大样本条件下随机性检测的误差分析及参数建议

孙月玥, 范丽敏

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.240700006

大样本条件下随机性检测的误差分析及参数建议

孙月玥[1]　范丽敏[2]
1 北京理工大学数学与统计学院　北京 100081
2 中国科学院软件研究所可信计算与信息保障实验室　北京 100190
(yysunnya@163.com)

摘　要　在信息安全领域,随机性检测在确保密码系统的安全性中起着至关重要的作用。这些测试的稳定性和可靠性直接影响密码系统的整体安全性。检测过程中的误差问题一直是学术界和工业界关注的焦点,特别是在处理大规模样本时,误差的累积更容易导致随机性检测的可靠性降低。因此,研究如何提高随机性检测的准确性和可靠性具有重要意义。GM/T 0005-2021标准中包含了9个具有可变参数的检测项目。针对大样本二元数据的随机性检测问题,根据其特点进行分类,并进行误差量化分析。当待检二元序列比特长度为1×10^8时,GM/T 0005-2021标准中的检测参数建议基本合理。对于Maurer通用统计检测,子序列长度取6时p值误差上界为0.0014928,相较于GM/T 0005-2021中建议的参数表现出更高的准确性。对于线性复杂度检测,更小的子序列长度同样会导致更小的误差。随着样本长度的增加,扩展研究了1×10^9时的参数选择,分析了不同样本长度和参数下的误差,并给出了样本长度为1×10^9时的检测参数建议。

关键词:随机性检测;大样本;误差分析;检测参数;GM/T 0005-2021
中图分类号　TN918.1

Error Analysis and Parameter Recommendations for Randomness Test Under Large Sample Conditions

SUN Yueyue[1] and FAN Limin[2]
1 School of Mathematics and Statistics,Beijing Institute of Technology,Beijing 100081,China
2 Trusted Computing and Information Assurance Laboratory,Institute of Software,Chinese Academy of Sciences,Beijing 100190,China

Abstract　In the field of information security,randomness tests play a crucial role in ensuring the security of cryptographic systems. The stability and reliability of these tests directly impact the overall security of cryptographic systems,making error issues during the testing process a focal point for both academia and industry. Particularly when handling large-scale samples,the accumulation of errors can more readily lead to reliability issues in randomness testing. Consequently,studying methods to enhance the accuracy and reliability of randomness testing is of significant importance. The GM/T 0005-2021 standard outlines 9 tests with variable parameters designed for randomness testing of large binary data samples. This study categorizes these tests according to their characteristics and conducts a quantitative error analysis. Specifically,when the bit length of the binary sequence under test is 1×10^8,the parameters recommended by the GM/T 0005-2021 standard are generally reasonable. For the Maurer universal statistical test,a subsequence length of 6 results in upper bound p-value error of 0.0014928,demonstrating higher accuracy compared to the parameters suggested in the GM/T 0005-2021 standard. Similarly,for the linear complexity test,using smaller subsequence lengths results in smaller errors. With the increase in sample length,this study extends the analysis to parameter selection for a sample length of 1×10^9. It systematically examines the errors associated with different sample lengths and parameter configurations,providing refined parameter recommendations for randomness testing when the sample length reaches 1×10^9.

Keywords　Randomness test,Large sample,Error analysis,Test parameters,GM/T 0005-2021

1　引言

随机性是信息安全领域的一个重要研究课题,尤其在密码学方面具有重要意义。密码算法的随机性在算法设计中扮演着关键角色,随机性检测技术被广泛应用于验证随机数的质量,以确保其满足各种应用的要求。不同的标准和对密码算法随机性的测试要求被提出,相应的随机性测试工具包也随之出现。随机性统计检验是评估随机数质量的主要手段。

到稿日期:2024-07-01　返修日期:2024-10-01
基金项目:国家密码科学基金(2025NCSF02057)
This work was supported by the National Cryptography Science Foundation of China(2025NCSF02057).
通信作者:范丽敏(fanlimin@iscas.ac.cn)

许多国家和组织制定了相关检验标准,用于密码产品的评估,例如美国标准协会的 NIST SP800-22[1],德国 BSI 的 AIS31[2],Diehard[3] 和 TestU01[4],以及我国的 GM/T 0005-2021[5] 等。

统计方法的重要性在于通过对样本的研究来推断总体的特征。利用相对较少的随机数样本即可评估随机数生成算法产生的随机数的安全性,判断其是否存在显著的统计相关性。在大数据的时代背景下,数据的存储和处理更加便捷,这引发了对使用更多数据进行更准确特征获取和深度信息挖掘的思考[6]。然而,在处理大样本数据时,现有随机性统计检验的可靠性和准确性面临挑战。

