Computer Science and Application 计算机科学与应用, 2022, 12(10), 2370-2383
Published Online October 2022 in Hans. http://www.hanspub.org/journal/csa
https://doi.org/10.12677/csa.2022.1210243

Hans汉斯

分组密码随机序列的不变量研究

罗钰舒，彭圣宇，余愉先，郑智捷

云南大学国家示范性软件学院，云南 昆明

收稿日期：2022年9月22日；录用日期：2022年10月19日；发布日期：2022年10月27日

摘　要

对于分组密码而言，密码的安全性来自混淆、扩散性，而两者主要来自分组密码的轮函数操作。为满足加密算法标准化、加密算法本土化及各方的需求，本文使用图论的相关知识，利用最新向量逻辑——变值体系来对分组密码安全性进行研究。文章通过获取分组密码的中间随机序列，分析随机序列的特征，得到随机序列的不变量，进一步对不变量进行统计分析研究，研究方向为通过控制变量法，观察对比不变量统计可视化结果，探索不变量特征；对于同一分组密码而言，改变不变量数据量，得出同一密码不变量具有饱和态的特性的结论，为分组密码的安全性提供进一步的理论依据。

关键词

分组密码，随机序列，不变量，统计，图论，变值体系

Research on Invariant of Random Sequences in Block Ciphers

Yushu Luo, Shengyu Peng, Yuxian Yu, Jeffrey Zheng

National Pilot School of Software, Yunnan University, Kunming Yunnan

Received: Sep. 22nd, 2022; accepted: Oct. 19th, 2022; published: Oct. 27th, 2022

Abstract

For block cipher, the security of cipher comes from confusion and diffusion, and the two mainly come from round function operation of block cipher. In order to meet the standardization of encryption algorithms, localization of encryption algorithms and the needs of all parties, this paper uses the relevant knowledge of graph theory and the latest vector logic variable system to study the security of block ciphers. This paper obtains the intermediate random sequence of block cipher, analyzes the characteristics of the random sequence, obtains the invariant of the random sequence, and further conducts statistical analysis and research on the invariant. The research

文章引用: 罗钰舒, 彭圣宇, 余愉先, 郑智捷. 分组密码随机序列的不变量研究[J]. 计算机科学与应用, 2022, 12(10): 2370-2383. DOI: 10.12677/csa.2022.1210243

direction is to observe and compare the statistical visualization results of the invariant through the control variable method, and explore the characteristics of the invariant; For the same block cipher, by changing the amount of invariant data, it is concluded that the same cipher invariant has the characteristics of saturation state, which provides a further theoretical basis for the security of block cipher.

Keywords

Block Cipher, Random Sequence, Invariant, Statistics, Graph Theory, Variable Value System

1. 引言

1.1. 分组密码简介

分组密码学的研究主要包含两部分，一是分组密码编码学，二是分组密码分析学。前者是研究如何设计安全的加密算法和密钥生成算法。后者是研究如何利用算法的特征来推导出特定的明文或者特定的密钥[1]。

分组密码的数学模型是将明文消息编码表示后的数字(简称明文数字)序列，划分成长度为 n 的组(可看成长度为 n 的矢量)，每组分别在密钥的控制下变换成等长的输出数字(简称密文数字)序列。密码安全性来自混淆、扩散性，而两者主要来自分组密码的轮函数操作。不同分组密码使用不同轮函数，不同分组密码轮函数产生的中间序列具有不同特征。

1.2. 不变量理论

数学不变性是理解和发展新的科学理论和技术的关键。大多数科学理论依靠群体行为和转换的不变性质来描述我们所生活的世界的规则。相对论和量子力学等理论都依赖于不变性性质来构建它们的结构[2]。

希尔伯特在 1893 年发表的一篇很有影响的论文《论不变量的完全系》中，发展出了解决不变量理论问题的新方法。他强调，这一方法根本上不同于他的前辈们的方法，因为他把代数不变量理论当作代数函数域的一般理论的组成部分来处理[3]。

数学中的一个主要风险是，数学理论主要在句法层面上运作，它们可能本质上是处理一个空集。除非目的是通过证明集合为空来证明安全性。目前关于对称密码中多项式不变量应用的研究缺乏实质或材料，无法以现实生活中有效的实例的形式工作。许多结果都是关于密码组件，而不是完整密码。例如，对于类似 AES 的 S 盒，我们可以使用所谓的交叉比(在更一般的非线性 ø 情况下，它已经是一个不变量，在数学中很少研究)。然而，这种类型的不变量仍然非常简单，或者我们只使用一个变量[4]。

1.3. 研究目的

分组密码备受关注，是密码学研究的热点课题之一。目前，已有大量的文献对各种分组密码的安全性进行讨论研究。为满足加密算法标准化、加密算法本土化及各方的需求，本文使用图论的相关知识，利用最新向量逻辑——变值体系[5]来对分组密码安全性进行研究。

从相空间结构和组织的角度，变值逻辑是共轭逻辑从二维黑白图像投影到一维 0~1 向量之后获得的成果。在二维图中，利用不变量在规则平面格连接上完成共轭分类和变换；而在一维向量上，对任意 N > 0，对 2N 状态群集都能利用各种不变量参数结合输入输出关系进行分类和变换，形成具有任意分划特性的量化描述模式，灵活地适配不同的应用。

本文将通过获取分组密码中的随机序列，并对所获取的各个序列进行不变量检测，得到不变量统计可视化结果，进行分析总结，探索不变量反映出分组密码的性质，从而达到由分组密码产生的随机序列的不变量研究的目的。

2. 分组密码随机序列获取

本文选择四种分组密码作为研究对象，分别是 DES、AES、RC5、SM4。整体工作流程如图 1，对于所获取的中间序列，需要具备两个要求：

要求一：确保中间序列足够长，以此获取完备特征；

要求二：确保用来对比的不同分组密码所产生的中间序列等长。

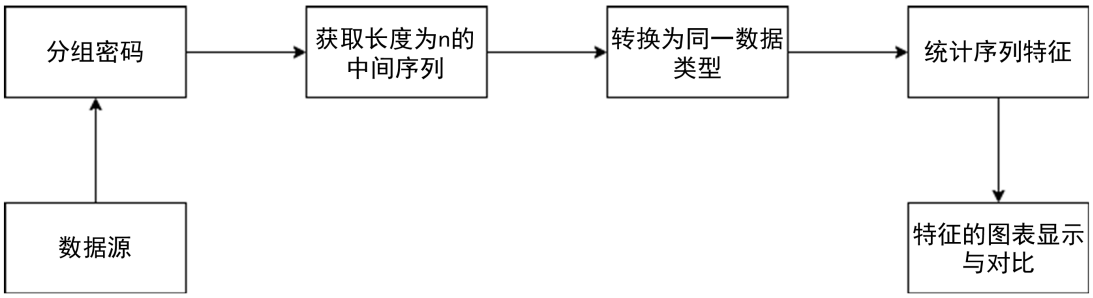


Figure 1. Random sequence acquisition process of block cipher
图 1. 分组密码随机序列获取流程

对于随机序列的特征选取上，以 0-1 向量变值逻辑为基础，构造特征向量。

在随机序列特征的统计上，采用以下方法来获取序列特征：对序列长度为 n 的中间序列等长分割为长度为 m (m < n)的多个分段[6]，由此得到⌈n/m⌉个分段，再分别统计⌈n/m⌉个分段中 1 和 01 的个数，统计结果形式如整个序列分段中 1 的个数为 k $\left(k \in [0,m]\right)$ 的段数有 i 段，则记$(k,\ i)$为对应结果。特征获取过程如图 2。

