披露规避系统的隐私与偏差分析

Keyu Zhu[1] , **Ferdinando Fioretto**[2] , **Pascal Van Hentenryck**[1] , **Saswat Das**[3] 和 **Christine Task**[4]1佐治亚理工学院2雪城大学3印度国家科学教育与研究学院4Knexus Research Corporation
keyu.zhu@gatech.edu, ffiorett@syr.edu, pvh@isye.gatech.edu,
saswat.das@niser.ac.in, christine.task@knexusresearch.com

arXiv:2301.12204v1 [cs.CR] 28 Jan 2023

摘要

披露避免（DA）系统用于保护数据的保密性，同时允许对数据进行分析和传播，以服务于分析目的。这些方法，例如单元格抑制、交换和k-匿名性，通常得到应用，并可能对社会和经济产生重大影响。然而，对其隐私和偏差保障的形式化分析一直处于缺失状态。本文提出了一个填补这一空白的框架：为这些机制提出差分隐私版本，并推导其隐私界限。此外，本文还在美国人口普查数据发布和分类任务中，从准确性与公平性方面将这些机制的性能与传统差分隐私机制进行了比较。结果表明，与普遍看法相反，就准确性和公平性而言，传统差分隐私技术可能优于广泛使用的数据匿名化机制的差分隐私对应版本。

1 引言

披露避免（DA）系统是用于保护保密性，同时仍支持数据分析和传播的方法。这些技术应用于经济学、公共卫生、社会科学和数据科学等多个领域，并且在普查及其他数据收集工作中拥有悠久的历史。例如，美国人口普查局自1930年十年一次的人口普查数据发布起，就一直采用各种传统数据匿名化技术。这些技术包括根据特定区域内的人口或家庭数量抑制某些表格，以及交换具有相似特征的记录中的数据。

尽管传统的保密性措施（例如抑制 [Kelly 等、 1992], 交换 [Dalenius 和 Reiss、1982], 以及 k-匿名性 [Sweeney）对于防止意外或蓄意披露十分重要，但它们缺乏能够量化个人在数据发布时所承担隐私风险的形式化保证。这一点很重要，因为它限制了参与者评估这些保护措施对已发布数据影响的能力。

相比之下，差分隐私（DP） [Dwork等人， 2006]是一种相对较新的披露避免方法（DA），它对隐私作出了严格定义，并允许量化隐私保证。在差分隐私中，通过以受控的方式向个人数据添加噪声，可以保护个人的隐私。这一过程确保个人参与某个数据集不会对后续查询的结果产生显著影响。美国人口普查局最近在 2020年人口普查数据发布中采用了差分隐私，这标志着隐私保护朝着更严格的方向迈出了重要一步。不过值得注意的是，许多其他数据机构和组织仍然依赖传统披露避免系统来保护其数据的保密性。

尽管这些方法能够有效防止意外或故意的信息披露，但与差分隐私相比，它们能提供何种隐私保证尚不明确。另一方面，差分隐私虽然比传统的披露避免系统提供更强的隐私保证，但可能会以准确性和公平性为代价 [Kuppam 等人，2019年；Tran 等人，2021年；Fioretto等人， 2022], 这一议题近来引发了广泛讨论。

鉴于这些披露避免方法被用于发布数据产品，并为具有重大社会和经济影响的决策提供依据，因此有必要从隐私、偏差和公平性等方面，对传统披露避免方法与差分隐私进行严格比较。然而，这种比较面临的挑战之一，是缺乏用于评估隐私保护的标准化框架。差分隐私对隐私作出了严格定义，并使隐私保证能够量化。另一方面，传统的披露避免技术可能没有一套明确的隐私度量指标，因此难以直接比较它们所提供的隐私保护水平。

贡献 本文旨在应对这一挑战：提出一个用于比较传统披露避免与差分隐私的框架，并作出四项具体贡献。(1) 首先，针对三种广泛采用的经过精心随机化的 传统 披露避免方法提出相应版本：抑制、交换和 k-匿名性。由此得到的随机化机制可以进行严格分析。具体而言，本文推导出这些新机制的 (ϵ, δ)-差分隐私界限，并证明其在准确性方面接近传统对应方法。(2) 随后，本文推导出新披露避免机制的偏差界限，从而能够将其与已有此类界限的经典差分隐私技术直接比较。(3) 接下来，本文分析所考察的披露避免系统产生的公平性影响，并



图1：各种传统披露避免机制的示意图。

表明随机化披露避免算法造成的公平性违规接近其传统对应方法。(4) 最后，本文对新披露避免机制的性能进行了广泛的实证分析，并将其与两种经典的差分隐私算法在数据发布和分类任务上的表现进行了比较。

从更广泛的角度来看，本文表明，与普遍看法相反，在隐私级别相同的情况下，经典差分隐私机制在重要的数据发布和学习任务中，可能比传统披露避免系统具有更高的准确性和公平性。因此，本研究的结果有望影响数据机构和组织处理披露避免问题的方式：事实上，本文首次提出了一个框架，用于比较传统 *DA* 与差分隐私的相对优势和局限性。。

2 问题设定

本文考虑包含 m 条记录和 d 个属性的数据集，A_1，……，A_d。每条记录都是一个 d维属性元组，与数据全集 $\mathcal{X}$中的某个唯一个体相关联：$= \prod_{i=1}^{d} \mathrm{dom}\,(A_i)$，其中dom (*A*)表示属性 A的所有可能取值的集合。为方便起见，假设 $\mathcal{X} = \{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n\}$，其中 n 是数据全集 $\mathcal{X}$的大小，并考虑数据集 D的直方图 $\boldsymbol{x}(D) \in \mathbb{N}_+^n$，其 i^{th} 项 $x_i(D)$ 表示具有属性组合 $\boldsymbol{a}_i$的个体记录数。若不存在歧义，为简洁起见，在表达式 $\boldsymbol{x}(D)$中省略数据集 D 。此外，在不失一般性的前提下，假设直方图 $\boldsymbol{x}$ 按某种递增顺序排列，即 $x_i \le x_j$，对任意$i < j$均成立。最后，假设直方图 $\boldsymbol{x}$ 的每一项均以值 $B > 0$为上界，即 $\boldsymbol{x} \in [B]^n$。

例如，考虑图1(a)中的示意图；数据集 D 包含具有三个属性的记录：（地理）“区块”、“性别”和“投票年龄”。相关直方图如图1(b)所示。在此例中，属性“性别”与“邮政编码”等外部信息结合后，可能成为个人身份识别信息，因此被称为准标识符（QI），而其余属性则称为非准标识符。本文中，准标识符和非准标识符的集合分别记为 Q 和 N。给定一条记录 $\boldsymbol{a}$ 和一个属性集合 S，$\boldsymbol{a}[S]$ 表示 $\boldsymbol{a}$中属性 S 的取值向量。

本文旨在分析传统披露避免系统（将在下一节回顾）在发布直方图 $\boldsymbol{x}$（D）的隐私保护版本 $\tilde{\boldsymbol{x}}$（D）这一任务中的隐私、效用和公平性属性。本文所考虑的隐私概念是差分隐私，将在下一节回顾。分析中至关重要的效用和公平性概念依赖于（统计）偏差的概念。对于任一条目 $i \in [n]$，与机制 $\mathcal{M}$ 相关的偏差为

$$\mathcal{B}(\mathcal{M})_i = \mathbb{E}\left[\mathcal{M}(D)_i\right] - x_i(D)\,,$$

其中，期望值是对机制的随机性取平均，且 $\boldsymbol{\mathcal{B}}(\mathcal{M})$ $= [\mathcal{B}(\mathcal{M})_1 \quad \ldots \quad \mathcal{B}(\mathcal{M})_n]$. 公平性定义为直方图条目之间偏差的最大差异。

定义 1 (α-公平性 [Zhu 等， 2022])。如果一个机制$\mathcal{M}$的偏差最大差异不超过 α，即称其满足 α-公平性，

$$\|\boldsymbol{\mathcal{B}}(\mathcal{M})\|_{\rightleftharpoons} = \max_{i\in[n]} \mathcal{B}(\mathcal{M})_i - \min_{i\in[n]} \mathcal{B}(\mathcal{M})_i \le \alpha.$$

3 披露避免与私有数据发布

本节概述数据机构用于保护数据集中敏感信息的常见披露避免方法。受篇幅限制，论文将所有定理的证明放在附录中。

差分隐私。差分隐私（DP） [Dwork 等， 2006]是一种强隐私概念，用于量化并界定个体参与某项计算所导致的隐私损失。通俗地说，当数据集中的一条记录发生变化时，任何输出的概率都不会有太大变化，从而限制输出所泄露的有关任何个体的信息量。将数据集中的一条记录 D更改后得到一个新数据集 D'，这定义了邻接性这一概念，记作 $D \sim D'$。

定义 2. 一种机制 $\mathcal{M}$： $\mathcal{D} \to \mathcal{R}$ 的定义域为 $\mathcal{D}$，值域为 $\mathcal{R}$，满足(ϵ, δ)-差分隐私；对于任意两个输入$D \sim D' \in \mathcal{D}$以及任意输出响应子集 $R \subseteq \mathcal{R}$：

$$\Pr[\mathcal{M}(D) \in R] \le e^{\epsilon} \Pr[\mathcal{M}(D') \in R] + \delta.$$