目前,研究者们已经对随机性检测中的显著性水平、统计量偏差及其对检测结果的影响进行了广泛的探讨。例如,Demirhan 等[7]强调了在随机性检测中,选择和解释预设显著性水平时需特别谨慎,指出低于 0.01 的显著性水平会增加拒绝原假设的难度。Zhu 等[8]针对 NIST SP 800-22 二级检验中分布不一致的问题,提出用 Q 值替代 P 值。实验结果表明,基于 Q 值的方法在相同参数下能够识别出 NIST SP 800-22 未能检测出的统计缺陷。Chen 等[9]进一步对 NIST SP 800-22 测试套件中常用的两级随机性测试进行了两阶段误差分析,针对卡方近似和正态近似给出了 NIST SP 800-22 检验中 p 值偏差的估计。Pareschi[10] 通过 Berry-Esséen 定理[11]对频数检测中二项分布近似正态分布的偏差进行了上界估计,并在二级检验中引入了这种偏差,从而提升了统计量的精度。随机性检测的有效性不仅依赖于统计量本身,还涉及各检测项目之间的独立性。Luengo 等[12]通过推断研究,审查了 NIST SP800-22 套件中的依赖性。针对部分特定的随机性检测项目,研究者们提出了多种改进方法,以提升检测的准确性和有效性。Rukhin[13]将近似熵改进为基于增量对比的方法,证明了在固定模板长度 m 时,其分布收敛于 χ^2 随机变量;而当 m 趋于无穷大时,极限分布为正态分布。针对离散傅里叶变换(Discrete Fourier Transform,DFT)检测中统计量的理论参考分布缺失问题,Iwasaki 等[14]提出了以功率谱方差作为测试统计量的新方法,以检测周期性特征。实验结果显示,该方法较传统 DFT 检测具有更强的检测能力。此外,Iwasaki[15]对 DFT 检测中的统计量方差进行了深入研究,通过 Parseval 定理解释了方差偏差的来源,并推导出在特定假设下的可靠理论方差。特别地,当待检序列较长时,现存的随机性检测方法会出现检测偏差。Haramoto[16]指出,当样本量过大时,NIST SP800-22 中的二级测试会因 p 值计算中的近似误差而错误拒绝表现优异的伪随机数生成器。Akcengiz[17]发现,将为相对较短序列设计的测试应用于长序列时,近似方法的使用会破坏序列的结构。为解决这一问题,Akcengiz 通过详细的计算将现有的随机性检测方法改进为长序列测试,并提出了新的轻量级测试套件用于测试长序列。大多数长序列的随机性检测使用数学近似来计算随机变量的期望值,或将长序列划分为多个短序列再使用统计拟合优度测试来评估序列的随机性,这可能导致结果误差或信息丢失。为了进一步提高长序列测试的准确性,Akcengiz 等[18]在前期研究的基础上,提出了一个新的测试套件(LS-14)来评估长序列的随机性。Chen 等[19]发现,由于 DFT 检测统计量的实际分布与假设的正态分布之间存在偏差,当序列较长或序列数量较大时,即便是已知表现良好的比特序列,也有较高的概率被错误拒绝。为此,他们重构了符合卡方分布的统计量,使得 DFT 检测更适用于长序列。针对 Maurer 通用统计检测及其改进测试近似分布的合理性问题,Hikima 等[20]证明随着被测试序列长度的增加,Maurer 通用统计检测的真实参考分布会收敛于正态分布。然而,这些研究主要针对特定的随机性检测方法进行优化,且未能系统地探讨在大样本条件下参数可变的检测项目的表现。

当样本量足够大时,对于参数可变的随机性检测项目,参数的选取是否合适,以及参数该如何选取以使随机性检测的准确性提高,是值得深入研究的问题。然而,现有研究对此关注较少。本文关注了 GM/T 0005-2021 中 9 项参数可变的随机性检测项目,具体分析了这些随机性检测项目在样本量为 1×10^8 时的误差,从而给出 GM/T 0005-2021 中的参数选取是否合适的结论。同时,考虑到当样本长度增加到 1×10^9 时,GM/T 0005-2021 标准中并未提供各随机性检测项目具体的参数选择建议,本文进一步研究了在样本量为 1×10^9 时,这 9 个随机性检测项目在不同参数设置下的误差表现,并给出了在样本量达到 1×10^9 时各检测项目的参数选择建议,以提高随机性检测项目的准确性和适用性。

本文第 2 章介绍了本文的背景和基础知识,以及文章中用到的符号;第 3 章中结合 GM/T 0005-2021 中给出的参数建议,按照类别分析了其中 9 个随机性检测项目的误差;第 4 章扩展研究了样本量增大到 1×10^9 时参数可变的检测项目的误差,并相应地给出了参数建议;最后总结全文并展望未来。

2 背景和基础知识

2.1 符号及定义

本文中,用 ε 表示待检的二元序列,对应的 ε_i 就表示待检序列的第 i 位;n 表示待检二元序列的比特长度,特别地,取 $n=1\times10^8$ 或 $n=1\times10^9$;m 表示子序列的比特长度;N 表示一个待检测的 n 比特序列中 m 位子序列的个数,即 $N=\lfloor n/m\rfloor$;$\pi_i(i=1,\cdots,N)$ 表示第 i 个子序列中 1 所占的比例;常数 $C=0.4785$;对于二元推导检测,k 为二元推导的次数;对于自相关检测,b 为逻辑左移的位数;M_1(M_2)为二元矩阵秩检测中矩阵的行(列)数;对于 Maurer 通用统计检测,L 表示子序列的长度,Q 和 K 分别为初始序列和测试序列中 L 位非重叠子序列的个数。

2.2 随机性检测和假设检验

随机数发生器是生成随机二元序列的硬件或程序,一般可以分为伪随机数发生器和真随机数发生器。伪随机数发生器通常借助密码算法来生成随机数,例如线性同余发生器、MT 随机数发生器等。与伪随机数发生器不同,真随机数发生器通过不可预测的物理过程来生成随机数,比如环形振荡器、量子随机数发生器等都是基于物理基本原理来生成随机数的,是常见的真随机数发生器[21]。

目前,随机性检测主要通过检查输出序列的随机性来实

现，所采用的检测方法基于假设检验。一个假设检验问题中两个互补的假设称为原假设和备择假设。在随机性检测中，原假设 H_0 假定被测试的序列是随机的，备择假设 H_1 假设被测试的序列不是随机序列。在测试过程中，计算样本序列的相应统计量，并将其与预设的阈值进行比较。由于概率分布与阈值相关，如果统计值超过阈值，从假设检验的角度来看，这种小概率事件不应发生，因此拒绝 H_0，否则接受随机性假设。p 值统计量的值是报告假设检验结果的一种方法。$p(X)$是一个满足对每一个样本点 x，都有 $0 \leqslant p(x) \leqslant 1$ 的检验统计量，如果 $p(X)$的值小，则备择假设 H_1 为真。

2.3 误差分析模型

大多数随机性检测方法（如 GM/T 0005-2021）都是通过近似分布进行的检验，并未使用检验统计量的真实精确分布。这种近似是有必要的，因为大部分检验统计量的真实分布是离散的，其计算复杂度较高，分析难度较大，而近似分布多为连续型分布，计算复杂度较低，分析难度较小。然而，这种近似的可靠性是值得考量的。

GM/T 0005-2021 统计检验包中的检测项目主要使用了两种近似分布，即近似正态分布和近似卡方分布。文献[9]根据 Berry-Esséen 定理及连续性约束，给出了近似正态分布检验和近似卡方分布检验的误差分析方法。具体地，根据 Berry-Esséen 定理[11]，在中心极限定理的假设下，可由一串独立的 Bernoulli 随机变量 $X_i(i=1,2,\cdots,n)$ 构造近似服从正态分布的统计量。该近似正态分布的近似 p 值(p) 与真实 p 值(p_0) 之间的偏差上界为[10]：

$$\sup_{0\leqslant p_0\leqslant 1} |p-p_0| \leqslant 2\frac{CE[|X_i|^3]}{\sigma^3\sqrt{n}} \tag{1}$$