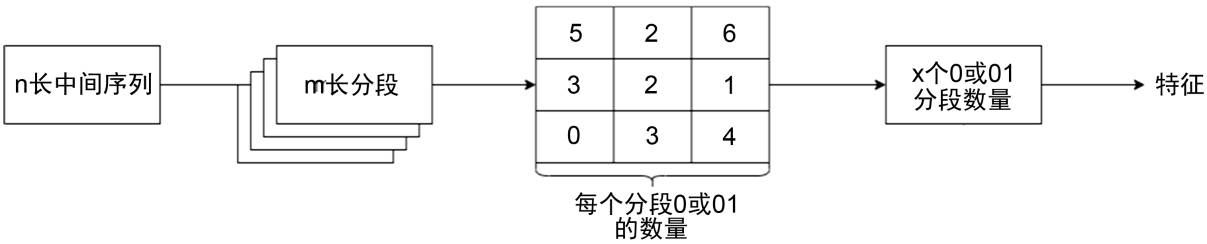


Figure 2. Random sequence feature acquisition process
图 2. 随机序列特征获取过程

3. 随机序列不变量获取

不同于经典统计方法稳健统计法[7]的统计适应性问题和贝叶斯统计学[8]归纳推理，本文使用的是基于概率的特征统计方法，将不变量视作一个特征，对其进行概率处理，这样做的优点是能够通过可视化

结果比较客观的得出需要的结论和看出研究对象的特性，且较为简便，易于操作。

根据不变量定义，参数 k 为所得到的多组随机序列，变换 T 为计数每组序列中 1 和 01 的个数。则函数 $I(P)$为首先对每个 1 和 01 的数量进行概率统计，得到各段的概率 p_i，且有

$$\sum_{i=0}^{m} p_i = 1,$$

再分别对各个概率进行开根号处理，之后进行求和，最终得到一个具体的数值，这个数值即是函数 $I(P)$的值，也是需要检测得出的不变量。获取不变量流程如图 3。

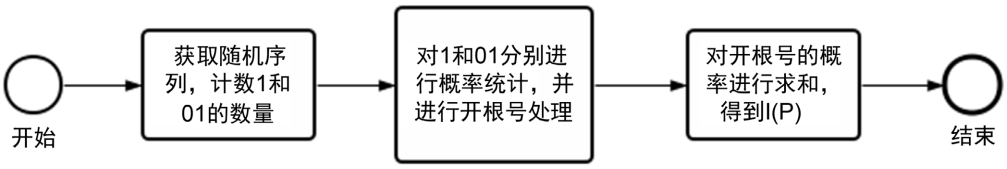


Figure 3. Invariant acquisition process
图 3. 不变量获取流程

4. 不变量统计

利用 0~1 向量逻辑，从测度空间出发，形成状态群聚特征值，针对选择的不变量特征状构造特征向量，展现群聚效应和共轭表示结构[9]。

以 SM4 密码为示例，对多段随机序列进行求解不变量值，并进行统计，最后可视化结果。

数据类型为论文文本，分段长度 m = 64，选取 1500 段随机序列计算分别得出 1 和 01 的开根号概率和(即不变量) p_i 和 q_i $\left(i \in [1,1500]\right)$，保留两位小数后，乘 100 得到一个整数，将不变量作为横坐标，对应不变量的段数 h 作为纵坐标，整理后得到可视化结果，如图 4。

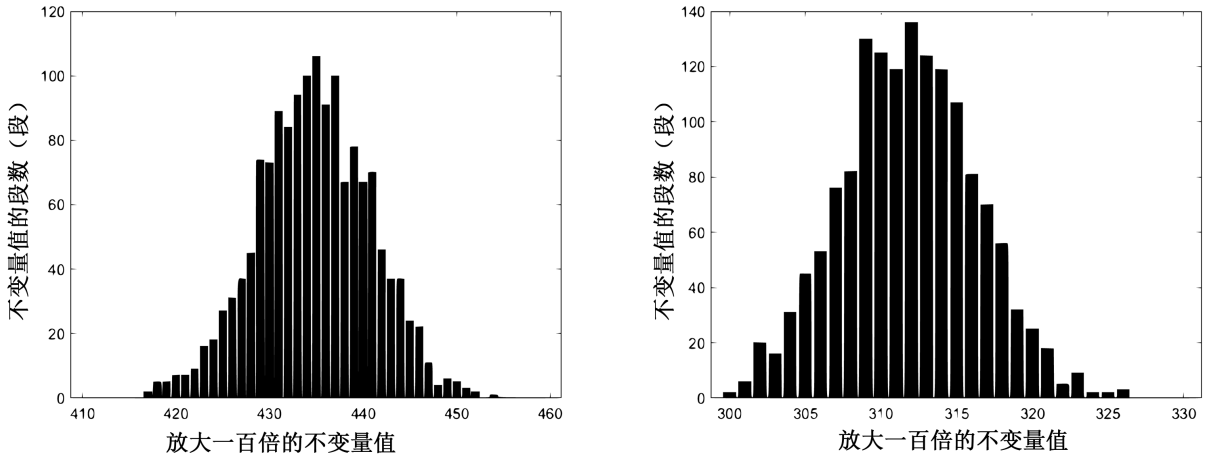


Figure 4. SM4 block cipher, segment length m = 64, data volume x = 1500, left count 1, right count 01
图 4. SM4 分组密码，分段长度 m = 64，数据量 x = 1500，左计数 1，右计数 01

5. 不变量饱和临界值研究

由于不变量样本数据量较少，分析得到的结论具有一定局限性，后续增加数据量，进行更进一步的研究。

在增加数据量的情况下，发现当数据量到达一定量时，图形的基本形状就不会再发生改变，即达到

饱和态，推测饱和态出现的这种情况是与密码的稳定性有关，于此进行探究。

由于分组长度 m 越大，所得到的不变量值所属范围长度会增大，如果 m 值过大，所得到的临界值会受更多种因素的影响，临界值也就越难以得到。因此，为了控制重复运行的次数，选择适中的 m 值，便于在一定时间内对不变量临界值进行研究。

在研究过程中，选择的分组长度 m 值为 64，具体流程如图 5。

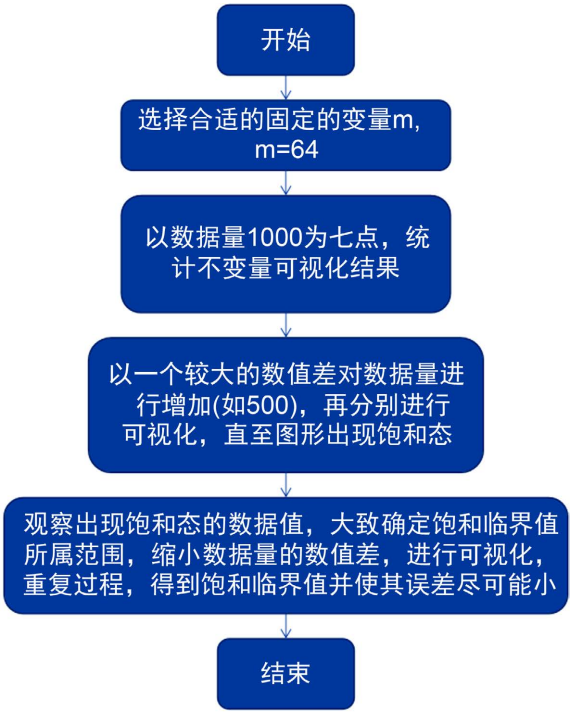


Figure 5. Research process of invariant saturation critical value
图 5. 不变量饱和临界值研究流程