参数 $\epsilon > 0$ 描述了算法的隐私损失，其值越接近 0，隐私性越强；而参数 $\delta \in [0, 1)$ 则表示算法未能满足 ϵ-差分隐私的概率。具体而言，用于直方图数据发布的拉普拉斯机制 定义为 $\mathcal{M}_{\mathrm{Lap}}(\boldsymbol{x}) = \boldsymbol{x} +$

Lap($2/\epsilon$)，其中 Lap(η) 是以 0 为中心、缩放因子为 η 的拉普拉斯分布，满足 (ϵ，0)-差分隐私。此外，Canonne等人提出的离散高斯机制 [， 2020], 定义为 $\mathcal{M}_{\text{Gaus}}(\boldsymbol{x}) = \boldsymbol{x} + \mathcal{N}_{\mathbb{Z}}(0, 4/\epsilon^2)$，其中 $\mathcal{N}_{\mathbb{Z}}(0, \sigma)$ 是均值为 0、标准差为 σ 的离散高斯分布，满足 ($\frac{1}{2}\epsilon^2 + \epsilon\sqrt{2\log(1/\delta)}$ δ)-差分隐私。

接下来，我们讨论三种主流的传统披露避免系统。与差分隐私不同，这些系统无法为隐私泄露提供形式化界限。

单元格抑制。 单元格抑制技术 [Kelly 等 1992], 常由统计机构采用（例如，[新西兰统计局 2020]），旨在数据发布前隐藏直方图中的低频计数。

定义 3。 给定一个直方图 $\boldsymbol{x}$ 和一个阈值 k，单元格抑制会返回一个隐私保护直方图 $\tilde{\boldsymbol{x}}$，其条目为

$$\tilde{\boldsymbol{x}}_i = \max\{\boldsymbol{x}_i, k/2\}. \tag{1}$$

图1(c)展示了如何将单元格抑制应用于图1(b)中的直方图，阈值为 $k = 2$。受影响的行计数以红色突出显示。这种方法的一个重要局限是，它只保护记录数量较少的敏感属性，而忽略其他属性。

交换。 交换 [Dalenius 和 Reiss 1982] 是一种机制，它会将某条记录中一组敏感属性（准标识符）的值与另一条记录中相应的值进行交换。简而言之，这些算法的基本步骤可概括如下：

1. 按照差异程度成比例的概率，从直方图中选取多对记录；2. 交换每对选定记录的准标识符属性值。

与单元格抑制类似，交换会生成一个保护隐私的直方图 $\tilde{\boldsymbol{x}}$。然而，与单元格抑制（以及差分隐私机制）不同，交换还需要一项额外信息：数据集的准标识符属性。图1(d)展示了交换的应用：将两行互换，并使用“性别”作为准标识符属性（见图(a)）。受影响的行计数以红色突出显示。虽然交换已被广泛使用，例如美国人口普查局会在较小的地理区域内交换相似个体的记录，但它并不能抵御重构攻击 [Garfinkel等人， 2019]。

k-匿名性。接下来， k-匿名性通过确保数据集中的每条记录都与至少 $k-1$ 条其他记录无法区分，从而保护数据集中的敏感数据。

定义 4(k-匿名 [Sweeney， 2002])。当数据集投影后仅包含准标识符，且每条记录至少出现 k 次时，当且仅当该数据集满足相对于一组准标识符的 k-匿名性。

k-匿名性的基本思想是，对数据集中个人的某些识别属性进行泛化，使得每个具有相似特征的群体至少包含 k 个人。以下概述了该算法（正式描述见附录 B）：

1. 为每个准标识符定义一个层次结构 H； 2. 构建一个直方图，列出每种准标识符组合对应的记录数量；

3. 抑制泛化直方图中实例数少于 k 的组合。

4. 发布生成的直方图 $\tilde{\boldsymbol{x}}(D, H)$。

一个重要的观察是，与前文回顾的方法不同， k-匿名性会在不同于 $\mathcal{X}$的空间中生成保护隐私的直方图 $\tilde{\boldsymbol{x}}(D, H)$。这一重要观察与误差分析相关。此外，该方法还需要访问准标识符属性以及泛化直方图。

图1(e)展示了将 k-匿名性与$k = 2$ 应用于图1(b)的直方图，使用泛化层次结构将男性和女性归为单一属性。尽管 k-匿名性被广泛用于发布统计数据和医疗数据，但它无法阻止利用外部公开数据发起的重新识别攻击 [Li等，2011]。

4 DAAnalysis 路线图

本节概述了本文其余部分所采用的方法。第 5 节介绍了传统 DA系统的差分隐私对应机制，包括单元格抑制、交换和 k-匿名性。其目的是说明，这些差分隐私对应机制保留了原始机制的主要特征，并在统一的直方图数据发布隐私设置下，对其隐私和误差进行分析 $\tilde{\boldsymbol{x}}(D)$。需要注意的是，经典差分隐私算法（例如，拉普拉斯机制）和单元格抑制均不对数据属性作任何假设。相比之下，交换依赖于准标识符的使用，而 k-匿名性还要求具备泛化层次结构。该层次结构迫使 k-匿名性在不同于 $\tilde{\boldsymbol{x}}(D)$ 的空间中生成直方图 $\tilde{\boldsymbol{x}}(D, H)$。虽然这不会影响隐私分析，从而可以在所有机制之间进行有意义的比较，但它给评估这些技术的性能带来了挑战。本文还通过提出一个统一的实证框架，根据原始数据空间比较各种技术的误差和偏差，来应对这一挑战$\mathcal{X}$。这需要对 k-匿名性执行重构步骤，具体内容见附录 B.3。需要认识到，尽管差分隐私披露避免机制与其传统披露避免对应机制具有许多共同特征，但不应将其视为传统机制的“加噪”版本。因此，文中呈现的分析和实验结果未必会显示误差随着隐私预算增加而降低。事实上，在某些隐私预算下，它们甚至可能比传统机制更加精确。

接下来，我们介绍传统披露避免系统的差分隐私版本及其隐私分析。这些分析针对给定的 ϵ 值，明确了 δ 参数的取值。第 6 节分析公平性结果。最后，第 7 节基于美国社区调查（ACS）数据的一个子集进行实验评估 [NIST， 2021]。

5 隐私与误差分析

本节介绍本文的第一项主要贡献。它提出了前文所述披露避免（DA）的差分隐私对应方法，并分析了其隐私保证和误差。本节首先给出一个技术引理，该引理规定了满足（ϵ， δ）-DP 的充分条件。该引理是

推导下文所述披露避免（DA）随机化版本的隐私保证的关键工具。

引理 1. 设 D、 D' 为数据集，满足 $D \sim D'$，令 S 定义为

$$S \coloneqq \left\{ \boldsymbol{o} \;\middle|\; \frac{\Pr\left(\mathcal{M}(D) = \boldsymbol{o}\right)}{\Pr\left(\mathcal{M}(D') = \boldsymbol{o}\right)} \leq \exp(\epsilon) \right\}$$

并令 S 表示 S的补集。如果

$$\Pr\left(\mathcal{M}(D) \in S^{\complement}\right) \leq \delta\,,$$

那么机制 $\mathcal{M}$ 是(ϵ, δ)-差分隐私的。

5.1 差分隐私单元格抑制

虽然单元格抑制保护了数据集中少数群体的隐私，却忽视了多数群体的隐私保护，因此不满足差分隐私的要求。事实上，这种机制具有确定性，无法针对两个相邻数据集生成不同的输出。更详细的讨论见附录 B。为解决这一问题，本文提出了单元格抑制的随机化版本，称为差分隐私单元格抑制。该机制记作 $\mathcal{M}_{\text{CS}}$，其为每个 $i \in [n]$ 发布一个隐私计数，具体如下：

$$\mathcal{M}_{\text{CS}}(D)_i = \hat{x}_i = \begin{cases} x_i & \text{if } x_i + \eta_i \geq k, \\ k/2 & \text{otherwise} \end{cases}, \tag{2}$$

其中， $\eta_i \sim \text{Lap}(2/\epsilon)$是一个加性噪声变量，取自中心为 0、尺度因子为 $2/\epsilon$ 的拉普拉斯分布； k为单元格抑制阈值。

机制 $\mathcal{M}_{\text{CS}}$ 与稀疏向量技术 （SVT） [Dwork 等人 2014] 有相似之处。给定一系列查询和一个实数阈值，SVT 会输出一个向量，指示每个（带噪声的）查询答案高于还是低于相应的（带噪声的）阈值。然而，两者存在三项根本差异：（1） $\mathcal{M}_{\text{CS}}$不会扰动阈值 k；（2）它生成数值输出，而*SVT*生成的是二元输出；（3）只要噪声计数高于阈值（公式 (2) 的第一个条件），它报告的就是真实计数，而不是噪声计数。

图 2（左）报告了 $\mathcal{M}_{\text{CS}}$ 在若干阈值 k （x轴）和参数 ϵ 下的实测误差。这些误差针对 ACS 马萨诸塞州数据集 [NIST 2021]（详见附录 C.1）给出：它们反映了单元格抑制机制与其差分隐私对应机制的直方图之间的 ℓ_1 距离 $\|\tilde{\boldsymbol{x}} - \boldsymbol{x}\|_1$。请注意， $\mathcal{M}_{\text{CS}}$ 产生的误差与原始机制的误差非常接近。这一点很重要，因为它使我们能够有意义地比较$\mathcal{M}_{\text{CS}}$ 与其他差分隐私机制： $\mathcal{M}_{\text{CS}}$ 的偏差与统计机构和组织目前广泛采用的传统单元格抑制机制相近。