其中，$E[|X_i|^3]$和 σ 分别表示 X_i 的三阶矩和标准差；C 为常数，Tyurin[22]已证明 C 可以取到 0.4785。

Pearson 卡方检验[23]：考虑 n 次独立实验，每次实验结果为 k 个类别中的一类。令向量 $\boldsymbol{P}=(p_1,\cdots,p_k)$ 表示非零的各类别概率，$(D_1,\cdots,D_k)$ 表示各类别的样本观测量，则 Pearson 卡方统计量：

$$V=\sum_{j=1}^{k}\frac{(D_j-np_j)^2}{np_j} \tag{2}$$

近似服从自由度为 $k-1$ 的卡方分布，即 $\chi^2_{(k-1)}$。引入统计量：

$$V'=\sum_{j=1}^{k}\frac{(D_j+\delta_j-np_j)^2}{np_j} \tag{3}$$

其中，$-1\leqslant\delta_j\leqslant 1, j=1,\cdots,k$ 且 $\sum_{j=1}^{k}\delta_j=0$（在下文中若不加特殊说明，均假设观测偏差 δ_j 满足该性质）。可以得到 p 值误差上界估计为[9]：

$$\begin{aligned}|p-p_0| &\leqslant \frac{1}{2}\max_{x>0} f_{\chi^2_{(k-1)}}(x)\,|E(V'-V)|\\ &\leqslant \frac{d_k}{2np_*}f_{\chi^2_{(k-1)}}(k-3)\end{aligned} \tag{4}$$

其中，$d_k\leqslant k$ 表示由于 1 比特变动引起计数变动的类别数，$p_*=\min_{1\leqslant i\leqslant k} p_i$。

3 随机性检测项目的误差分析

表 1 列出了样本量为 1×10^8 时，GM/T 0005-2021 中 15 个检测项目的参数范围以及检验统计量的近似分布和构造方法。这一章将结合 GM/T 0005-2021 中的参数建议，对其中参数可变化的 9 个检测项目分别进行误差上界量化分析。注：表 1 中的参数范围参考了 GM/T 0005-2021 以及 NIST SP800-22。

表 1 15 个随机性检测项目的统计量近似分布及参数范围($n=1\times10^8$)

Table 1 Statistical approximations and parameter ranges for 15 randomness testing metrics($n=1\times10^8$)

检测项目	近似分布	统计量构造方法	参数参考范围
单比特频数检测	半正态分布	二项分布统计量	—
块内频数检测	卡方分布	正态统计量的平方和	$n\geqslant100, m\geqslant20, m>0.01n, N<100$
扑克检测	卡方分布	Pearson 统计量	$m=4,8$
重叠子序列检测	卡方分布	统计量的二次型	$m<[\log_2 n]-2$
游程总数检测	半正态分布	二项分布统计量	$n>100$
游程分布检测	卡方分布	Pearson 统计量	—
块内最大游程检测	卡方分布	Pearson 统计量	—
二元推导检测	半正态分布	二项分布统计量	$k=3,7,15$
自相关检测	半正态分布	二项分布统计量	$b=1,2,8,16,32$
矩阵秩检测	卡方分布	Pearson 统计量	$n\geqslant38*M_1*M_2$
累加和检测	正态分布	随机游走	$n\geqslant100$
近似熵检测	卡方分布	统计量的二次型	$m<[\log_2 n]-5$
线性复杂度检测	卡方分布	Pearson 统计量	$n>10^6, 500<m<5000, N\geqslant200$
Maurer 通用统计检测	半正态分布	随机变量的和	$n\geqslant(Q+K)*L, 6\leqslant L\leqslant16, Q=10*2^L, K=[n/L]-Q\approx1000*2^L$
离散 Fourier 检测	半正态分布	二项分布统计量	$n\geqslant1000$

3.1 半正态分布

3.1.1 二元推导检测

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。所谓二元推导，即依次将二元初始序列中两个相邻比特做异或操作。对初始序列进行 k 次二元推导，即可得到长度为 $n-k$ 的 k 次二元推导序列。对于一个随机的序列，无论进行多少次推导，其 0 和 1 的个数都应该接近一致。

二元推导检测方法的统计量为：

$$V=\frac{S_{n-k}}{\sqrt{n-k}}\sim\mathcal{N}(0,1) \tag{5}$$

其中，S_{n-k} 表示将新序列转换为 -1 和 1 之后的累加和，即 $S_{n-k}=\sum_{i=1}^{n-k}X_i=\sum_{i=1}^{n-k}(2\varepsilon_i^{(k)}-1)$，$\varepsilon_i^{(k)}$ 表示 k 次二元推导序列的第

i 位。该随机性检测项目利用二项分布随机变量构造了近似服从标准正态分布的统计量,我们用式(1)对其 p 值的误差上界进行量化。因为 $X_i \in \{-1,1\}$,在这里 X_i 的三阶矩与标准差均为 1,所以 p 值误差上界为:

$$|p-p_0| \leqslant \frac{2C}{\sqrt{n}} \tag{6}$$

若利用式(4),可得到 p 值的误差上界为:

$$|p-p_0| \leqslant \frac{\phi(0)}{\sqrt{n-k}} \tag{7}$$

事实上,此时 $V'=\frac{S_{n-k}+\delta}{\sqrt{n-k}}$,其中 $-1 \leqslant \delta \leqslant 1$,且统计量 V 近似服从标准正态分布。故,由式(4)可得式(7)。当 n 固定为 1×10^8 时,若用式(1),得 p 值误差上界为 0.0000957;若用式(4),取 $k=3,7,15$,误差上界均近似为 0.0000399。

通过这些计算结果可以看出,在样本量 n 较大的情况下,二元推导次数 k 的变化对误差上界的影响相对较小。这表明,当样本量足够大时,二元推导的次数在一定范围内变动并不会显著改变检测方法的准确性。进一步分析,根据误差上界的估计式,我们可以观察到,较小的二元推导次数 k 保证了较低的 p 值误差上界。这一现象背后的原因是,在推导次数较少时,序列相对较长,序列中的随机性信息得以更好地保留,从而使得检测的误差上界相对较小。因此,GM/T 0005-2021 标准中对 k 值的选择是合理且有效的,这样的选择既能确保足够的随机性检测能力,又能控制误差上界在一个较低的范围内。