以 SM4 密码为例，数据类型为论文文本，选择的分段长度 m 值为 64，通过改变所统计的不变量数据量 x 来进行对不变量数据饱和临界值的研究。

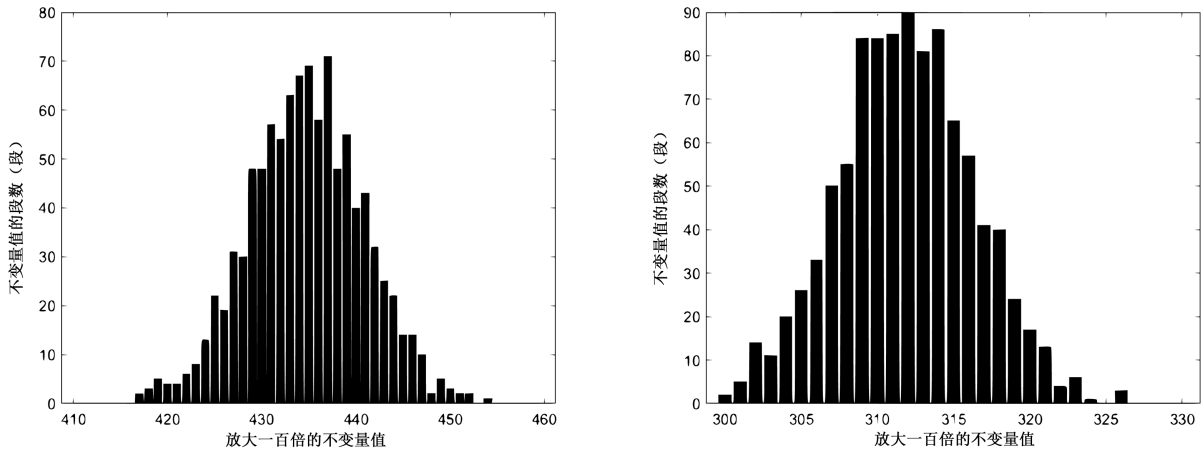


Figure 6. SM4 block cipher, segment length m = 64, data volume x = 1000, left count 1, right count 01
图 6. SM4 分组密码，分段长度 m = 64，数据量 x = 1000，左计数 1，右计数 01

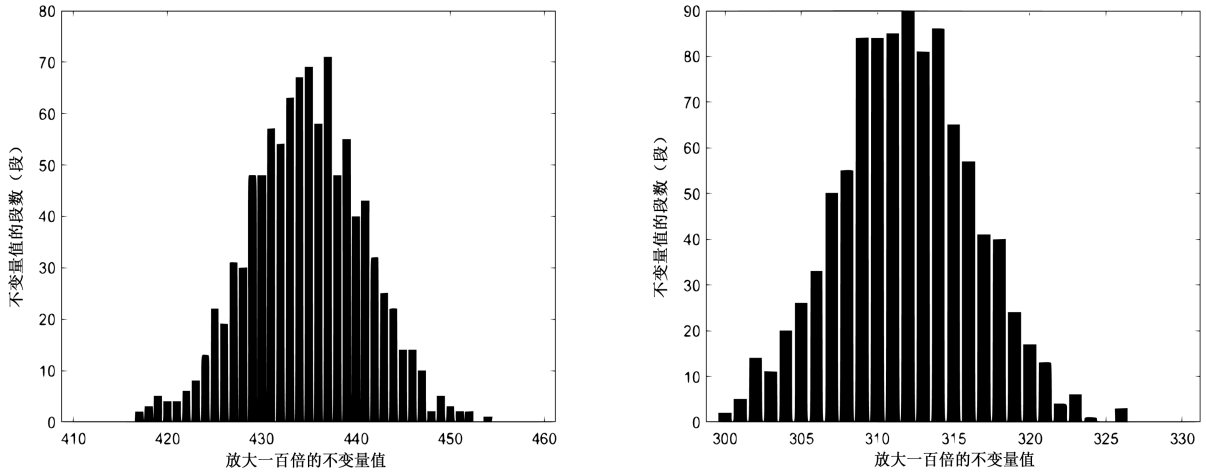


Figure 7. SM4 block cipher, segment length m = 64, data volume x = 1500, left count 1, right count 01
图 7. SM4 分组密码，分段长度 m = 64，数据量 x = 1500，左计数 1，右计数 01

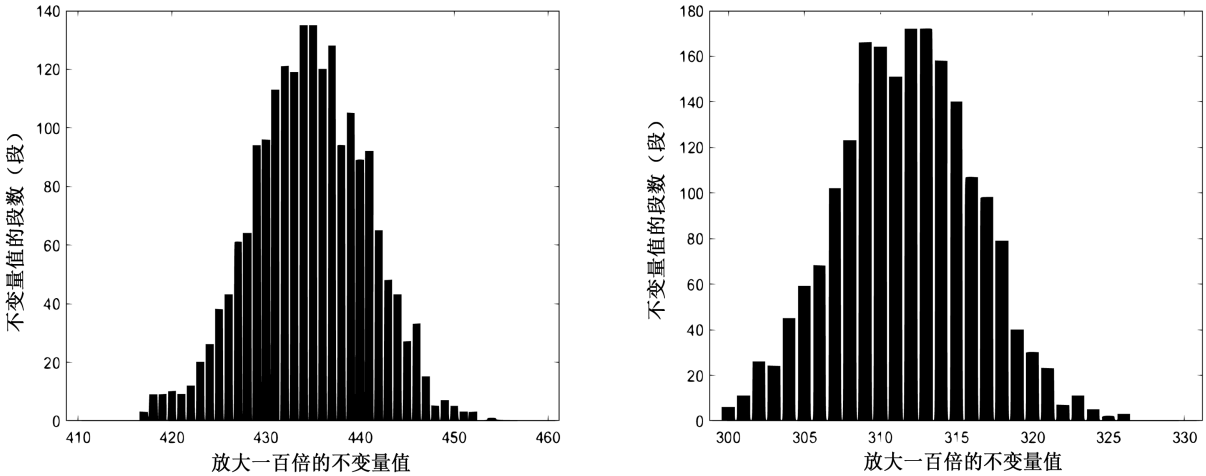


Figure 8. SM4 block cipher, segment length m = 64, data volume x = 2000, left count 1, right count 01
图 8. SM4 分组密码，分段长度 m = 64，数据量 x = 2000，左计数 1，右计数 01

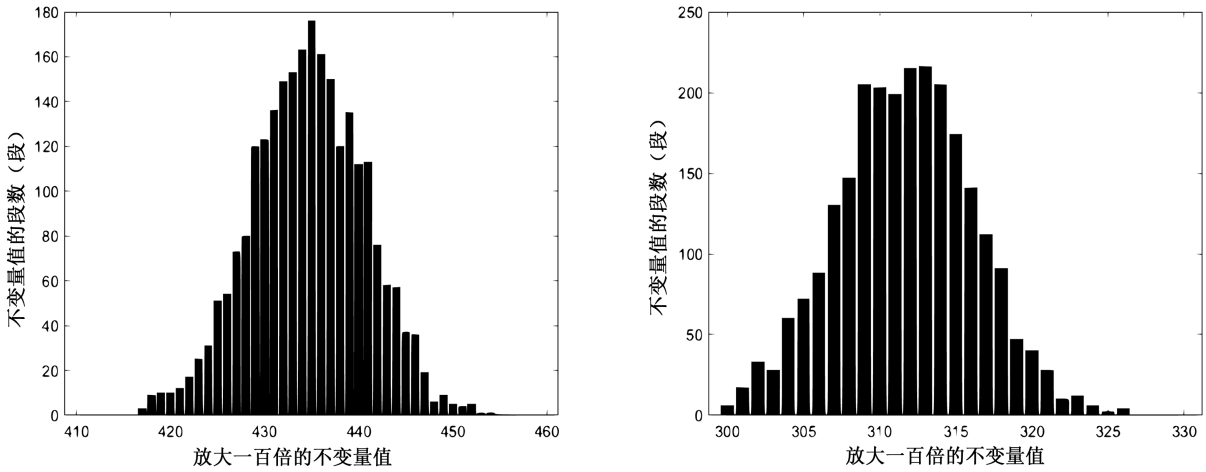


Figure 9. SM4 block cipher, segment length m = 64, data volume x = 2500, left count 1, right count 01
图 9. SM4 分组密码，分段长度 m = 64，数据量 x = 2500，左计数 1，右计数 01