隐私分析。</style}下一个定理给出了 $\mathcal{M}_{\text{CS}}$ 所提供的隐私保证。

定理1. 给定一个值 $\epsilon > 0$ 和阈值 $k < B$，机制 $\mathcal{M}_{CS}$ 是(ϵ, δ)-差分隐私的，其满足

$$\delta = 1 - \frac{1}{4}\exp\left(-\epsilon(B-k)\right)\,,$$

其中， B 是直方图条目的上界。

误差分析. 在考察隐私后，论文展示了 $\mathcal{M}_{\text{CS}}$ 返回的直方图 $\tilde{\boldsymbol{x}}$ 与原始直方图 $\boldsymbol{x}$有多接近。误差分析关注统计偏差，对于每个条目 $i \in [n]$，它可以表示为

$$\mathcal{B}\left(\mathcal{M}_{\text{CS}}\right)_i = \mathbb{E}\left[\mathcal{M}_{\text{CS}}(D)_i\right] - x_i = \left(\frac{k}{2} - x_i\right) \cdot \Pr\left(x_i + \eta_i < k\right).$$

请注意，只有当噪声计数低于阈值 k 时才会产生误差，其大小量化为阈值一半与真实计数之差。因此，下一个定理将与 $\mathcal{M}_{\text{CS}}$ 相关的误差同噪声计数低于阈值的概率，以及阈值一半与原始直方图计数之间的差值联系起来。

定理 2。 差分隐私单元格抑制机制 $\mathcal{M}_{CS}$ 的统计偏差可界定如下，

$$\|\boldsymbol{\mathcal{B}}\left(\mathcal{M}_{CS}\right)\|_1 \leq \|k/2 \cdot \mathbf{1}_n - \boldsymbol{x}\|_2 \cdot \|\boldsymbol{p}\|_2\ ,$$

其中， $\boldsymbol{p}$是向量的简写形式

$$\boldsymbol{p} \coloneqq \left[\Pr\left(x_1 + \eta_1 < k\right) \quad \ldots \quad \Pr\left(x_n + \eta_n < k\right)\right]. \tag{3}$$

5.2 差分隐私交换

尽管交换机制具有随机性，它仍不满足差分隐私的要求。为说明其为何失效，我们来看一对相邻数据集 D 和 D'。假设 D' 中有一条记录，即 $\boldsymbol{a}_1$，其在任何属性 $A \in Q$上都不与 D 中的任何记录匹配。无论如何执行交换，都无法从输入数据集 D生成记录 $\boldsymbol{a}_1$。

为了获得交换的差分隐私对应机制，因此，推理准标识符全集而非数据库中实际存在的准标识符集合至关重要。令$\mathcal{X}_Q = \{\boldsymbol{q}_1, \ldots, \boldsymbol{q}_{n_Q}\}$表示准标识符的数据全集。该机制不会交换准标识符，而是从 $\mathcal{X}_Q$中随机选择一些准标识符。该机制称为差分隐私交换，记作$\mathcal{M}_{\text{SW}}$，其工作方式如下：对于每个 $\boldsymbol{a}_i \in \mathcal{X}$，考虑对 $(\boldsymbol{a}_i,\ x_i)$，表示元组及其在直方图 $\boldsymbol{x}$中的相应计数。 $\mathcal{M}_{\text{SW}}$ 定义 $\tilde{\boldsymbol{a}}_i[N] = \boldsymbol{a}_i[N]$ 并

$$\tilde{\boldsymbol{a}}_i[Q] = \begin{cases} \boldsymbol{a}_i[Q] & \text{w.p. } \gamma = \frac{\exp(\epsilon)}{\exp(\epsilon) + n_Q - 1}, \\ \text{Uniform}(\mathcal{X}_Q \setminus \boldsymbol{a}_i[Q]) & \text{w.p. } 1 - \gamma \end{cases} \tag{4}$$

其中，Uniform(C) 表示事件空间 C 上的均匀概率。上述步骤的结果可能会产生多个条目（$\tilde{\boldsymbol{a}}_i, x_i$）和（$\tilde{\boldsymbol{a}}_j, x_j$），且具有 $\tilde{\boldsymbol{a}}_i = \tilde{\boldsymbol{a}}_j$。该过程通过对各种 x_i 和 x_j 求和，将所有此类元组合并。随后，将得到的子直方图扩展为直方图$\tilde{\boldsymbol{x}}$。请注意， $\mathcal{M}_{\text{SW}}$ 只修改准标识符，并生成隐私保护直方图 $\tilde{\boldsymbol{x}}(D)$，与原始交换算法的做法类似。

图 2（中）比较了在交换不同比例（%）的行以及不同参数 ϵ 的情况下，由 $\mathcal{M}_{\text{SW}}$ 生成的直方图与其传统对应方法生成的直方图之间的 ℓ_1 距离 $\|\tilde{\boldsymbol{x}} - \boldsymbol{x}\|_1$。再次观察可知，这两种机制的误差非常接近。

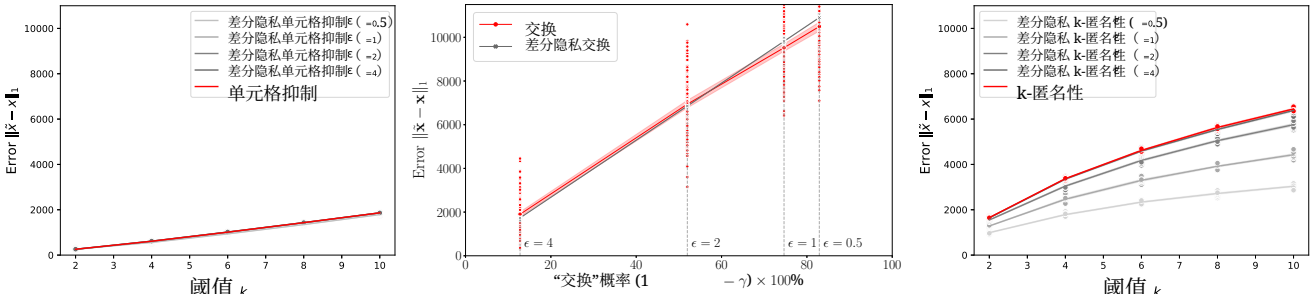


图 2：MA ACS 数据集：单元格抑制（左）、交换（中）和 k-匿名性（右）及其差分隐私对应方法的误差（重复 200 次的平均值）。

隐私分析。请回想，差分隐私保护的是参与数据集的任何个人用户的信息披露。另一方面，差分隐私交换作用于直方图计数这一层级，其运作方式会妨碍依赖纯 $(\epsilon, 0)$-差分隐私的分析。$\mathcal{M}_{\text{SW}}$ 的隐私分析见下述定理。

定理3。对于给定的 $\epsilon > 0$，$\mathcal{M}_{SW}$ 算法是（ϵ、δ）-差分隐私的，其中 δ 为

$$1 - \frac{1-\gamma^2}{n_Q - 1} - \left(\frac{1-\gamma}{n_Q - 1}\right)^2,$$

其中，γ 在方程(4) 中定义，且 $n_Q = |\mathcal{X}_Q|$,

误差分析。接下来，我们讨论交换及其差分隐私对应方法返回的直方图误差有多接近。

命题1。差分隐私交换直方图中每个元素 $i \in [n]$的偏差可表示为

$$\mathcal{B}\left(\mathcal{M}_{SW}\right)_i = \frac{\sum_{j\in\mathcal{I}_i} x_j - n_Q \cdot x_i}{\exp(\epsilon) + n_Q - 1},$$

其中，索引集 $\mathcal{I}_i$ 收集了数据全集 $\mathcal{X} = \{\boldsymbol{a}_j \mid j \in [n]\}$ 中所有与 $\boldsymbol{a}_i$具有相同非准标识符 N 的元素，即，

$$\mathcal{I}_i := \{j \in [n] \mid \boldsymbol{a}_j[N] = \boldsymbol{a}_i[N]\}. \tag{5}$$

定理4。差分隐私交换机制 $\mathcal{M}_{SW}$ 的统计偏差可表示如下，

$$\|\boldsymbol{\mathcal{B}}\left(\mathcal{M}_{SW}\right)\|_1 = \frac{n_Q}{\exp(\epsilon) + n_Q - 1} \sum_{i=1}^{n} D_{\text{mean}}\left(\boldsymbol{x}_{\mathcal{I}_i}\right),$$

其中，$\mathcal{I}_i$ 是在方程(5) 中定义的索引集，$\boldsymbol{x}_{\mathcal{I}_i}$ 是约化直方图，其中包含任意 $j \in \mathcal{I}_i$对应的计数 x_j。此外，D_{mean} $(\boldsymbol{x}_{\mathcal{I}_i})$ 是直方图 $\boldsymbol{x}_{\mathcal{I}_i}$的平均绝对偏差，即，

$$D_{\text{mean}}\left(\boldsymbol{x}_{\mathcal{I}_i}\right) := \frac{1}{n_Q} \sum_{j\in\mathcal{I}_i} \left| x_j - \frac{\sum_{l\in\mathcal{I}_i} x_l}{n_Q} \right|.$$