3.1.2 自相关检测

自相关检测用来检测待检序列与将其逻辑左移 b 位后所得新序列的关联程度。一个随机序列和将其逻辑左移任意位所得的新序列应该都是独立的,故其关联程度也应该很低,即初始序列与将其左移 b 位后所得新序列进行异或操作形成的新序列中 0 和 1 的个数应该接近。该检测方法的统计量为:

$$V=\frac{2(A(d)-((n-d)/2))}{\sqrt{n-b}} \sim \mathcal{N}(0,1) \tag{8}$$

其中,$A(d)=\sum_{i=0}^{n-b-1}(\varepsilon_i \oplus \varepsilon_{i+b})$。类似于式(4),得到 p 值的误差上界估计式为:

$$|p-p_0| \leqslant \frac{2\phi(0)}{\sqrt{n-b}} \tag{9}$$

当 n 取 1×10^8,b 取 1,2,8,16,32 时,误差上界均近似为 0.0000798。

通过这些结果可以看出,逻辑左移位数 b 的变化对误差上界的影响相对较小,表明在随机序列中,初始序列与逻辑左移后的新序列之间的独立性得以很好地保持。因此,在较大的样本量下,自相关检测方法对左移位数的选择较为宽松,不同的 b 值不会显著影响检测结果的准确性。与二元推导检测类似,较小的逻辑左移位数 b 不仅能有效降低检测中的误差上界,还能保持检测过程的简洁性和高效性,因此 GM/T 0005-2021 中的参数选择是较合理的。

3.1.3 Maurer 通用统计检测

Maurer 通用统计检测用于检测待检序列能否被无损压缩。因为随机序列不能被显著压缩,所以如果待检序列能被显著地压缩,则认为该序列不随机。该检测方法的统计量为:

$$V=\frac{\frac{sum}{K}-E(L)}{\sigma} \sim \mathcal{N}(0,1) \tag{10}$$

其中,$sum=\sum_{i=Q+1}^{Q+K}\log_2(i-T_j)$,$j$ 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,$E(L)=2^{-L}\sum_{i=1}^{+\infty}(1-2^{-L})^{i-1}\log_2 i$,整数 Q 和 K 分别表示初始序列和测试序列中包含的 L 位非重叠子序列的数量总和,$\sigma=c(L,K)\sqrt{\frac{Var(\log_2 G)}{K}}$,这里 $G=G_L$ 是参数为 $1-2^{-L}$ 的几何分布,$c(L,K)$是一个影响因子,Coron 和 Naccache[24] 给出了其准确值。类似于式(4),我们得到 p 值的误差上界估计式为:

$$|p-p_0| \leqslant \phi(0) \cdot \frac{d_L}{K \cdot c(L,K)\sqrt{Var[a_n]/K}} \tag{11}$$

其中,$d_L=L \cdot \log_2(L)$。同样地,Coron 和 Naccache[24] 给出了 $Var[a_n]$的准确值。对于样本量 $n=1\times10^8$,参考 NIST SP 800-22 以及 GM/T 0005-2021 给出的参数建议,分别选取子序列长度 $6 \leqslant L \leqslant 12$,对应的初始序列及测试序列个数分别为 $Q=10*2^L$,$K=\lfloor n/L \rfloor - Q$。

表 2 展示了子序列长度 L 变化时 Maurer 通用统计检测 p 值的误差上界。根据表 2 第二列,p 值误差上界随着 L 的增加而增大。因此,在进行 Maurer 通用统计检测时,选择较小的 L 值可以有效降低误差上界,确保检测的准确性。例如,表 2 中,当 $L=6$ 时,p 值误差上界仅为 0.0014928;而当 $L=12$ 时,误差上界显著增加至 0.0195567。这表明,在实际应用中,为了维持检测方法的精确性和可靠性,选择较小的子序列长度是至关重要的。对比子序列长度为 6 和 7 的误差上界,GM/T 0005-2021 中的参数选择($L=7$) 是较合理的。

表 2 Maurer 通用统计检测 p 值的误差上界

Table 2 Upper bound of p-value error in Maurer's universal statistical test

L	$n=1\times10^8$	$n=1\times10^9$
16		0.0028044
15		0.0024808
14		0.0021792
13		0.0018973
12	0.0195567	0.0016343
11	0.0044019	0.0013906
10	0.0036905	0.0011665
9	0.0030444	0.0009625
8	0.0024635	0.0007789
7	0.0019470	0.0006157
6	0.0014928	0.0004720

3.2 卡方分布

3.2.1 扑克检测

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。该检

测方法将待检序列划分成 N 个长度为 m 的非重叠子序列，通过统计各类子序列模式出现的频数来判断待检序列的随机性。扑克检测方法的检验统计量为：

$$V=\sum_{i=1}^{2^m}\frac{\left(n_i-\frac{N}{2^m}\right)^2}{\frac{N}{2^m}}=\frac{2^m}{N}\sum_{i=1}^{2^m}n_i^2-N\sim\chi^2_{(2^m-1)} \tag{12}$$

类似于式(4)，得到 p 值误差上界的估计式为：

$$|p-p_0|\leqslant f_{\chi^2_{(2^m-1)}}(2^m-3)\cdot\frac{2^m}{N} \tag{13}$$

当 $n=1\times10^8$ 时，分别取子序列长度 $m=4,8$，得到 p 值误差上界分别为 0.000 000 05 和 0.000 000 4。

通过这些结果可以看出，随着子序列长度 m 的增加，p 值误差上界有所增大，但总体上仍保持在一个非常低的范围内。这表明，无论是较小的 m 值还是较大的 m 值，在扑克检测中都能表现出较高的检测准确性。这进一步验证了 GM/T 0005-2021 标准中参数选择的合理性。

3.2.2 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，从而影响对二元序列随机性好坏的评价。构造统计值：

$$V=\frac{(F_M-0.2888N)^2}{0.2888N}+\frac{(F_{M-1}-0.5776N)^2}{0.5776N}+\frac{(N-F_M-F_{M-1}-0.1336N)^2}{0.1336N}\sim\chi^2_{(2)} \tag{14}$$