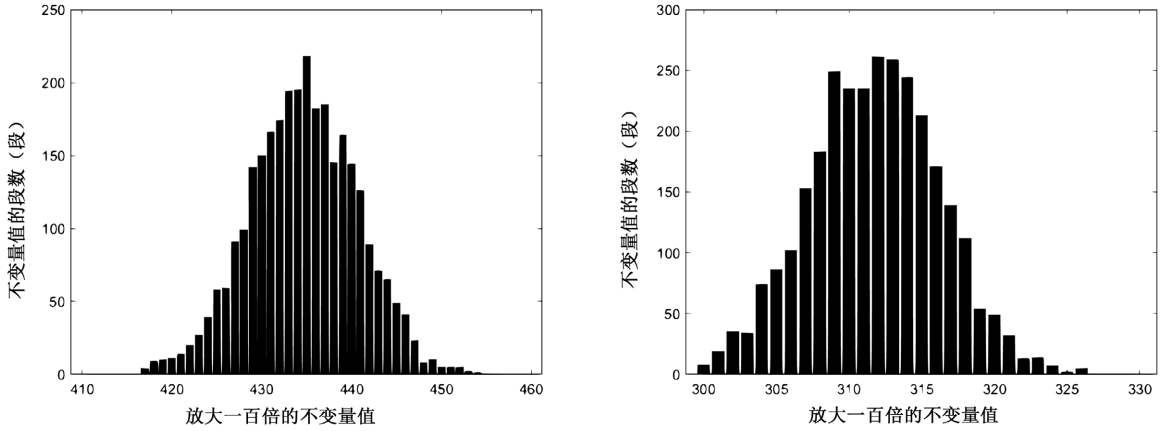


Figure 10. SM4 block cipher, segment length m = 64, data volume x = 3000, left count 1, right count 01
图 10. SM4 分组密码，分段长度 m = 64，数据量 x = 3000，左计数 1，右计数 01

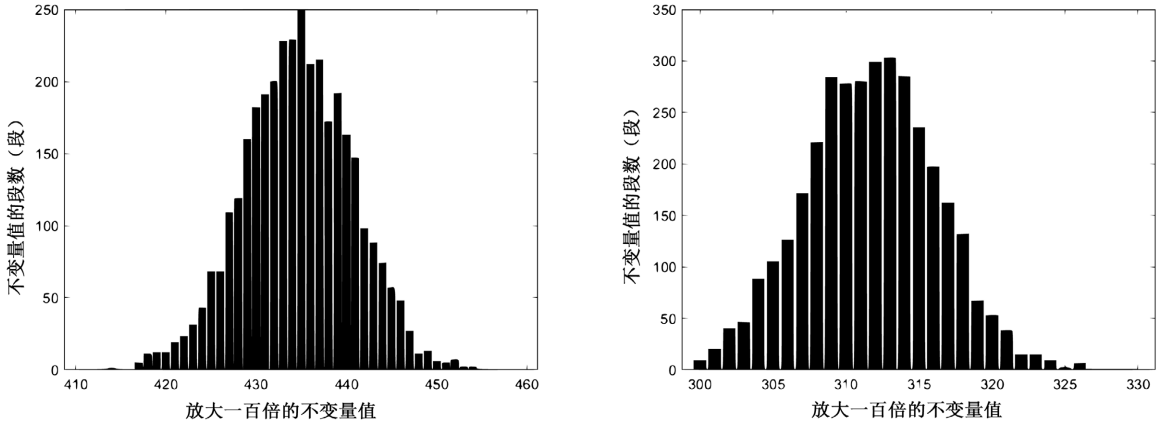


Figure 11. SM4 block cipher, segment length m = 64, data volume x = 3500, left count 1, right count 01
图 11. SM4 分组密码，分段长度 m = 64，数据量 x = 3500，左计数 1，右计数 01

从上述结果可以看出当数据量达到 2500 以上时，再加大数据量，所得到的图形形状不会再有较大的变化，判定 SM4 密码不变量饱和临界值在 2000~2500 之间。为增大临界值精度、缩小误差，在数据量 2000~2500 间缩小每次增加的数据量。

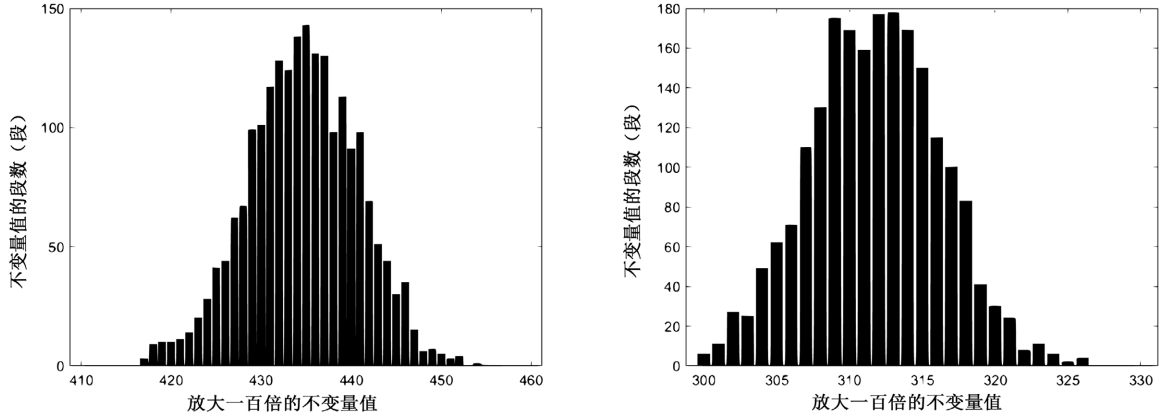


Figure 12. SM4 block cipher, segment length m = 64, data volume x = 2100, left count 1, right count 01
图 12. SM4 分组密码，分段长度 m = 64，数据量 x = 2100，左计数 1，右计数 01

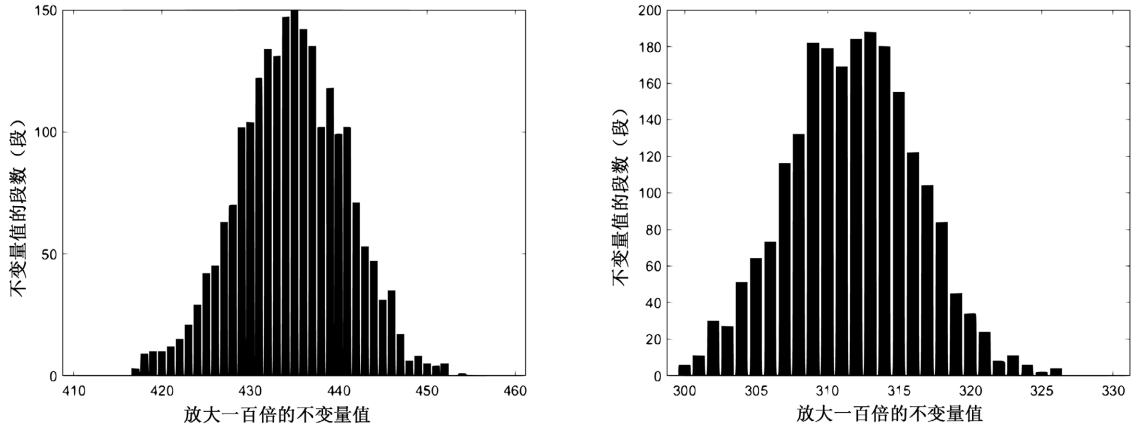


Figure 13. SM4 block cipher, segment length m = 64, data volume x = 2200, left count 1, right count 01
图 13. SM4 分组密码，分段长度 m = 64，数据量 x = 2200，左计数 1，右计数 01