5.3 差分隐私 k-匿名

与单元格抑制类似，k-匿名是一种确定性算法，无法产生满足差分隐私要求的输出。然而，近年来，已有多项尝试旨在将 k-匿名与差分隐私相结合。特别是，Li 等人 [2011] 提出了一种机制，将泛化直方图和参数为 k 的单元格抑制应用于原始数据集的一个子抽样版本。本文提出了该机制的一种推广。

差分隐私 k-匿名算法。本文对上一节介绍的 k-匿名算法进行了简单修改。该机制称为差分隐私-k-匿名，记作 $\mathcal{M}_{\text{KA}}$，分两步执行：

1. 生成数据集 D 的子采样版本 D_β，其中每一行以概率 $\beta \in (0, 1)$ 保留。2. 在 D_β 上应用经典的 k-匿名算法。

与 [Li 等人不同，2011]，$\mathcal{M}_{\text{KA}}$ 使得合并单元格时可以使用泛化层次结构，就像在其确定性对应机制中一样。图 2（右）报告了 $\mathcal{M}_{\text{KA}}$ 在多个 k 取值（x 轴）下的实测误差；β的取值由隐私参数 ϵ 的选择决定（见定理 5）。图中报告了以下直方图之间的 ℓ_1 距离：在原始空间 $\mathcal{X}$ 中，由泛化后的差分隐私 k 匿名化结果重构得到的直方图 $\boldsymbol{x}(D)$，以及通过其确定性对应机制得到的直方图。附录 B 介绍了所采用的重构步骤。再次注意，差分隐私与确定性 k-匿名对应机制造成的误差非常接近：事实上，当 ϵ 取值较小时，差分隐私版本的表现优于确定性机制，这是抽样过程所致，因为该过程使用的记录较少。

隐私分析。下一个结果推广了 [Li 等,2011] 并报告了 $\mathcal{M}_{\text{KA}}$所提供的隐私保证。

定理 5。对于给定的 $k > 0$ 和采样概率 $\beta \in (0, 1)$，$\mathcal{M}_{KA}$ 对于 $\delta = d(k, \beta, \epsilon)$满足$(\epsilon, \delta)$-差分隐私，其中函数 d定义为

$$d(k, \beta, \epsilon) = 1 - \min_{w\in[B]} \left(\sum_{j=0}^{\nu} f(j; w, \beta) \right)^2, \tag{6}$$

ν 是 $\lfloor (1 - \exp(-\epsilon))w \rfloor$ 的简写，f 表示二项随机变量的概率质量函数，

$$f(j; w, \beta) = \binom{w}{j} \beta^w (1-\beta)^{w-j}, \quad \forall\, j \in [w], \tag{7}$$

误差分析。本次分析的目标仍然是表明，k-匿名的 DP 版本在误差方面接近于其

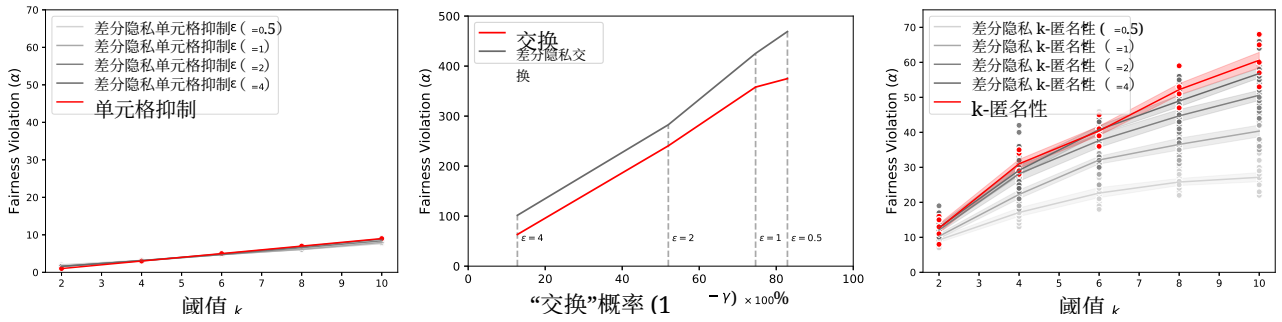


图 3：MA ACS 数据集：单元格抑制（左）、交换（中）和 k-匿名性（右）的公平性值 α 及其差分隐私对应方法（200 次重复的平均值）。

确定性对应机制。回顾一下，与单元格抑制和交换不同， k-匿名并不会返回一个处于属性全集相同空间中的直方图，这是由于其应用了泛化层次结构。因而， k-匿名的输出由相对于数据全集 $\mathcal{X}_H$ 泛化层次结构的个体记录计数组成。由于这一关键差异，本节其余部分通过界定直方图 $\boldsymbol{x}(D_\beta)$（在步骤 1 中生成）中的某个计数是否会在 k-匿名化步骤（步骤 2）中被合并，来分析该机制的误差。

令 $\mathcal{M}^{\beta}{}_{\mathrm{KA}}$ 表示二进制向量

$$\mathcal{M}^{\beta}_{\mathrm{KA}}(D_\beta) := [\mathbf{1}\{x_1(D_\beta) < k\} \quad \ldots \quad \mathbf{1}\{x_n(D_\beta) < k\}].$$

误差分析聚焦于统计偏差，即计数是否会被泛化层级合并，也就是 $\mathcal{M}^{\beta}{}_{\mathrm{KA}}(D_\beta)$ 与 $\mathcal{M}^{\beta}{}_{\mathrm{KA}}(D)$ 之间的差异：

$$\begin{aligned}\mathcal{B}\left(\mathcal{M}^{\beta}_{\mathrm{KA}}\right) &= \mathbb{E}\left[\mathcal{M}^{\beta}_{\mathrm{KA}}(D_\beta)\right] - \mathcal{M}^{\beta}_{\mathrm{KA}}(D) \\ &= \begin{bmatrix} \mathbb{E}\left[\mathbf{1}\{x_1(D_\beta) < k\}\right] - \mathbf{1}\{x_1 < k\} \\ \vdots \\ \mathbb{E}\left[\mathbf{1}\{x_n(D_\beta) < k\}\right] - \mathbf{1}\{x_n < k\} \end{bmatrix}^{\top}.\end{aligned}$$

接下来，我们证明计数 $x_i(D_\beta)$ 与二项随机变量 $B(x_i, \beta)$ 等价，并给出刻画偏差的数学表达式。

定理 6。 差分隐私 k-匿名机制 $\mathcal{M}^{\beta}{}_{KA}$ 的统计偏差可表示为：对每个 $i \in [n]$,

$$\begin{aligned}\mathcal{B}\left(\mathcal{M}^{\beta}_{KA}\right)_i &= \mathbb{E}\left[\mathbf{1}\{x_i(D_\beta) < k\}\right] - \mathbf{1}\{x_i < k\} \\ &= \begin{cases} \sum_{j=x_i-k+1}^{x_i} \binom{x_i}{j} \beta^{x_i-j}(1-\beta)^j, & x_i \geq k, \\ 0, & \text{otherwise.} \end{cases}\end{aligned}$$

6 公平性分析

本文的第二项主要贡献是分析各种差分隐私 DA 算法相较于传统差分隐私的公平性。本文采用的公平性定义（定义 1）是隐私保护直方图中偏差的最大差值。需要指出的是，本文在不同于其他机制的背景下，考察了利用泛化直方图的 DP k-匿名算法的偏差

。因此，本文特别侧重于对 DP k-匿名算法公平性的分析比较。

下一个结果量化了差分隐私单元格抑制和交换以及拉普拉斯机制所带来的不公平性。

定理 7 (α-公平性，适用于 $\mathcal{M}_{\mathrm{CS}}$)。 差分隐私单元格抑制算法满足 α_{CS}-公平性，其中 α_{CS} 为

$$(x_n - x_1)\, p_1 + \max\left\{\left|\frac{k}{2} - x_1\right|, \left|\frac{k}{2} - x_n\right|\right\}(p_1 - p_n),$$

其中， p_1 和 p_n 分别是方程(3) 中定义的 $\boldsymbol{p}$ 的首项和末项。

定理 8 (α-公平性适用于 $\mathcal{M}_{\mathrm{SW}}$)。 差分隐私交换算法 $\mathcal{M}_{SW}$ 是 α_{SW}-公平的，其中 α_{SW} 由下式给出

$$\frac{2n_Q \|\boldsymbol{x}\|_{\rightleftharpoons}}{\exp(\epsilon) + n_Q - 1} = \frac{2n_Q}{\exp(\epsilon) + n_Q - 1}(x_n - x_1).$$

定理 9 (α-公平性适用于 $\mathcal{M}_{\mathrm{Lap}}$)。 拉普拉斯机制$\mathcal{M}_{Lap}$ 是 α_{Lap}-公平的，其中 α_{Lap} 由下式给出

$$\frac{\exp(-\epsilon x_1/2)}{2}\|\boldsymbol{x}\|_{\rightleftharpoons} = \frac{\exp(-\epsilon x_1/2)}{2}(x_n - x_1).$$

图 3 展示了公平性违规值，即 α的值，比较了单元格抑制、交换和 k-匿名，以及它们对应的差分隐私机制；这些结果涵盖不同的隐私参数 ϵ 和 k的不同取值（针对单元格抑制和 k-匿名），或不同的交换行的百分比（针对交换）。可以观察到，差分隐私机制的公平性违规程度与传统对应机制相当，甚至更低。随着隐私参数 ϵ 增大，这一点尤其值得注意。如前所述，重要的是要记住，差分隐私机制并非传统对应机制的"加噪"版本，而是在概念上相似的机制。因此，与传统对应机制相比，它们的公平性违规程度可能更低，差分隐私 k-匿名就是一例。