其中，$N=\left\lfloor\frac{n}{M_1*M_2}\right\rfloor$，即 n 比特长的待检序列被分为 N 个 M_1*M_2 的二元矩阵，特别地，取 $M_1=M_2=M$。F_M 表示秩为 M 的矩阵的个数，F_{M-1} 表示秩为 $M-1$ 的矩阵的个数，$N-F_M-F_{M-1}$ 表示秩小于 $M-1$ 的矩阵的个数。V 是一个自由度为 2 的 Pearson 卡方统计量。类似于式(4)，得到 p 值的误差上界估计为：

$$|p-p_0|\leqslant\frac{1}{2}F_{\chi^2_{(2)}}\left(\frac{dM^2}{0.1336n}\right) \tag{15}$$

其中，$d\leqslant2$。根据式(15)，该检测方法的 p 值误差上界随着矩阵的行(列)数增加而增大。当样本量 n 为 1×10^8 时，取 $M=32$，得到误差上界为 0.000 038 3。

上述结果表明，即使在较大的矩阵尺寸下，检测方法依然能够提供精确的评估结果，说明随机性检测标准 GM/T 0005-2021 中的参数选择较为合理。

3.2.3 线性复杂度检测

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。构造统计值：

$$V=\sum_{i=0}^{K}\frac{(v_i-N\pi_i)^2}{N\pi_i}\sim\chi^2_{(K)} \tag{16}$$

其中，$v_i(i=0,\cdots,K)$ 是计数变量。类似于式(4)，得到 p 值的误差上界为：

$$|p-p_0|\leqslant f_{\chi^2_{(K)}}(K-2)\cdot\frac{d_K}{2N\pi_*} \tag{17}$$

其中，$\pi_*=\min_{0\leqslant i\leqslant K}\pi_i$，$d_K\leqslant2$，$K=6$。因为 $N=\lfloor n/m\rfloor$，所以该检验的误差上界随着子序列长度 m 的增加而增大。取 $m=5\,000$ 时，误差上界为 0.000 649 6；取 $m=1\,000$，$m=500$ 时，误差上界分别为 0.000 129 9，0.000 065。

这表明，虽然子序列长度 m 增大时误差上界增加，但在合理的 m 值范围内，误差上界的波动幅度并不大。考虑到检验的效率，GM/T 0005-2021 中对 $m(m=5\,000)$ 的选择是较为合理的。

3.2.4 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 $m/2$。对于随机序列来说，其任意长度的 m 位子序列中 1 的个数都应该接近 $m/2$。该检测方法的统计量为：

$$V=4m\sum_{i=1}^{N}(\pi_i-1/2)^2\sim\chi^2_{(N)} \tag{18}$$

其中，$\pi_i(i=1,\cdots,N)$ 表示第 i 个子序列中 1 所占的比例，这是一个标准正态分布随机变量的平方和构造，服从自由度为 N 的卡方分布。类似于式(4)，得到 p 值的误差上界估计式为：

$$|p-p_0|\leqslant\frac{2}{m}f_{\chi^2_{(N)}}(N-2) \tag{19}$$

由式(19)可见，块内频数检测的 p 值误差上界随着子序列长度 m 的增加而减小。这一现象表明，当选择较大的 m 值时，每个子序列中 1 的个数与 $m/2$ 的偏差会更接近正态分布的预期，从而提高检测的精确度。具体来说，当 $n=1\times10^8$ 时，取 $m=100\,000$，得到误差上界为 0.000 000 2，这说明了该检测方法较高的准确度以及 GM/T 0005-2021 中参数建议的合理性。

3.2.5 重叠子序列检测

对于任意的正整数 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。对于随机二元序列来说，由于其具有均匀性，因此 m 位可叠加子序列的每一类模式出现的概率应该接近。该检测的统计量为：

$$\begin{aligned}&\nabla\Psi_m^2=\Psi_m^2-\Psi_{m-1}^2\sim\chi^2_{(2^{m-1})}\\&\nabla^2\Psi_m^2=\Psi_m^2-2\Psi_{m-1}^2+\Psi_{m-2}^2\sim\chi^2_{(2^{m-2})}\end{aligned} \tag{20}$$

其中：

$$\begin{aligned}\Psi_m^2&=\sum_{i_1,i_2,\cdots,i_m}\frac{\left(v_{i_1,i_2,\cdots,i_m}-\frac{n}{2^m}\right)^2}{\frac{n}{2^m}}\\&=\frac{2^m}{n}\sum_{i_1,i_2,\cdots,i_m}\left(v_{i_k,\cdots,i_m}-\frac{n}{2^m}\right)^2\\&=\frac{2^m}{n}\sum_{i_1,i_2,\cdots,i_m}v_{i_k,\cdots,i_m}^2-n\end{aligned} \tag{21}$$

$v_{i_1,i_2,\cdots,i_m}$ 表示模式为 $(i_1,i_2,\cdots,i_m)$ 的子序列出现的次数。令 $v'_{i_1,\cdots,i_m}=v_{i_1,\cdots,i_m}+\delta_{i_1,\cdots,i_m}$，则：

$$E(\nabla\Psi_m'^2-\nabla\Psi_m^2)\leqslant\frac{2^m}{n}\sum_{i_1,i_2,\cdots,i_m}\delta^2_{i_1,i_2,\cdots,i_m} \tag{22}$$

$$E(\nabla^2\Psi_m'^2-\nabla^2\Psi_m^2)\leqslant\frac{2^m}{n}\sum_{i_1,i_2,\cdots,i_m}\delta^2_{i_1,i_2,\cdots,i_m}+\frac{2^{m-2}}{n}\sum_{i_1,i_2,\cdots,i_{m-2}}\delta^2_{i_1,i_2,\cdots,i_{m-2}} \tag{23}$$

类似于式(4),该检验的 p 值误差上界为:

$$|p-p_0|\leqslant f_{\chi^2_{(2^{m-1})}}(2^{m-1}-2)\cdot\frac{d_m 2^m}{2n}+f_{\chi^2_{(2^{m-2})}}(2^{m-2}-2)\cdot\frac{d_m 2^m}{2n}+f_{\chi^2_{(2^{m-2})}}(2^{m-2}-2)\cdot\frac{d_{m-2}2^{m-2}}{2n} \tag{24}$$