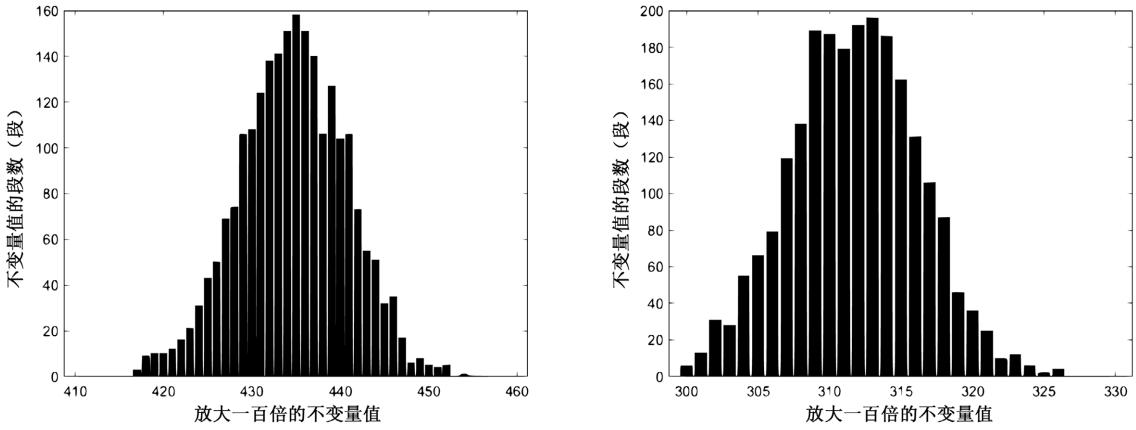
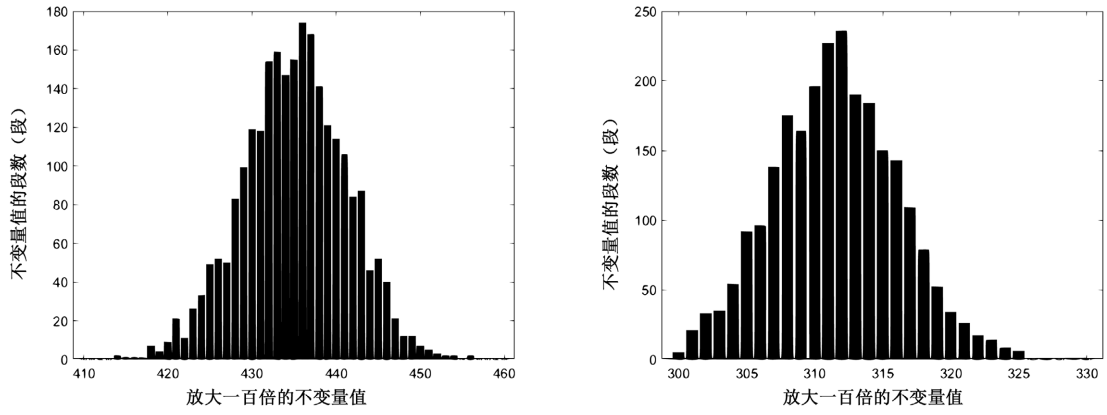


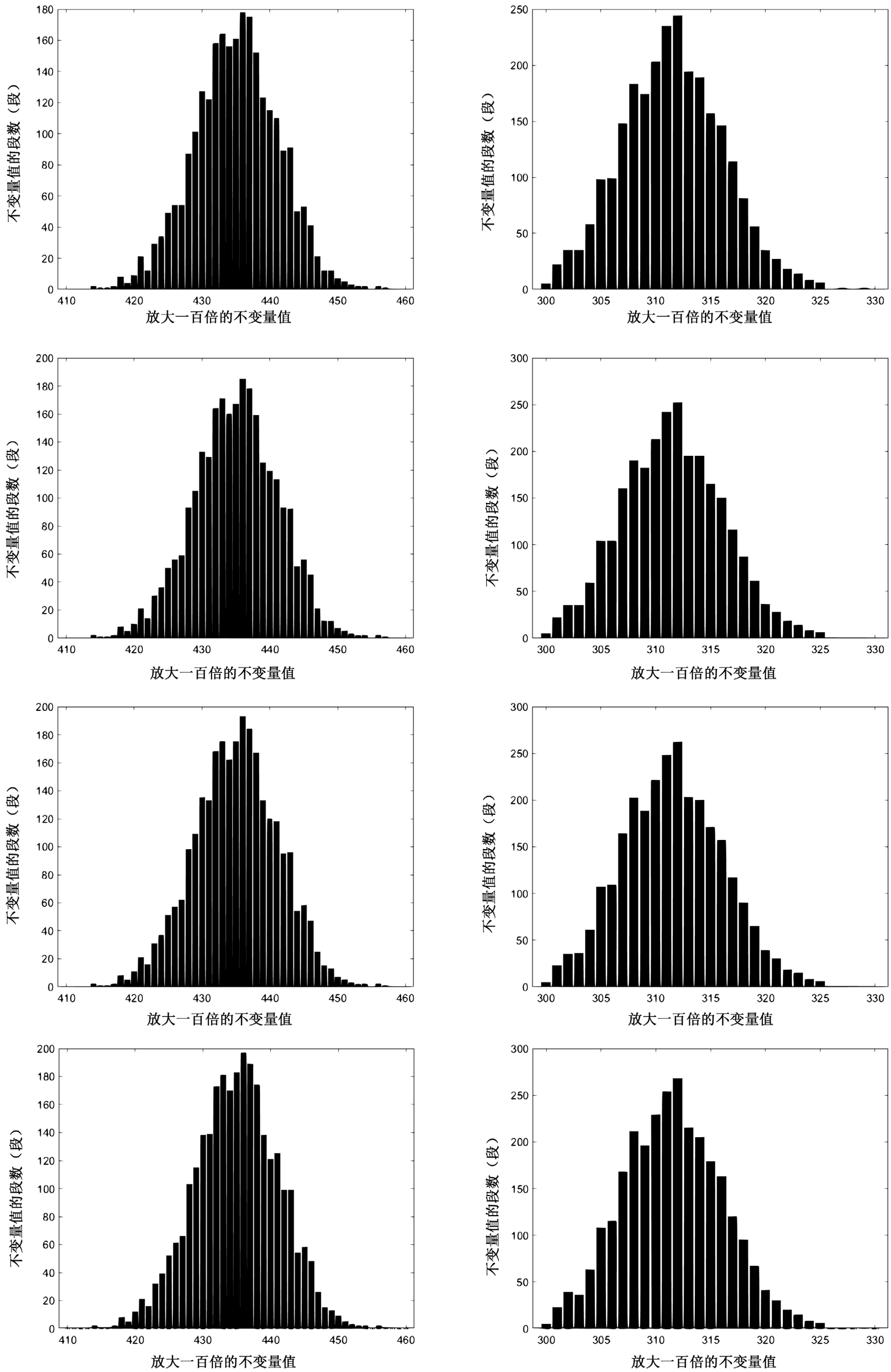
Figure 14. SM4 block cipher, segment length m = 64, data volume x = 2300, left count 1, right count 01
图 14. SM4 分组密码，分段长度 m = 64，数据量 x = 2300，左计数 1，右计数 01