以下定理是本文的第三个关键结果。它证明了在公平性误差方面，拉普拉斯机制优于差分隐私单元格抑制和交换。

定理 10。 假设原始直方图 $\boldsymbol{x}(D)$ 的最小计数介于2 和阈值 k 之间，即 $2 \leq$

ϵ	机制	δ	偏差（ℓ_1 范数）	α-公平性
0.5	拉普拉斯	**0**	**763.775**	**3.655**
	离散高斯机制	0.363	980.81	4.945
	差分隐私抑制	0.999	935.525	4.345
	差分隐私交换	0.868	10906.79	469.015
	差分隐私 k-匿名	0.878	2337.8	22.65
1	拉普拉斯	**0**	**342.885**	**1.845**
	离散高斯机制	0.132	659.215	3.065
	差分隐私抑制	0.999	1003.035	4.5
	差分隐私交换	0.874	9859.26	425.335
	差分隐私 k-匿名性	0.906	3297.4	32.1
2	拉普拉斯	**0**	**154.78**	**0.905**
	离散高斯机制	0.017	436.925	2.2
	差分隐私抑制	0.999	1018.335	4.72
	差分隐私交换	0.899	6841.19	282.73
	差分隐私 k-匿名性	0.981	4175.5	37.65
4	拉普拉斯	**0**	**67.34**	**0.465**
	离散高斯机制	3E-4	290.715	1.495
	差分隐私抑制	0.999	1014.63	4.92
	差分隐私交换	0.969	1664.63	101.645
	差分隐私 k-匿名	0.999	4590.7	40.75

表 1：MA 数据集数据发布：从经验偏差的 δ、ℓ_1 范数和 α-公平性方面比较差分隐私机制。

$x_1 \leq k$。那么，拉普拉斯机制所对应的公平性误差不大于差分隐私单元格抑制或差分隐私交换机制所对应的误差，即：

$$\alpha_{Lap} \leq \alpha_{CS} \quad and \quad \alpha_{Lap} \leq \alpha_{SW}.$$

值得注意的是，k-匿名机制作用于不同于原始直方图空间的空间，因此无法从理论上比较拉普拉斯机制与 DP k-匿名机制。不过，本文接下来展示了实证证据，表明拉普拉斯机制同样显著优于 DP k-匿名机制。

7 实验评估

本研究评估传统数据匿名化机制的差分隐私变体的性能，并将其与第 3 节介绍的两种关键差分隐私机制——拉普拉斯机制和离散高斯机制——进行比较。实验使用了马萨诸塞州、得克萨斯州和 Outlier[NIST, 2021] 的 2019 年美国社区调查 IPUMS 数据集。所有实验均报告 200 次重复的平均结果。后两个数据集的结果列于附录中，因为它们的趋势与前一个数据集相似。附录还对数据集和实验设置作了更详尽的说明。本节重点评估这几种机制在两种场景中的表现：数据发布和分类。

数据发布。第一项任务比较了由所研究的各种差分隐私机制生成的直方图重建得到的数据集。有关重建算法的详细信息，请参阅附录 B。表 1 评估了传统披露避免机制的差分隐私变体的性能，
机制以及拉普拉斯机制和离散高斯机制在误差和公平性违规方面的表现。对于可能产生负计数的机制，采用简单的后处理，将结果投影到非负正交象限。

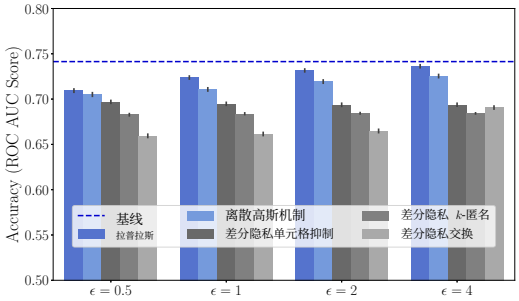


图 4：MA 数据集：逻辑回归结果。

这些结果尤为重要：与普遍看法相反，它们表明，经典的差分隐私算法不仅能提供强有力的隐私保证（见 δ数值），还能够生成直方图，在准确性（见偏差列）和公平性指标（见 α-公平性列）方面均优于传统数据增强机制。因此，机构希望发布数据集时，宜考虑采用传统差分隐私机制。

分类。比较所有差分隐私机制在分类任务中的表现同样十分重要。该设置采用通过数据发布查询获得的私有数据集来训练逻辑回归分类器。任务是预测某人年收入是否超过 50,000 美元，图 4 中的结果以其在原始非私有数据集上的准确性表示。请注意，与其他传统披露避免机制生成的数据相比，拉普拉斯机制和离散高斯机制所产生的分类器具有高得多的准确率。值得注意的是，拉普拉斯机制和离散高斯机制的分类准确率与基线方法（在非私有数据集上训练）更加接近，优于任何其他方法。同样，这一点意义重大：尽管这些任务十分简单，但它们是数据机构和组织日常开展的大量统计分析的基础。

8 结论

本文提出了一个框架，用于比较传统披露规避系统（DA）与差分隐私。文中提出了三种广泛采用的传统DA方法——抑制、交换和k-匿名性——经过精心随机化的版本，并推导出这些机制的(ϵ, δ)-差分隐私界限。本文还对这些差分隐私算法进行了实证分析，结果表明，无论在准确性还是公平性方面，它们都与相应的传统方法相近。随后，本文通过实验，将这些差分隐私披露避免机制与传统差分隐私机制（即拉普拉斯机制或离散高斯机制）在数据发布和分类任务上的表现进行了比较。与普遍看法相反，实验评估表明，在隐私级别相同的情况下，经典差分隐私机制在准确性和公平性方面可能优于传统DA方法。这项研究有望影响数据机构和组织未来采取披露避免措施的方式，因为它提供了一个框架，使人们能够比较传统DA与差分隐私各自的优势和局限。

伦理声明

从伦理角度来看，本研究的目的并非认可那些尚未充分考虑其行为所带来隐私影响的机构发布数据。本研究不应被视为贬低传统DA方法的手段。此外，文中所呈现的实证分析应被理解为仅适用于本研究讨论的机制和数据集。

还必须考虑该研究的潜在益处，例如，采用传统差分隐私工具可能提高数据发布的准确性和公平性。此外，该研究还有望推动更有效的隐私保护技术的发展。

致谢

本研究部分得到了美国国家科学基金会资助项目 2133169、人工智能研究所项目 2112533 以及美国国家科学基金会 CAREER 奖 2143706 的支持。Fioretto 还获得了谷歌研究学者奖和亚马逊研究奖的支持。文中观点和结论仅代表作者本人。作者还要感谢 Matt Williams，就交换及其他传统数据匿名化技术所展开的有益讨论。

参考文献

Cl´ement L Canonne、Gautam Kamath 和 Thomas Steinke。用于差分隐私的离散高斯机制。神经信息处理系统进展，33:15676–15688，2020年。FH Clarke 和 Yu S Ledyaev。中值不等式。美国数学学会会刊，第 1075–1083 页，1994年。Tore Dalenius 和 Steven P Reiss。数据交换：一种披露控制技术。统计规划与推断杂志，6(1):73–85，1982年。Cynthia Dwork、Frank McSherry、Kobbi Nissim 和 Adam Smith。为私有数据分析校准噪声与敏感度。在 密码学理论会议，第 265–284 页。施普林格，2006年。Cynthia Dwork、Aaron Roth 等。差分隐私的算法基础。理论计算机科学基础与趋势®，9(3–4):211–407，2014年。Ferdinando Fioretto、Cuong Tran、Pascal Van Hentenryck 和 Keyu Zhu。决策与学习任务中的差分隐私和公平性：综述。在 国际人工智能联合会议（*IJCAI*）论文集，第 5470–5477 页，2022年。Simson Garfinkel、John M Abowd 和 Christian Martindale。理解针对公开数据的数据库重构攻击。美国计算机协会通讯，62(3):46–53，2019年。James P Kelly、Bruce L Golden 和 Arjang A Assad。单元格抑制：敏感表格数据的披露保护。网络，22(4):397–417，1992年。

Satya Kuppam、Ryan McKenna、David Pujol、Michael Hay、Ashwin Machanavajjhala 和 Gerome Miklau。使用隐私保护数据进行公平决策。*arXiv* 预印本 *arXiv:1905.12744*，2019年。K. LeFevre、D.J. DeWitt 和 R. Ramakrishnan。Mondrian 多维 k-匿名。在第 *22* 届数据工程国际会议（*ICDE'06*），第 25–25 页，2006年。Ninghui Li、Wahbeh H Qardaji 和 Dong Su。可证明的隐私数据匿名化：或者，k-匿名性遇上差分隐私。*CoRR*，*abs/1101.2604*，49：55，2011年。NIST。SDNist v1.4 测试版：合成数据报告工具。美国国家标准与技术研究院，2021年。Latanya Sweeney。k-匿名性：一种保护隐私的模型。国际不确定性、模糊性与知识型系统杂志，10(05)：557–570，2002年。新西兰统计局。微观数据输出指南。https://www.stats.govt.nz/assets/Methods/Microdata-Output-Guide-2020-v5-Sept22update.pdf，2020年。Cuong Tran、Ferdinando Fioretto、Pascal Van Hentenryck 和 Zhiyan Yao。公平性视角下的差分隐私决策。在国际人工智能联合会议（*IJCAI*），第 560–566 页，2021年。Keyu Zhu、Ferdinando Fioretto 和 Pascal Van Hentenryck。差分隐私数据的后处理：公平性视角。在 Lud De Raedt 主编的第三十一届国际人工智能联合会议论文集，*IJCAI-22*，第 4029–4035 页。国际人工智能联合会议组织，2022年 7 月。主会场。