其中,$d_m\leqslant m(d_{m-2}\leqslant m-2)$。因为子序列是可重叠的,子序列长度为 m,所以 1 比特变动最多可导致 m 类发生计数变动。

从式(24)可以看出,p 值的误差上界与子序列长度 m 直接相关。当样本长度 n 取 1×10^8 时,根据 NIST SP 800-22 的参数建议,子序列长度 $m<24$。对于 GM/T 0005-2021 中建议的$m=3,5,7$,对应的误差上界为 0.000 000 09,0.000 000 2,0.000 000 4。该检测的 p 值误差上界随着 m 的增加而略有提升,但总体上在 GM/T 0005-2021 推荐的参数设置下仍然处于极低水平。因此,GM/T 0005-2021 的参数选择是较为合理的。

3.2.6 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。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。对于任意一个 m 来说,随机序列的近似熵应该近似等于 $\ln 2$。构造统计量:

$$V=2n(\ln 2-ApEn(m))\sim\chi^2_{(2^m)} \tag{25}$$

其中,$ApEn(m)$表示 $m+1$ 位子序列模式与 m 位子序列模式相对频数分布的熵之差。该检验的分析过程与重叠子序列类似,所以该检测方法的 p 值误差上界为:

$$|p-p_0|\leqslant f_{\chi^2_{(2^m)}}(2^m-2)\cdot\frac{d_{m+1}2^{m+1}}{2n} \tag{26}$$

其中,$d_{m+1}\leqslant m+1$。

与重叠子序列检测方法类似,当样本长度 n 取 1×10^8 时,子序列长度 $m<21$。分别取 GM/T 0005-2021 建议的 $m=5,7$,得到误差上界分别为 0.000 000 1,0.000 000 3。这些结果表明,在 GM/T 0005-2021 建议的参数下,检测方法具有较高的精度,能够在极低的误差范围内评估序列的随机性。因此,GM/T 0005-2021 的参数选择是较为合理的。

4 实验结果和分析

4.1 $n=1\times10^8$ 时检测项目的参数建议

表 3 列出了当样本长度为 1×10^8 时,9 个随机性检测项目取不同参数对应的 p 值误差上界。其中非加粗字体为 GM/T 0005-2021 中建议的参数,加粗字体为满足表 1 要求的其他参数选择。

在采用 GM/T 0005-2021 中建议的参数时,大部分检测项目的 p 值误差上界均较小,表现出较高的准确性,所以 GM/T 0005-2021 中的参数选取基本合理。对于 Maurer 通用统计检测,选择 $L=6$ 表现出更高的准确性。当子序列长度 m 减小时,线性复杂度检测的 p 值误差上界也在减小。因此,对于线性复杂度检测,建议选择更小的 m 值。

表 3 不同参数对应的误差上界($n=1\times10^8$)

Table 3 Upper bound of errors for different parameters($n=1\times10^8$)

随机性检测项目	检测参数	p 值误差上界
二元推导检测	$k=3,7,15$	0.000 039 90
自相关检测	$b=1,2,8,16,32$	0.000 079 80
Maurer 通用统计检测	$L=7$	0.001 947 00
	$L=6$	**0.001 492 80**
扑克检测	$m=4$	0.000 000 05
	$m=8$	0.000 000 40
矩阵秩检测	$M=32$	0.000 038 30
线性复杂度检测	m=5 000	0.000 649 60
	$m=1\,000$	**0.000 129 90**
	$m=500$	**0.000 065 00**
块内频数检测	$m=100\,000$	0.000 000 20
重叠子序列检测	$m=3$	0.000 000 10
	$m=5$	0.000 000 20
	$m=7$	0.000 000 40
近似熵检测	$m=5$	0.000 000 10
	$m=7$	0.000 000 30

4.2 $n=1\times10^9$ 时的误差分析及参数建议

GM/T 0005-2021 标准中的大部分随机性检测项目的参数范围可结合 NIST SP800-22 来确定。然而,对于扑克检测、二元推导检测以及自相关检测,GM/T 0005-2021 中仅提供了样本量为 1×10^8 时的参数选择建议,参数选择的参考依据有限。为解决这一问题,在样本量扩展到 1×10^9 时,我们结合 GM/T 0005-2021 中的参数规律,适当扩充了参数选择范围,并通过误差分析实验逐步验证和优化,以确定合理且准确的参数。本节旨在从 p 值误差的角度,给出第 3 章中 9 个随机性检测项目在样本长度为 1×10^9 时的参数建议。在这一节中,如无特殊说明,均默认样本长度 n 为 1×10^9。

对于二元推导检测以及自相关检测,p 值误差上界对二元推导次数 k 或逻辑左移位数 b 并不敏感,且理论上较小的 k 或 b 更有利于提高检测方法的准确性以及效率,所以建议取 $k=3,7,15,31$,$b=1,2,8,16,32,64$。

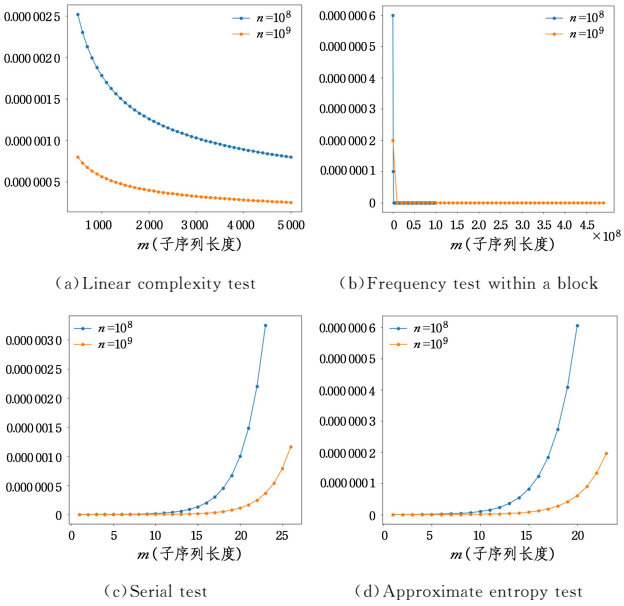
对于 Maurer 通用统计检测,表 2 第三列给出了不同子序列长度 L 对应的 p 值误差上界,同样地,误差随着 L 的增加而增大。因此,建议选择较小的 L 进行检测,例如 $L=6,7,8,9$。