由图 6~14 的可视化结果可以看出，数据量在 2200~2300 之间即可显示整体形状，于是得出 SM4 密码不变量饱和临界值约为 2250。

基于上述流程，我们还分别对 AES、RC5、DES 进行了不变量饱和临界值探究，它们的关键截图分别如图 15~17 所示。

AES：




不变量值的段数（段）
180 160 140 120 100 80 60 40 20 0
410 420 430 440 450 460
放大一百倍的不变量值
不变量值的段数（段）
250 200 150 100 50 0
300 305 310 315 320 325 330
放大一百倍的不变量值
不变量值的段数（段）
200 180 160 140 120 100 80 60 40 20 0
410 420 430 440 450 460
放大一百倍的不变量值
不变量值的段数（段）
300 250 200 150 100 50 0
300 305 310 315 320 325 330
放大一百倍的不变量值
不变量值的段数（段）
200 180 160 140 120 100 80 60 40 20 0
410 420 430 440 450 460
放大一百倍的不变量值
不变量值的段数（段）
300 250 200 150 100 50 0
300 305 310 315 320 325 330
放大一百倍的不变量值
不变量值的段数（段）
200 180 160 140 120 100 80 60 40 20 0
410 420 430 440 450 460
放大一百倍的不变量值
不变量值的段数（段）
300 250 200 150 100 50 0
300 305 310 315 320 325 330
放大一百倍的不变量值

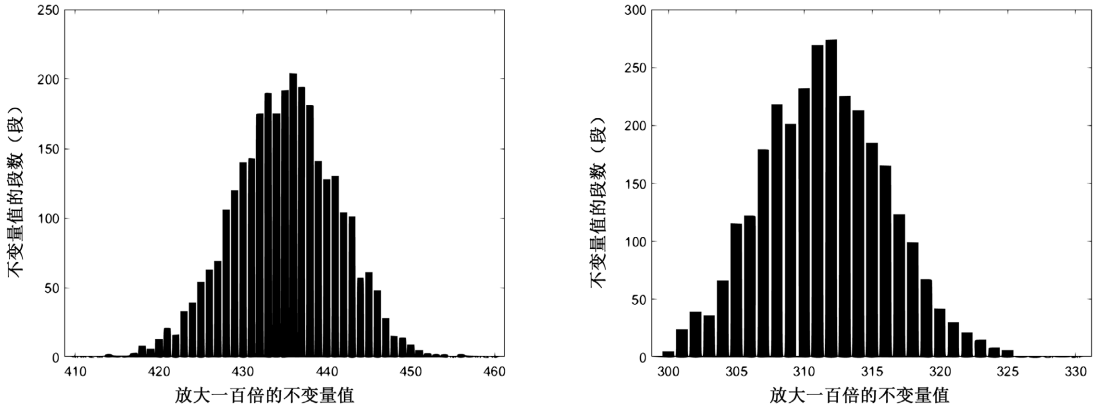
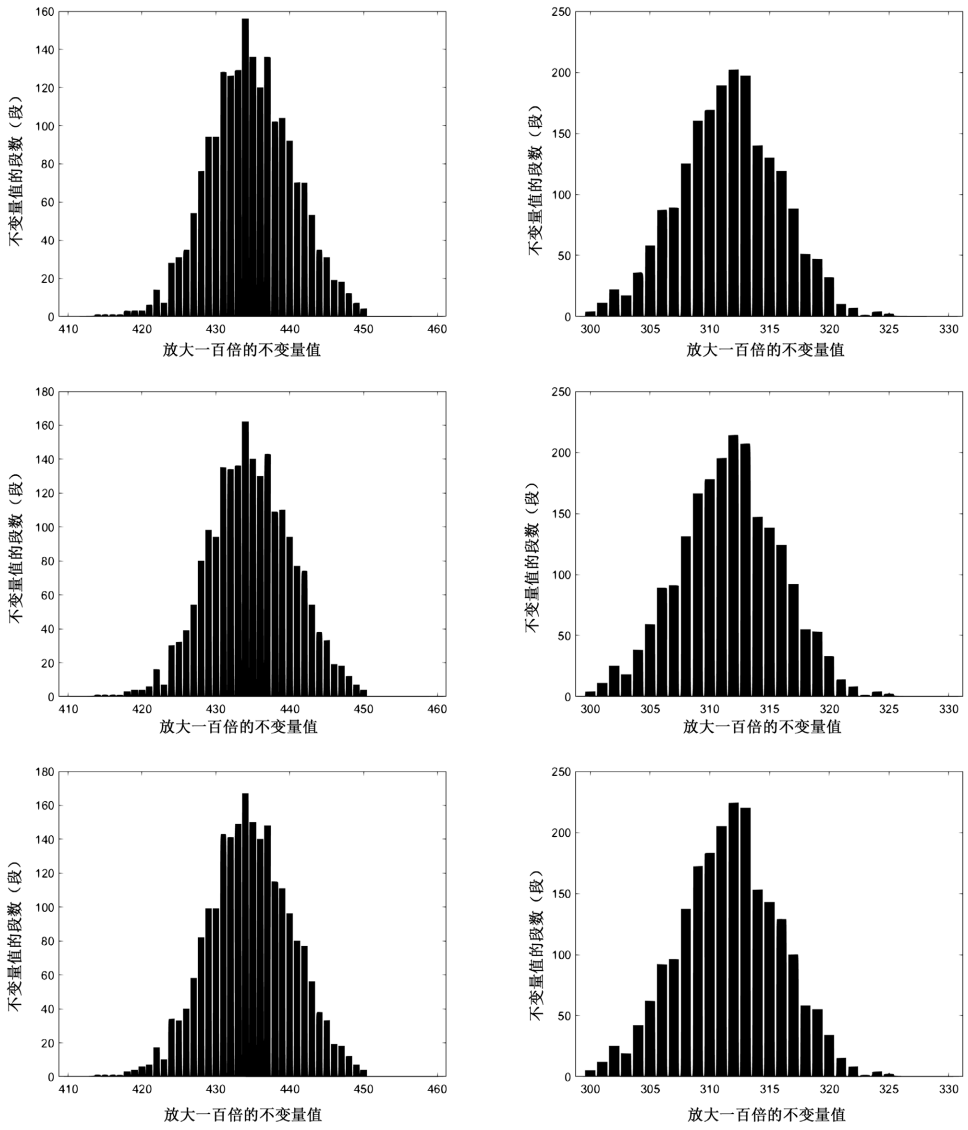


Figure 15. The data volume x is from 2500 to 3000, the increment interval is 100, the left count is 1, and the right count is 01
图 15. 数据量 x 从 2500 到 3000，递增间隔 100，左计数 1，右计数 01