缺失的证明

引理 *1* 的证明。对于任意输出 $O \subseteq \mathcal{R}$ 和相邻数据集 D, $D' \in \mathcal{X}^m$,

$$\begin{aligned}
&\Pr(\mathcal{M}(D) \in O) \\
=\ &\Pr(\mathcal{M}(D) \in (O \cap S)) + \Pr\left(\mathcal{M}(D) \in \left(O \cap S^{\complement}\right)\right) \\
\leq\ &\int_{\boldsymbol{o} \in (O \cap S)} \Pr(\mathcal{M}(D) = \boldsymbol{o})\, d\boldsymbol{o} + \Pr\left(\mathcal{M}(D) \in S^{\complement}\right) \\
\leq\ &\exp(\epsilon) \cdot \int_{\boldsymbol{o} \in (O \cap S)} \Pr(\mathcal{M}(D') = \boldsymbol{o})\, d\boldsymbol{o} + \delta \\
\leq\ &\exp(\epsilon) \cdot \int_{\boldsymbol{o} \in O} \Pr(\mathcal{M}(D') = \boldsymbol{o})\, d\boldsymbol{o} + \delta \\
=\ &\Pr(\mathcal{M}(D') \in O) + \delta.
\end{aligned}$$

□

定理*1*的证明。不失一般性，假设相邻数据集 D, $D' \in \mathcal{X}^m$ 的最后一条记录不同，其中 $\boldsymbol{a}_n$ 位于 D 中，$\boldsymbol{a}_{n-1}$ 位于 D' 中。这意味着 D 和 D' 的直方图在最后两个条目上有所不同，即

$$\begin{aligned}
\boldsymbol{x}(D') &= [x'_1 \quad \dots \quad x'_{n-1} \quad x'_n] \\
&= [x_1 \quad \dots \quad x_{n-1} + 1 \quad x_n - 1].
\end{aligned}$$

考虑以下集合

$$\underline{S} = \left\{ \boldsymbol{o} \in \mathbb{N}_+^n \;\middle|\; o_{n-1} = o_n = \frac{T}{2} \right\}.$$

对于任意元素 $\boldsymbol{o} \in \underline{S}$，可得

$$\begin{aligned}
&\frac{\Pr(\mathcal{M}_{\mathsf{CS}}(D) = \boldsymbol{o})}{\Pr(\mathcal{M}_{\mathsf{CS}}(D') = \boldsymbol{o})} \\
=\ &\frac{\Pr(x_{n-1} + \eta_{n-1} < k)}{\Pr\left(x'_{n-1} + \eta'_{n-1} < k\right)} \cdot \frac{\Pr(x_n + \eta_n < k)}{\Pr(x'_n + \eta'_n < k)} \qquad (8) \\
=\ &\frac{\int_{-\infty}^{k} \exp\left(-\frac{\epsilon}{2}\left|v - x_{n-1}\right|\right) dv}{\int_{-\infty}^{k} \exp\left(-\frac{\epsilon}{2}\left|v - x'_{n-1}\right|\right) dv} \cdot \frac{\int_{-\infty}^{k} \exp\left(-\frac{\epsilon}{2}\left|v - x_n\right|\right) dv}{\int_{-\infty}^{k} \exp\left(-\frac{\epsilon}{2}\left|v - x'_n\right|\right) dv} \\
\leq\ &\exp\left(\frac{\epsilon}{2}\right) \cdot \exp\left(\frac{\epsilon}{2}\right) \qquad (9) \\
=\ &\exp(\epsilon),
\end{aligned}$$

其中，在公式 (8) 中，η_{n-1}、η'_{n-1}、η_n 和 η'_n 均为参数为 $2/\epsilon$ 的独立同分布拉普拉斯分布随机变量。此外，公式 (9) 来源于三角不等式。该不等式意味着，在引理 1 中，集合 S 是 S 的子集。因此，概率 $\Pr\left(\mathcal{M}_{\mathsf{CS}}(D) \in S\right)$ 可按如下方式计算

$$\begin{aligned}
&\Pr\left(\mathcal{M}_{\mathsf{CS}}(D) \in S^{\complement}\right) \\
\leq\ &\Pr\left(\mathcal{M}_{\mathsf{CS}}(D) \in \underline{S}^{\complement}\right) = 1 - \Pr(\mathcal{M}_{\mathsf{CS}}(D) \in \underline{S}) \\
=\ &1 - \Pr(x_{n-1} + \eta_{n-1} < k) \cdot \Pr(x_n + \eta_n < k) \\
\leq\ &1 - \Pr(B + \eta_{n-1} < k) \cdot \Pr(B + \eta_n < k) \qquad (10) \\
=\ &1 - \frac{1}{4}\exp(-\epsilon(B - k)), \qquad (11)
\end{aligned}$$

其中，公式 (10) 来源于以下事实：函数 $x \mapsto \Pr(x + \eta < k)$ 在 $\eta \sim \mathsf{Lap}(2/\epsilon)$ 条件下是递减的，并且假设直方图 $\boldsymbol{x}(D)$ 的每个元素均由常数 B 限制。此外，公式 (11) 源于假设 $k < B$。根据引理 1，机制 $\mathcal{M}_{\mathsf{CS}}$ 的隐私保证得以确立，从而完成证明。

□

定理 *2* 的证明。根据柯西-施瓦茨不等式，

$$\begin{aligned}
\|\boldsymbol{\mathcal{B}}(\mathcal{M}_{\mathsf{CS}})\|_1 &= \sum_{i=1}^{n} \left| \left(\frac{k}{2} - x_i\right) \cdot \Pr(x_i + \eta_i < k) \right| \\
&\leq \left\| \frac{k}{2} \cdot \mathbf{1}_n - \boldsymbol{x} \right\|_2 \cdot \|\boldsymbol{p}\|_2 .
\end{aligned}$$

□

定理*3*的证明。不失一般性，假设相邻数据集 D, $D' \in \mathcal{X}^m$ 的最后一条记录不同，其中 $\boldsymbol{a}_n$ 位于 D 中，$\boldsymbol{a}_{n-1}$ 位于 D' 中，且 $\boldsymbol{a}_n[Q] \neq \boldsymbol{a}_{n-1}[Q]$ 和 $\boldsymbol{a}_n[N] = \boldsymbol{a}_{n-1}[N]$。考虑以下集合

$$S = \left\{ \tilde{\boldsymbol{x}} \in \mathbb{N}_+^n \;\middle|\; \tilde{\boldsymbol{a}}_{n-1}[Q] = \tilde{\boldsymbol{a}}_n[Q] \right\},$$

其中，$\tilde{\boldsymbol{x}}$ 是与由差分隐私交换机制生成的 $\{\tilde{\boldsymbol{a}}_i \mid i \in [n]\}$ 相关联的结果直方图。显然，对于任意 $\tilde{\boldsymbol{x}} \in S$,

$$\exp(-\epsilon) \leq \frac{\Pr(\mathcal{M}_{\mathsf{SW}}(D) = \tilde{\boldsymbol{x}})}{\Pr(\mathcal{M}_{\mathsf{SW}}(D') = \tilde{\boldsymbol{x}})} \leq \exp(\epsilon).$$

因此可得

$$\begin{aligned}
&\Pr\left(\mathcal{M}_{\mathsf{SW}}(D) \in S^{\complement}\right) \\
=\ &1 - \Pr(\mathcal{M}_{\mathsf{SW}}(D) \in S) \\
=\ &1 - \sum_{j \in \mathcal{I}_n} \Pr(\tilde{\boldsymbol{a}}_{n-1}[Q] = \tilde{\boldsymbol{a}}_n[Q] = \boldsymbol{a}_j[Q]) \\
=\ &1 - 2\frac{\gamma(1-\gamma)}{n_Q - 1} - (n_N - 2)\left(\frac{1-\gamma}{n_Q - 1}\right)^2 \\
=\ &1 - \frac{1-\gamma^2}{n_Q - 1} - \left(\frac{1-\gamma}{n_Q - 1}\right)^2 .
\end{aligned}$$

根据引理 1，它为差分隐私交换机制提供了隐私保证。

□

定理 *5* 的证明。为了推导差分隐私 k匿名机制的隐私保证，只需证明样本数据集 D_β 满足 (ϵ, δ)-差分隐私，其中δ 在公式 (6) 中定义。这是因为后处理免疫性这一重要性质 [Dwork等， 2014]。

在分析之前，令 $\mathcal{M}_\beta$ 表示生成样本数据集的随机机制。不失一般性，假设相邻数据集 D, $D' \in \mathcal{X}^m$ 的最后一条记录不同，其中 $\boldsymbol{a}_n$ 位于 D，而 $\boldsymbol{a}_{n-1}$ 位于D'。这意味着 D 和 D' 的直方图在最后两个条目上不同，即，

$$\begin{aligned}
\boldsymbol{x}(D') &= [x'_1 \quad \dots \quad x'_{n-1} \quad x'_n] \\
&= [x_1 \quad \dots \quad x_{n-1} + 1 \quad x_n - 1].
\end{aligned}$$