扑克检测在 $n=1\times10^8$ 时就表现出较好的准确性,且其误差上界表达式主要由指数函数 2^m 决定,所以较小的 m 可以保证较小的误差。建议选择 $m=4,8,12$。

在矩阵秩检测中,矩阵的行(列)数越少,误差越小。当 $M=32$ 时,p 值误差上界为 0.000 003 8,误差较小。结合 NIST 的参数建议,建议选择 $M=32$。

图 1 分别展示了线性复杂度检测、块内频数检测、重叠子序列检测以及近似熵检测在样本量分别为 1×10^8 和 1×10^9 时 p 值误差上界随子序列长度 m 的变化情况。在同样的子序列长度取值下,这 4 种检测算法表现的趋势均为样本长度越大,误差越小,这表明样本长度的增大有利于提高检测的准确性。线性复杂度检测和块内频数检测的误差均随着 m 的增大而减小,重叠子序列检测和近似熵检测则相反。因此,当

样本长度达到 1×10^9 时，对于线性复杂度检测以及块内频数检测，建议选择更大的子序列长度 m；而对于重叠子序列检测以及近似熵检测，建议选择较小的 m。但我们同时也注意到，相较于 $n=1\times10^8$，在 NIST SP 800-22 以及 GM/T 0005-2021 建议的参数范围内，$n=1\times10^9$ 时各检测项目的误差都是较小的，所以考虑到检测时的效率问题以及其他问题，m 可在一定范围内灵活调整。



(a) Linear complexity test (b) Frequency test within a block

(c) Serial test (d) Approximate entropy test

图 1 不同子序列长度下的 p 值误差上界

Fig. 1 Upper bound of p-value errors for different subsequence lengths

综上，我们给出 $n=1\times10^9$ 时随机性检测项目的参数选取建议，如表 4 所列。

表 4 $n=1\times10^9$ 的参数建议

Table 4 Parameter suggestions for $n=1\times10^9$

随机性检测项目	检测参数
二元推导检测	$k=3,7,15,31$
自相关检测	$b=1,2,8,16,32,64$
Maurer 通用统计检测	$L=6,7,8,9$
扑克检测	$m=4,8,12$
矩阵秩检测	$M=32$
线性复杂度检测	$m=5000$
块内频数检测	$m=10000000$
重叠子序列检测	$m=3,5,7,9,11,13$
近似熵检测	$m=5,7,9,11$

为了更清楚地展示样本数量对随机性检测的影响，表 5 中详细展示了在不同样本数量下不同的参数选择机制及其对应的误差。从表 5 中可以看到，当样本数量不同时，各随机性检测项目参数的选择机制是不同的，例如重叠子序列检测和近似熵检测，样本数量越大，参数可选择范围越广。同时，对于同一个检测项目，不同的样本数量会导致误差不同，样本量越大，误差越小。根据极限定理，各检测项目统计量的真实分布在样本量趋于无穷时近似服从正态分布或卡方分布。

相比于文献[10,13,15]对单一随机性检测项目的研究，本文的研究范围更广，主要集中于 GM/T 0005-2021 标准中参数可变的 9 个检测项目。与文献[8-9]侧重误差估计的研究不同，本文重点探讨了大样本条件下的参数选择问题。通过实验验证，本文不仅弥补了现有文献中关于大样本条件下 GM/T 0005-2021 标准参数选择的空白，还提供了不同参数设置下检测误差表现的实证结果，为大规模应用提供了更具实用性的建议。

表 5 不同样本量下随机性检测项目的误差

Table 5 Errors of randomness test items under different sample sizes

检测项目	检测参数	p 值误差上界		
		$n=1\times10^6$	$n=1\times10^8$	$n=1\times10^9$
二元推导检测	$k=3,7$	0.0003989		
	$k=3,7,15$		0.0000399	
	$k=3,7,15,31$			0.0000126
自相关检测	$b=1,2,8,16$	0.0007979		
	$b=1,2,8,16,32$		0.0000798	
	$b=1,2,8,16,32,64$			0.0000252
Maurer 通用统计检测	$L=6$	0.0149562	0.0014928	0.0004720
	$L=7$	0.0195567	0.0019470	0.0006157
	$L=8$		0.0024635	0.0007789
	$L=9$		0.0030444	0.0009625
	$L=10$		0.0036905	0.0011665
	$L=11$		0.0044019	0.0013906
	$L=12$		0.0195567	0.0016343
	$L=13$			0.0018973
	$L=14$			0.0021792
	$L=15$			0.0024808
	$L=16$			0.0028044
扑克检测	$m=4$	0.000004900	0.000000050	0.000000005
	$m=8$	0.0000363 0	0.000000 40	0.000000 04
	$m=12$			0.000000 2
矩阵秩检测	$M=32$	0.0038177	0.0000383	0.0000038
线性复杂度检测	$m=500$	0.0064959	0.0000650	0.0000065
	$m=1000$	0.0129918	0.0001299	0.0000130
块内频数检测	$m=5000$	0.0649589	0.0006496	0.0000650
	$m=10000$	0.0000057		
	$m=100000$	0.0000020	0.0000002	
	$m=1000000$		0.000000 10	0.000000 02
	$m=10000000$			0.000000006
重叠子序列检测	$m=3$	0.0000087	0.0000001	0.000000 01
	$m=7$	0.0000431 0	0.000000 40	0.000000 04
近似熵检测	$m=16$	0.0022253	0.0000223	0.0000022
	$m=23$		0.0003637	0.0000364
	$m=26$			0.0001164
	$m=2$	0.0000022		
	$m=7$	0.0000257 0	0.000000 30	0.000000 03
	$m=13$	0.0003575	0.0000036	0.0000004
	$m=20$		0.0000607	0.0000061
	$m=23$			0.0000196