RC5：



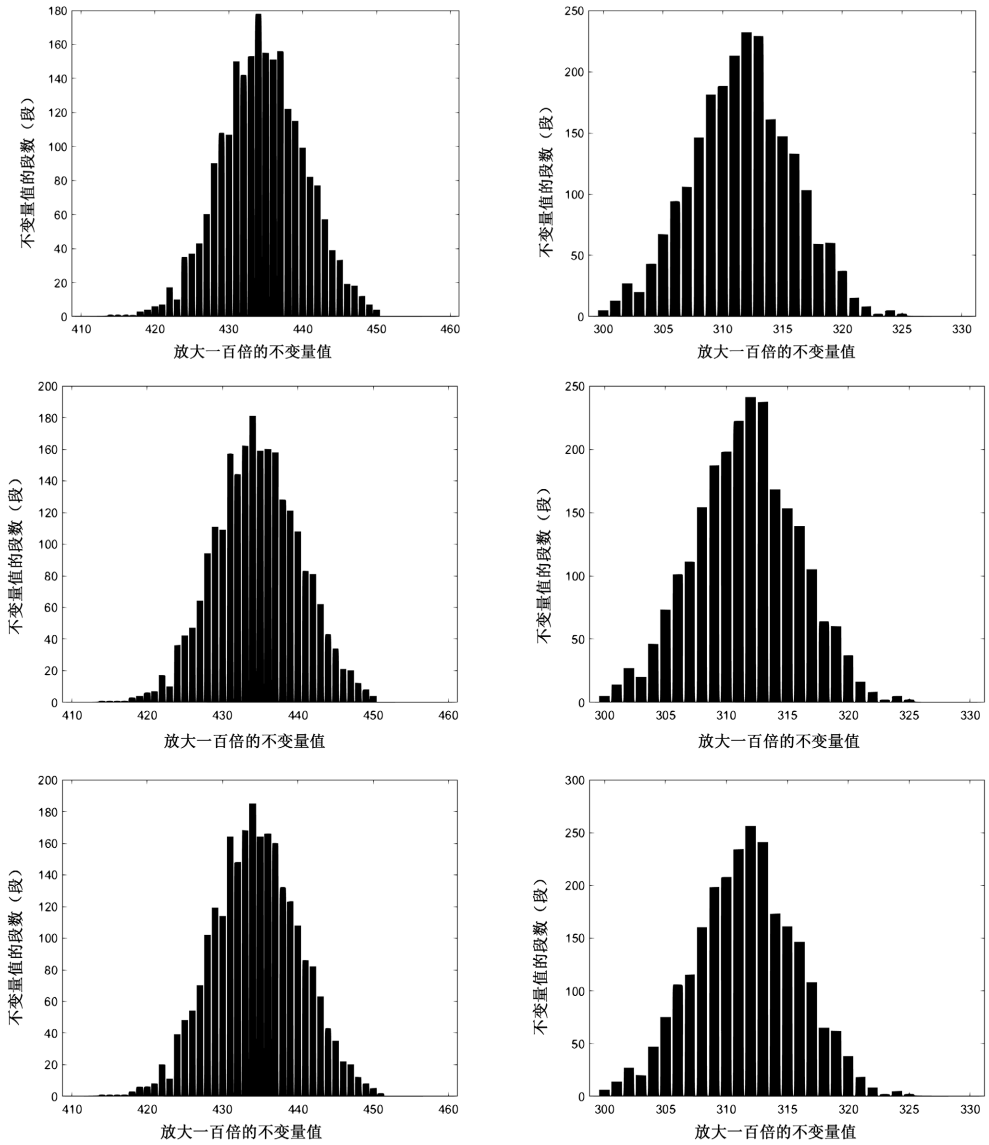
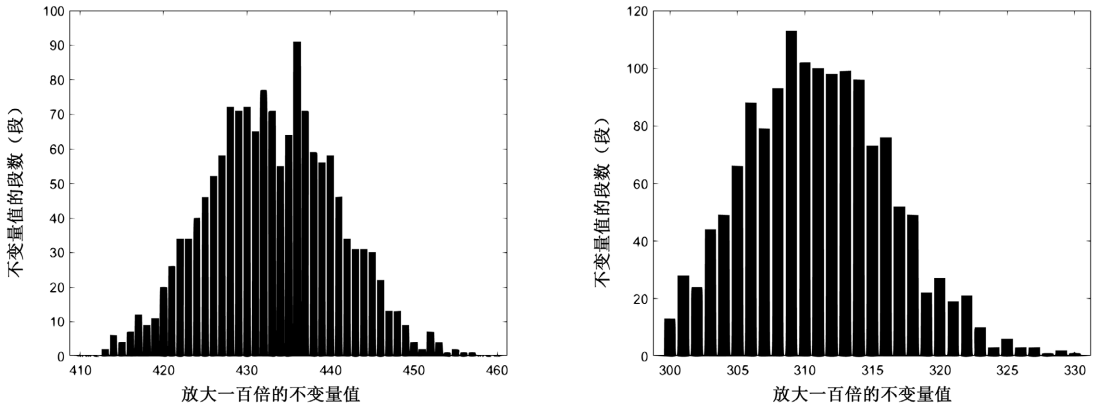
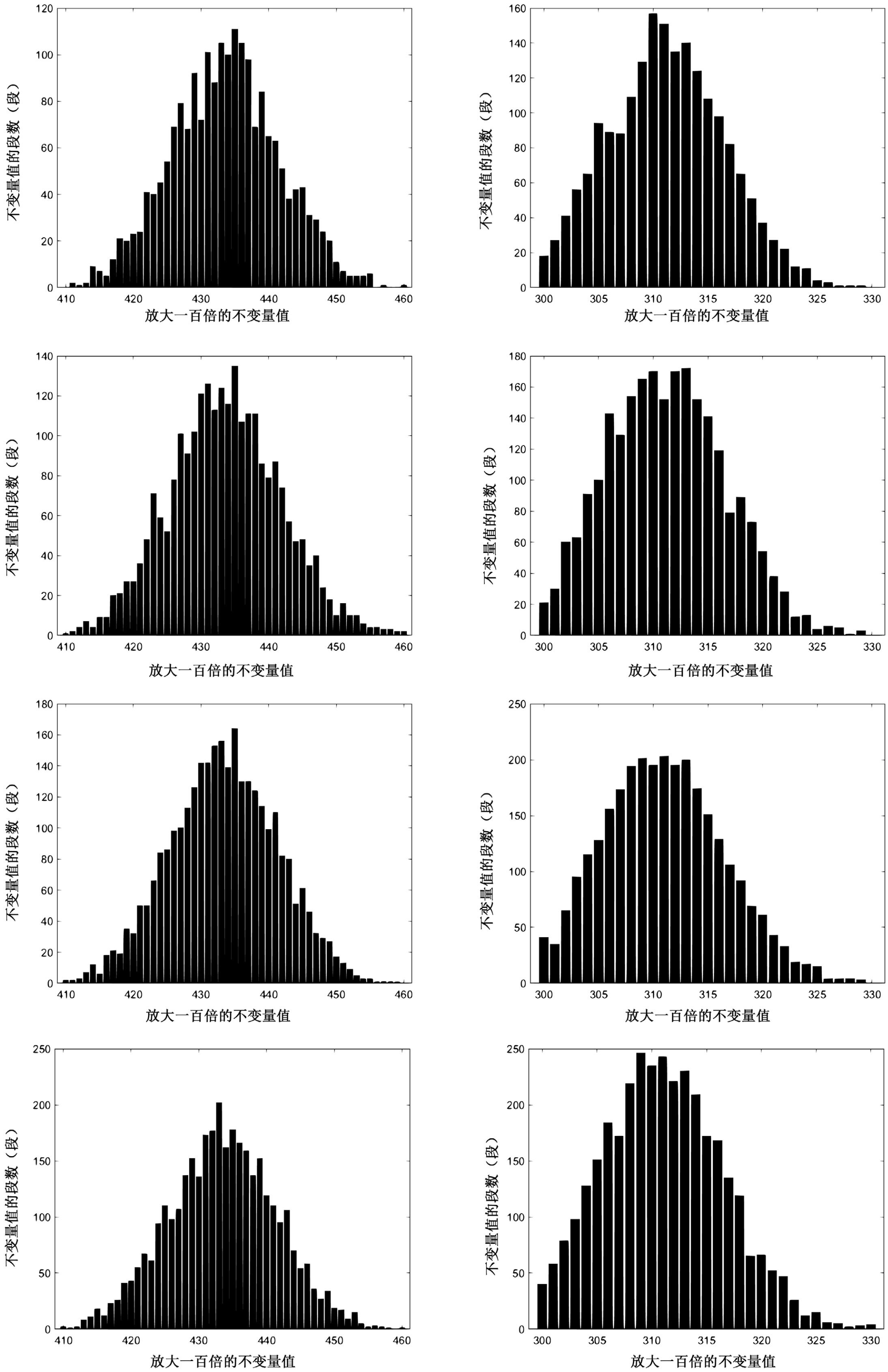