考虑集合

$$S_{D,D'} := \left\{ D_\beta \;\middle|\; \begin{array}{l} x(D_\beta)_n \le \lfloor (1-\exp(-\epsilon)) \cdot x_n \rfloor \\ x(D_\beta)_{n-1} \le \lfloor (1-\exp(-\epsilon)) \cdot (x_{n-1}+1) \rfloor \end{array} \right\}.$$

对于任意样本数据集$D_\beta \in S_{D,\ D'}$,

$$\begin{aligned} & x(D_\beta)_n \le \lfloor (1-\exp(-\epsilon)) \cdot x_n \rfloor , \\ \Longrightarrow \quad & x(D_\beta)_n \le (1-\exp(-\epsilon)) \cdot x_n , \\ \Longrightarrow \quad & \exp(-\epsilon)\, x_n \le x_n - x(D_\beta)_n \\ \Longrightarrow \quad & \frac{x_n}{x_n - x(D_\beta)_n} \le \exp(\epsilon) . \end{aligned}$$

注意，计数 $x(D_\beta)_n$ 是非负的，这意味着

$$1 \le \frac{x_n}{x_n - x(D_\beta)_n} \le \exp(\epsilon) . \tag{12}$$

同样，对于属性 $\boldsymbol{a}_{n-1}$，以下不等式成立

$$\exp(-\epsilon) \le \frac{x_{n-1}+1-x(D_\beta)_{n-1}}{x_{n-1}+1} \le 1 . \tag{13}$$

因此,

$$\begin{aligned} & \frac{\Pr\left(\mathcal{M}^\beta_{\mathrm{KA}}(D) = D_\beta\right)}{\Pr\left(\mathcal{M}^\beta_{\mathrm{KA}}(D') = D_\beta\right)} \\ = \; & \frac{\Pr\left(x(D_\beta)_{n-1} \mid x_{n-1}\right)}{\Pr\left(x(D_\beta)_{n-1} \mid x(D')_{n-1}\right)} \cdot \frac{\Pr\left(x(D_\beta)_n \mid x_n\right)}{\Pr\left(x(D_\beta)_n \mid x(D')_n\right)} \\ = \; & \frac{\binom{x_n}{x(D_\beta)_n}\binom{x_{n-1}}{x(D_\beta)_{n-1}}\left(\frac{\beta}{1-\beta}\right)^{x(D_\beta)_n + x(D_\beta)_{n-1}}(1-\beta)^{x_n+x_{n-1}}}{\binom{x_n-1}{x(D_\beta)_n}\binom{x_{n-1}+1}{x(D_\beta)_{n-1}}\left(\frac{\beta}{1-\beta}\right)^{x(D_\beta)_n + x(D_\beta)_{n-1}}(1-\beta)^{x_n+x_{n-1}}} \\ = \; & \frac{x_n}{x_n - x(D_\beta)_n} \cdot \frac{x_{n-1}+1-x(D_\beta)_{n-1}}{x_{n-1}+1} . \end{aligned}$$

于是，由公式 (12) 和 (13),

$$\exp(-\epsilon) \le \frac{\Pr\left(\mathcal{M}_\beta(D) = D_\beta\right)}{\Pr\left(\mathcal{M}_\beta(D') = D_\beta\right)} \le \exp(\epsilon) .$$

因此，对于任意相邻数据集 D 和 D', $S_{D,D'}$ 结果是一个集合，其中的每个元素都由 $\mathcal{M}_\beta(D)$ 和 $\mathcal{M}_\beta(D')$ 以相近的概率生成。根据引理 1，错误概率 δ 可计算为

$$\begin{aligned} \delta &= \max_{D \sim D'} \Pr\left(\mathcal{M}_\beta(D) \in S^{\complement}_{D,D'}\right) \\ &= 1 - \min_{D \sim D'} \Pr\left(\mathcal{M}_\beta(D) \in S_{D,D'}\right) \\ &= 1 - \min_{w \in [B]} \left(\sum_{j=0}^{\nu} f(j; w, \beta) \right)^2 , \end{aligned}$$

其中， ν 是 $\lfloor (1-\exp(-\epsilon))w \rfloor$ 的简写，而 f 是公式（7）中定义的概率质量函数。由此，DPk-匿名机制的隐私保证得以确立。 □

定理 7 的证明。 注意，对于任意 $i < j$,

$$\begin{aligned} & \mathcal{B}(\mathcal{M}_{\mathrm{CS}})_i - \mathcal{B}(\mathcal{M}_{\mathrm{CS}})_j = \left(\frac{k}{2} - x_i\right) p_i - \left(\frac{k}{2} - x_j\right) p_j \\ = \; & \left[\left(\frac{k}{2} - x_i\right) p_i - \left(\frac{k}{2} - x_j\right) p_i\right] + \\ & \left[\left(\frac{k}{2} - x_j\right) p_i - \left(\frac{k}{2} - x_j\right) p_j\right] \\ = \; & (x_j - x_i)\, p_i + \left(\frac{k}{2} - x_j\right)(p_i - p_j) \end{aligned} \tag{14}$$

因此，公平性误差 α_{CS} 可以计算为

$$\begin{aligned} & \|\boldsymbol{\mathcal{B}}(\mathcal{M}_{\mathrm{CS}})\|_{\rightleftharpoons} = \max_{i \in [n]} \mathcal{B}(\mathcal{M}_{\mathrm{CS}})_i - \min_{j \in [n]} \mathcal{B}(\mathcal{M}_{\mathrm{CS}})_j \\ = \; & \max_{1 \le i < j \le n} \left| \mathcal{B}(\mathcal{M}_{\mathrm{CS}})_i - \mathcal{B}(\mathcal{M}_{\mathrm{CS}})_j \right| \\ = \; & \max_{1 \le i < j \le n} \left| (x_j - x_i)\, p_i + \left(\frac{k}{2} - x_j\right)(p_i - p_j) \right| \qquad (15) \\ \le \; & \max_{1 \le i < j \le n} |(x_j - x_i)\, p_i| + \max_{1 \le i < j \le n} \left| \left(\frac{k}{2} - x_j\right)(p_i - p_j) \right| \qquad (16) \\ = \; & (x_n - x_1) p_1 + \max_{1 \le i < j \le n} \left| \left(\frac{k}{2} - x_j\right)(p_i - p_j) \right| \qquad (17) \\ \le \; & (x_n - x_1) p_1 + (p_1 - p_n) \max_{j \in [n]} \left| \left(\frac{k}{2} - x_j\right) \right| \\ \le \; & (x_n - x_1)\, p_1 + \max\left\{ \left|\frac{k}{2} - x_1\right|, \left|\frac{k}{2} - x_n\right| \right\} (p_1 - p_n) , \end{aligned}$$

其中，公式（15）由公式（14）推导得出，公式（16）来自三角不等式。此外，公式（17）是由于直方图 $\boldsymbol{x}$ 按递增顺序排列，即 $x_1 \le \cdots \le x_n$，因此，公式（3）中的概率 $\boldsymbol{p}$ 按递减顺序排列，即 $p_1 \ge \cdots \ge p_n$。除计数全部相同这一平凡情况，即 $x_1 = \cdots = x_n$ 外，当原始直方图的最大计数恰好为阈值的一半，即 $x_n = k/2$ 时，该不等式也是紧的。 □

命题 *1* 的证明。 对于任意 $\boldsymbol{a}_i \in \mathcal{X}$，可得

$$
\begin{aligned}
&\mathcal{B}\left(\mathcal{M}_{\mathrm{SW}}\right)_i = \mathbb{E}\left[\mathcal{M}_{\mathrm{SW}}(D)_i\right] - x_i \\
=\ & \sum_{j=1}^{n} \mathbb{E}\left[x_j \cdot \mathbf{1}\left\{\tilde{\boldsymbol{a}}_j = \boldsymbol{a}_i \mid \boldsymbol{a}_i\right\}\right] - x_i \\
=\ & \mathbb{E}\left[x_i \cdot \mathbf{1}\left\{\tilde{\boldsymbol{a}}_i = \boldsymbol{a}_i \mid \boldsymbol{a}_i\right\}\right] + \\
& \sum_{j \in \mathcal{I}_i \setminus \{i\}} \mathbb{E}\left[x_j \cdot \mathbf{1}\left\{\tilde{\boldsymbol{a}}_j = \boldsymbol{a}_i \mid \boldsymbol{a}_i\right\}\right] + \\
& \sum_{j \in [n] \setminus \mathcal{I}_i} \mathbb{E}\left[x_j \cdot \mathbf{1}\left\{\tilde{\boldsymbol{a}}_j = \boldsymbol{a}_i \mid \boldsymbol{a}_i\right\}\right] - x_i \\
=\ & x_i \cdot \gamma + \sum_{j \in \mathcal{I}_i \setminus \{i\}} \left(x_j \cdot \frac{1-\gamma}{n_Q - 1}\right) + 0 - x_i \\
=\ & \frac{x_i \exp(\epsilon)}{\exp(\epsilon) + n_Q - 1} + \frac{\sum_{j \in \mathcal{I}_i} x_j - x_i}{\exp(\epsilon) + n_Q - 1} + 0 - x_i \quad (18) \\
=\ & \frac{\sum_{j \in \mathcal{I}_i} x_j - n_Q x_i}{\exp(\epsilon) + n_Q - 1},
\end{aligned}
$$