结束语 本文针对 GM/T 0005-2021 标准中参数可变的 9 个检测项目，按照检验统计量的构造方法将其分为两类，对各检测项目的误差进行了量化分析，并从检测误差的角度对 GM/T 0005-2021 中的参数选取是否合适进行了验证及评价。特别地，在样本量达到 1×10^9 的情况下，针对 GM/T 0005-2021 标准中未提供参数选择建议的问题，我们扩展了研究，分析了在样本量为 1×10^9 时这 9 个检测项目的误差随参数变化的情况，并提出了有针对性的参数建议，填补了这一参数空白。本文的研究不仅验证了 GM/T 0005-2021 标准中部分参数选择的合理性，还提出了适用于更大样本量以及其他随机性检测标准的误差分析和参数选择的参考方法，具有重要的实际应用价值。未来可以进一步探讨更合理、紧凑的误差估计方法，并完善各个检测项目的参数选择机制和检测方

法,以进一步降低误差,提高随机性检测的准确性。这些研究将为提升密码学领域的随机性检测标准提供重要的理论支持和实践参考。

参 考 文 献

[1] BASSHAM L E, RUKHIN A L, SOTO J, et al. A Statistical Test Suite for Randomand Pseudorandom Number Generators for Cryptographic Applications: Technical Report [EB/OL]. https://tsapps. nist. gov/publication/get_pdf. cfm? pub_id = 906762.

[2] ILLMANN W, SCHINDLER W. A Proposal for: Functionality Classes for Random Number Generators [EB/OL]. https://www. bsi. bund. de/EN/Home/home_node. htm.

[3] MARSAGLIA G. The Marsaglia Random Number Cdrom Including the Diehard Battery of Tests of Randomness[EB/OL]. https://ani. stat. fsu. edu/diehard/.

[4] L'ECUYER P, SIMARD R. Testu01: Ac Library for Empirical Testing of Random Number Generators[J]. ACM Transactions on Mathematical Software(TOMS), 2007, 33(4): 1-40.

[5] 随机性检测规范:GM/T 0005-2021 [S]. 北京:国家密码管理局,2021.

[6] MAO C, SONG Y, CHEN J. A Lightweight Adaptive Random Testing Method for Deep Learning Systems[J]. Software: Practice and Experience, 2023, 53(11): 2271-2295.

[7] DEMIRHAN H, BITIRIM N. Statistical Testing of Cryptographic Randomness [J]. İstatistikçiler Dergisi: İstatistik ve Aktüerya, 2016, 9(1): 1-11.

[8] ZHU S, MA Y, LIN J, et al. More Powerful and Reliable Second-Level Statistical Randomness Tests for NIST SP 800-22 [C]// Advances in Cryptology-ASIACRYPT 2016: 22nd International Conference on the Theory and Application of Cryptology and Information Security, Hanoi, Vietnam, December 4-8, 2016, Proceedings, Part I 22. Berlin Heidelberg: Springer, 2016: 307-329.

[9] CHEN D, CHEN H, FAN L, et al. Error Analysis of NIST SP 800-22 Test Suite[J]. IEEE Transactions on Information Forensics and Security, 2023, 18: 3745-3759.

[10] PARESCHI F, ROVATTI R, SETTI G. OnStatistical Tests for Randomness Included in the NIST SP 800-22 Test Suite and Based on the Binomial Distribution[J]. IEEE Transactions on Information Forensics and Security, 2012, 7(2): 491-505.

[11] CHEN L H Y, GOLDSTEIN L, SHAO Q M. Normal Approximation by Stein's Method [M]. Berlin Heidelberg: Springer, 2010.

[12] LUENGO E A, OLIVARES B A, VILLALBA L J G, et al. Further Analysis of the Statistical Independence of the NIST SP 800-22 Randomness Tests[J]. Applied Mathematics and Computation, 2023, 459: 128222.

[13] RUKHIN A L. Approximate Entropyfor Testing Randomness [J]. Journal of Applied Probability, 2000, 37(1): 88-100.

[14] IWASAKI A, UMENO K. Randomness Test to Solve Discrete Fourier Transform Test Problems[J]. IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, 2018, 101(8): 1204-1214.

[15] IWASAKI A. Deriving the Variance of the Discrete Fourier Transform Test Using Parseval's Theorem[J]. IEEE Transactions on Information Theory, 2019, 66(2): 1164-1170.

[16] HARAMOTO H. Study on Upper Limit of Sample Size for a Two-Level Test in NIST SP800-22[J]. Japan Journal of Industrial and Applied Mathematics, 2021, 38(1): 193-209.

[17] AKCENGIZ Z. A New Lightweight Statistical Randomness Test Suite and Its Evaluationby Comparison with Other Test Suites [D]. Ankara: Middle East Technical University, 2021.

[18] AKCENGIZ Z, ASLAN M, DOĞANAKSOY A, et al. LS-14 Test Suite for Long Sequences[J]. Hacettepe Journal of Mathematics and Statistics, 2024, 53(1): 230-250.

[19] CHEN M, CHEN H, FAN L, et al. A New Discrete Fourier Transform Randomness Test[J]. Science China Information Sciences, 2019, 62: 1-16.

[20] HIKIMA Y, IWASAKI A, UMENO K. The Reference Distributions of Maurer's Universal Statistical Test and Its Improved Tests[J]. IEEE Transactions on Information Theory, 2021, 68(4): 2674-2683.

[21] MENG C, CAI M, YANG Y, et al. Generation of True Quantum Random Numbers with on-Demand Probability Distributions via Single-Photon Quantum Walks [J]. Optics Express, 2024, 32(11): 20207-20217.

[22] IL'YA S T. Refinement of the Upper Bounds of the Constants in Lyapunov's Theorem[J]. Russian Mathematical Surveys, 2010, 65(3): 586.

[23] PEARSON K X. On the Criterion That a Given System of Deviations from the Probable in the Case of a Correlated System of Variables is Such That it Can Be Reasonably Supposed to Have Arisen from Random Sampling[J]. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 1900, 50(302): 157-175.

[24] CORON J S, NACCACHE D. An Accurate Evaluation of Maurer's Universal Test[C]// International Workshop on Selected Areas in Cryptography. Berlin, Heidelberg: Springer Berlin Heidelberg, 1998: 57-71.

SUN Yueyue, born in 1999, postgraduate. Her main research interests include statistical test of randomness and so on.

FAN Limin, born in 1978, Ph.D, senior engineer. Her main research interests include side channel analysis and protection, and password detection.

(责任编辑:柯颖)