Figure 16. The data volume x is from 2000 to 2500, the increment interval is 100, the left count is 1, and the right count is 01
图 16. 数据量 x 从 2000 到 2500，递增间隔 100，左计数 1，右计数 01

DES：




不变量值的段数（段）
放大一百倍的不变量值

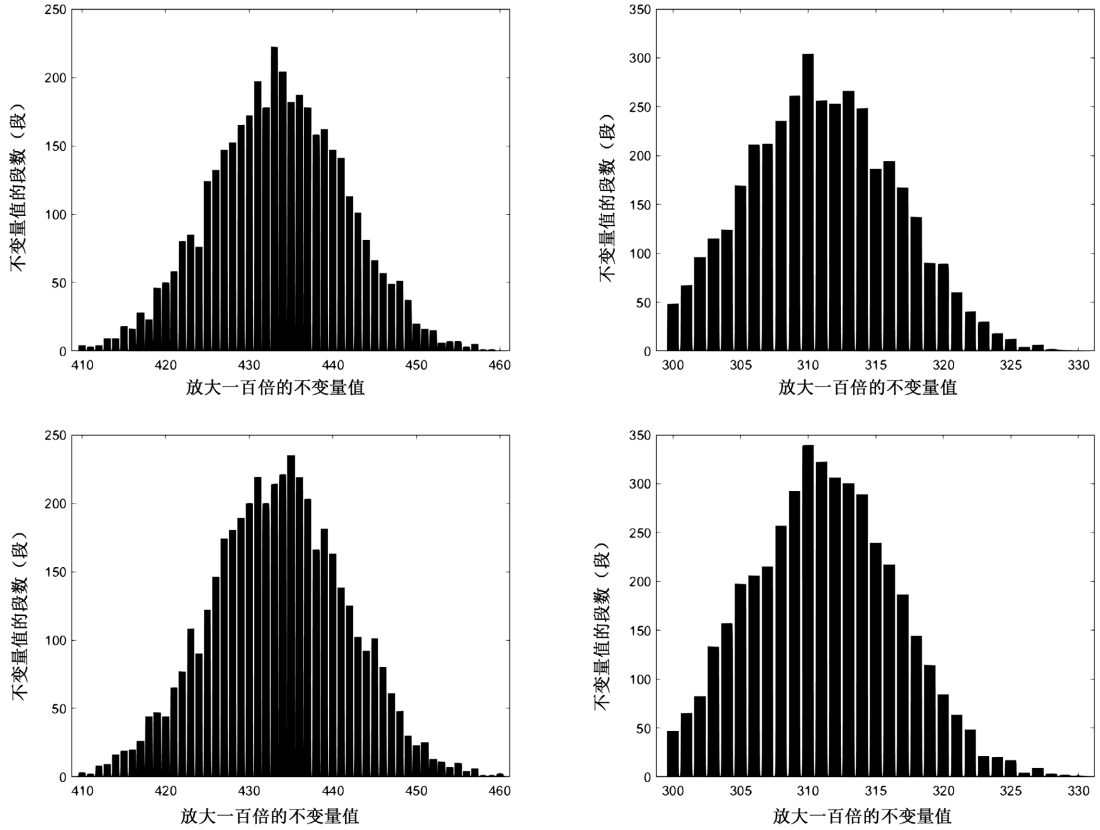


Figure 17. The data volume x is from 1500 to 4500, the increment interval is 500, the left count is 1, and the right count is 01
图 17. 数据量 x 从 1500 到 4500，递增间隔 500，左计数 1，右计数 01

6. 结论

本文仅展示 SM4 密码的研究可视化结果，其余三种密码均已做了相同的研究测试，通过研究可以知道，在 m 值固定下，这四种分组密码的不变量统计中，相同点在于：1 和 01 的计数均保持在一定的范围。而不同点在于：所得到的可视化形状有差异，拿计数 1 的不变量来阐述，DES 密码呈现类正态分布的曲线形状，且范围较宽；AES 密码也呈现类似正态分布的形状，但范围较窄；RC5 密码则展现为在峰值处有较为平缓的区域；SM4 密码则表现为阶梯式，上升逐渐变陡，而下降则较为平缓。

通过对各个分组密码不变量饱和态研究后，结果显示，对于一个单独的密码来说，当不变量数据量到达一定量时，其可视化图形形态会趋于稳定，当继续增大数据量时，图形形态不会发生较大的变化，即此时达到了饱和状态，称这个临界数据量为不变量饱和临界值；而对于不同的密码来说，它们的不变量饱和临界值也会有所不同，经过统计，SM4 密码和 RC5 密码的饱和临界值约为 2250，AES 密码的饱和临界值约为 2900，说明不同密码需要的使得统计图形形状呈现饱和态的数据量不同，另外，DES 密码出现了范围饱和临界值(即在不同范围内存在不同的饱和临界值)的特殊情况，猜测这种情况的出现与 DES 密码具有较为稳定的特性有关。

后续研究仍有展望，如改变分段长度 m 值，固定数据量进行不变量统计，观察分段长度对结果的影响；由于 DES 密码出现特殊情况，或继续深入探究特殊情况出现的原因，或找出其各段饱和态的数值等。

致　谢

感谢郑智捷教授的悉心指导，感谢云南大学软件学院对本项目的支持。

参考文献

[1] 崔婷婷. 分组密码算法和流密码算法的安全性分析[D]: [博士学位论文]. 济南: 山东大学, 2018.

[2] Zheng, J. (2019) Variant Construction from Theoretical Foundation to Applications, Springer Nature Press, Singapore. https://doi.org/10.1007/978-981-13-2282-2

[3] (美)卡尔·B·博耶. 数学史下修订版[M]. 北京: 中央编译出版社, 2012.

[4] Courtois, N.T. and Patrick, A. (2019) Lack of Unique Factorization as a Tool in Block Cipher Cryptanalysis. https://arxiv.org/abs/1905.04684

[5] 郑智捷. 变值体系理论及其应用第 1 卷理论基础及其应用[M]. 北京: 科学出版社, 2021.

[6] Zheng, J. and Zhu, M.H. (2021) Input-Output Types of Fifteen Modules on Discrete and Real Measurements for COVID-19. 71-85. https://doi.org/10.21203/rs.3.rs-65158/v2

[7] 王继荣, 王子亮. 稳健统计方法在实验室能力验证中的应用[J]. 质量与认证, 2022(2): 75-77. https://doi.org/10.16691/j.cnki.10-1214/t.2022.02.010

[8] 谷恒明. 经典统计学与贝叶斯统计学在回归模型中的比较研究[D]: [硕士学位论文]. 北京: 军事科学院, 2018.

[9] 郑智捷. 消解逻辑悖论建立元知识智能化体系[J]. 计算机科学, 2022, 49(1): 9-16.