95% 的 mod_j 小于某个门限值，此时门限值应为 $\sqrt{2.995\ 732\ 274*n}$。令 N_1 代表 mod_j 小于门限值的复数的个数，$N_0=0.95*n/2$，统计值 $V=(N_1-N_0)/\sqrt{0.95*0.05*\dfrac{n}{3.8}}$ 应服从标准正态分布。

附 录 C
（资料性）
随机性检测结果示例

C.1 单比特频数检测

输入：ε＝1100110000010101011011000100110011110000000000001001001101010100010001001111010110100000000110101111100110011100110110110000101100101
n＝128
中间值：V＝－1.237 437
输出：P_value＝0.215 925，Q_value＝0.892 038

C.2 块内频数检测

输入：ε＝1100100100001111111011010101000100010000101101000110000100011010011000100110001100110000101000010111000
n＝100
m＝10
中间值：N＝10
V＝7.200
输出：P_value＝0.706 438，Q_value＝0.706 438

C.3 扑克检测

输入：ε＝1100110000010101011011000100110011110000000000001001001101010100010001001111010110100000000110101111100110011100110110110000101100101
n＝128
m＝4
中间值：V＝19.000
输出：P_value＝0.213 734，Q_value＝0.213 734

C.4 重叠子序列检测

输入：ε＝1100110000010101011011000100110011110000000000001001001101010100010001001111010110100000000110101111100110011100110110110000101100101
n＝128
m＝2
中间值：$\nabla\Psi_m^2$＝1.656 250；$\nabla^2\Psi_m^2$＝0.125
输出：P_value1＝0.436 868；P_value2＝0.723 674
Q_value1＝0.436 868；Q_value2＝0.723 674

C.5 游程总数检测

输入：ε＝1100110000010101011011000100110011110000000000001001001101010100010001001111010110100000000110101111100110011100110110110000101100101
n＝128

中间值：$\pi=0.445\ 313$

$V=0.494\ 817$

输出：$P\_value=0.620\ 729$，$Q\_value=0.310\ 364$

C.6 游程分布检测

输入：$\varepsilon=1100110000010101011011000100110011110000000000001001001101010100010001001111010110100000001101011111001100111001101101100010110010$

$n=128$

中间值：$V=0.060\ 606$

输出：$P\_value=0.970\ 152$，$Q\_value=0.970\ 152$

C.7 块内最大游程检测

(1)块内最大“1”游程检测

输入：$\varepsilon=1100110000010101011011000100110011110000000000001001001101010100010001001111010110100000001101011111001100111001101101100010110010$

$n=128$

中间值：$V=4.882\ 605$

输出：$P\_value=0.180\ 598$，$Q\_value=0.180\ 598$

(2)块内最大“0”游程检测

输入：$\varepsilon=1100110000010101011011000100110011110000000000001001001101010100010001001111010110100000001101011111001100111001101101100010110010$

$n=128$

中间值：$V=0.842\ 410$

输出：$P\_value=0.839\ 299$，$Q\_value=0.839\ 299$

C.8 二元推导检测

输入：$\varepsilon=1100110000010101011011000100110011110000000000001001001101010100010001001111010110100000001101011111001100111001101101100010110010$

$n=128$

$k=3$

中间值：$V=-2.057\ 183$

输出：$P\_value=0.039\ 669$，$Q\_value=0.980\ 166$

C.9 自相关检测

输入：$\varepsilon=1100110000010101011011000100110011110000000000001001001101010100010001001111010110100000001101011111001100111001101101100010110010$

$n=128$

$d=1$

中间值：$V=0.266\ 207$

输出：$P\_value=0.790\ 080$，$Q\_value=0.395\ 040$

C.10 矩阵秩检测

输入：$\varepsilon=e$ 的二进制展开的前 1 000 000 比特

$n=1\ 000\ 000$

中间值：$V=2.358\ 278$

输出：$P\_value=0.307\ 543$，$Q\_value=0.307\ 543$

C.11 累加和检测

(1)前向累加和检测

输入：ε＝1100100100001111111011010101000100010000101101000110000100011010011000100110001100110001010001011100 0

$n=100$

输出：$P\_value=0.219\ 194$，$Q\_value=0.219\ 194$

(2)后向累加和检测

输入：ε＝1100100100001111111011010101000100010000101101000110000100011010011000100110001100110001010001011100 0

$n=100$

输出：$P\_value=0.114\ 866$，$Q\_value=0.114\ 866$

C.12 近似熵检测

输入：ε＝1100100100001111111011010101000100010000101101000110000100011010011000100110001100110001010001011100 0

$n=100$

$m=2$

中间值：$V=5.550\ 792$

输出：$P\_value=0.235\ 301$，$Q\_value=0.235\ 301$

C.13 线性复杂度检测

输入：ε＝e 的二进制展开的前 1 000 000 比特

$n=1\ 000\ 000$

$m=1000$

中间值：$V=2.706\ 147$

输出：$P\_value=0.844\ 721$，$Q\_value=0.844\ 721$

C.14 Maurer 通用统计检测

输入：ε＝e 的二进制展开的前 1 000 000 比特

$n=1\ 000\ 000$

中间值：$V=1.074\ 569$

输出：$P\_value=0.282\ 568$，$Q\_value=0.141\ 284$

C.15 离散傅立叶检测

输入：ε＝1100100100001111111011010101000100010000101101000110000100011010011000100110001100110001010001011100 0

$n=100$

中间值：$V=0.447\ 214$

输出：$P\_value=0.654\ 721$，$Q\_value=0.327\ 360$

C.16 样本分布均匀性判定示例

假设有50个样本序列，其*Q_value*值在10个区间上的分布情况如表C.1所示。

表C.1 *Q_value*分布示例

区间	个数	区间	个数
[0,0.1)	2	[0.5,0.6)	5
[0.1,0.2)	5	[0.6,0.7)	2
[0.2,0.3)	8	[0.7,0.8)	8
[0.3,0.4)	7	[0.8,0.9)	9
[0.4,0.5)	2	[0.9,1]	2

基于上表求得的$P_T=0.096\ 578$。