其中，公式 (18) 仅代入了公式 (4) 中定义的值 γ。因此，它成功确立了差分隐私交换机制$\mathcal{M}_{\mathrm{SW}}$偏差的数学表达式。 □

定理 *8* 的证明。首先注意，对于任意 $i \in [n]$,

$$
\sum_{j \in \mathcal{I}_i} \mathcal{B}\left(\mathcal{M}_{\mathrm{SW}}\right)_j = \sum_{j \in \mathcal{I}_i} \frac{\sum_{l \in \mathcal{I}_i} x_l - n_Q x_j}{\exp(\epsilon) + n_Q - 1} = 0\,,
$$

这意味着

$$
\max_{j \in \mathcal{I}_i} \mathcal{B}\left(\mathcal{M}_{\mathrm{SW}}\right)_j \geq 0 \geq \min_{j \in \mathcal{I}_i} \mathcal{B}\left(\mathcal{M}_{\mathrm{SW}}\right)_j\,. \quad (19)
$$

假设 g 和 h 分别是与最大偏差和最小偏差相关联的索引，即,

$$
\overline{g} = \arg\max_{l \in [n]} \mathcal{B}\left(\mathcal{M}_{\mathrm{SW}}\right)_l\,, \quad \underline{h} = \arg\min_{l \in [n]} \mathcal{B}\left(\mathcal{M}_{\mathrm{SW}}\right)_l\,.
$$

而 $\underline{g}$（或 $\overline{h}$）表示索引集 $\mathcal{I}_g$（或 $\mathcal{I}_h$）上与最小（或最大）偏差相关联的索引，即，

$$
\underline{g} = \arg\min_{l \in \mathcal{I}_{\overline{g}}} \mathcal{B}\left(\mathcal{M}_{\mathrm{SW}}\right)_l\,, \quad \overline{h} = \arg\max_{l \in \mathcal{I}_{\underline{h}}} \mathcal{B}\left(\mathcal{M}_{\mathrm{SW}}\right)_l\,.
$$

根据公式 (19)，以下不等式成立

$$
\mathcal{B}\left(\mathcal{M}_{\mathrm{SW}}\right)_{\underline{g}} \leq 0 \leq \mathcal{B}\left(\mathcal{M}_{\mathrm{SW}}\right)_{\overline{h}}\,. \quad (20)
$$

于是可得

$$
\begin{aligned}
&\left\|\boldsymbol{\mathcal{B}}\left(\mathcal{M}_{\mathrm{SW}}\right)\right\|_{\rightleftharpoons} = \mathcal{B}\left(\mathcal{M}_{\mathrm{SW}}\right)_{\overline{g}} - \mathcal{B}\left(\mathcal{M}_{\mathrm{SW}}\right)_{\underline{h}} \\
\leq\ & \left(\mathcal{B}\left(\mathcal{M}_{\mathrm{SW}}\right)_{\overline{g}} - \mathcal{B}\left(\mathcal{M}_{\mathrm{SW}}\right)_{\underline{g}}\right) + \left(\mathcal{B}\left(\mathcal{M}_{\mathrm{SW}}\right)_{\overline{h}} - \mathcal{B}\left(\mathcal{M}_{\mathrm{SW}}\right)_{\underline{h}}\right) \quad (21) \\
=\ & \frac{n_Q\left(x_{\underline{g}} - x_{\overline{g}}\right)}{\exp(\epsilon) + n_Q - 1} + \frac{n_Q\left(x_{\underline{h}} - x_{\overline{h}}\right)}{\exp(\epsilon) + n_Q - 1} \\
\leq\ & \frac{2n_Q\left(x_n - x_1\right)}{\exp(\epsilon) + n_Q - 1} \\
=\ & \frac{2n_Q \left\|\boldsymbol{x}\right\|_{\rightleftharpoons}}{\exp(\epsilon) + n_Q - 1}\,,
\end{aligned}
$$

其中，公式 (21) 直接由公式 (20) 推得。 □

定理 *9* 的证明。 根据 α-公平性的定义1，公平性违反系数 α 可计算为

$$
\begin{aligned}
\left\|\boldsymbol{\mathcal{B}}\left(\mathcal{M}_{\mathrm{Lap}}\right)\right\|_{\rightleftharpoons} &= \max_{j \in [n]} \mathcal{B}\left(\mathcal{M}_{\mathrm{Lap}}\right)_j - \min_{j \in [n]} \mathcal{B}\left(\mathcal{M}_{\mathrm{Lap}}\right)_j \\
&= \mathcal{B}\left(\mathcal{M}_{\mathrm{Lap}}\right)_1 - \mathcal{B}\left(\mathcal{M}_{\mathrm{Lap}}\right)_n \quad (22) \\
&= \frac{\exp\left(-\epsilon x_1/2\right) - \exp\left(-\epsilon x_n/2\right)}{\epsilon} \\
&\leq \frac{x_n - x_1}{\epsilon} \sup_{x \in (x_1, x_n)} \left|\frac{d \exp\left(-\epsilon x/2\right)}{dx}\right| \quad (23) \\
&= \frac{\exp\left(-\epsilon x_1/2\right)}{2}\left(x_n - x_1\right) \\
&= \frac{\exp\left(-\epsilon x_1/2\right)}{2}\left\|\boldsymbol{x}\right\|_{\rightleftharpoons}\,,
\end{aligned}
$$

其中，公式（22）源于以下事实：随着计数增加，偏差会减小，即 $\mathcal{B}(\mathcal{M}_{\mathrm{Lap}})_1 \geq \cdots \geq \mathcal{B}(\mathcal{M}_{\mathrm{Lap}})_n \geq 0$。此外，公式（23）是由中值不等式 [Clarke 和 Ledyaev, 1994] 得出的。至此，证明完成。 □

定理 *10* 的证明。 首先，注意到

$$
\begin{aligned}
\alpha_{\mathrm{CS}} &= \left(x_n - x_1\right) p_1 + \max\left\{\left|\frac{k}{2} - x_1\right|, \left|\frac{k}{2} - x_n\right|\right\}\left(p_1 - p_n\right) \\
&\geq \left(x_n - x_1\right) p_1 \\
&\geq \frac{1}{2}(x_n - x_1) \quad (24) \\
&\geq \frac{\exp\left(-\epsilon x_1/2\right)}{2}\left(x_n - x_1\right) \\
&= \alpha_{\mathrm{Lap}}\,,
\end{aligned}
$$

其中，公式（24）中的不等式源于以下事实：带有 $\eta \sim$ 拉普拉斯分布（$2/\epsilon$）的函数 $x \mapsto \Pr\left(x + \eta \leq k\right)$ 是递减的，并且 x_1 低于阈值 k，这意味着

$$
p_1 = \Pr\left(x + \eta \leq k\right) \geq \Pr\left(k + \eta \leq k\right) = \frac{1}{2}\,.
$$

此外， n_Q 是受限数据宇宙$\mathcal{X}_Q$的基数，该宇宙假定为非空，因此 n_Q 至少为 1。于是可得

$$
\begin{aligned}
\alpha_{\mathrm{SW}} &= \frac{2n_Q}{\exp(\epsilon) + n_Q - 1}(x_n - x_1) \\
&= 2\left(1 - \frac{\exp(\epsilon) - 1}{\exp(\epsilon) + n_Q - 1}\right)\left(x_n - x_1\right) \\
&\geq 2\exp\left(-\epsilon\right)\left(x_n - x_1\right) \\
&\geq 2\exp\left(-\epsilon x_1/2\right)\left(x_n - x_1\right) \quad (25) \\
&\geq \frac{\exp\left(-\epsilon x_1/2\right)}{2}\left(x_n - x_1\right) \\
&= \alpha_{\mathrm{Lap}}\,,
\end{aligned}
$$

其中，公式（25）基于 x_1 不小于 2 的假设。 □

B 传统 DA 算法

本节对本文采用的传统披露避免算法作出更正式的说明。

B.1 DP 单元格抑制

算法 1 和 2 分别给出了传统单元格抑制机制及其差分隐私版本的伪代码。

进一步来说，算法 1 描述了传统单元格抑制机制，它以直方图 $\boldsymbol{x}(D)$ 和阈值 $k \in \mathbb{Z}_+$ 作为输入，并返回 $\boldsymbol{x}(D)$ 的隐私版本 $\tilde{\boldsymbol{x}}(D)$。该算法遍历直方图中的每条记录：若 $x_i < k$，则将值为 $\lfloor k/2 \rfloor$ 的每个值 x_i 抑制；否则，发布原始值 x_i（第 1–3 行）。

算法 1单元格抑制

要求： 直方图 $\boldsymbol{x}(D)$，其计数向量为 $(x_i)^{i\in[n]}$，
阈值 $k \in \mathbb{Z}_+$。
函数 cellSuppress($\boldsymbol{x}(D)$ k):
1: **对** $i \in [n]$ **执行**
2: $\tilde{x}_i \leftarrow \begin{cases} \lfloor k/2 \rfloor \text{ 如果 } x_i < k \\ x_i \text{ 否则} \end{cases}$
3: **结束循环**
4: **返回**直方图 $\tilde{\boldsymbol{x}}(D)$，其计数为 $\tilde{\boldsymbol{x}} = (\tilde{x}_i)^{i\in[n]}$

算法2描述了单元格抑制的差分隐私版本。它以直方图 $\boldsymbol{x}(D)$、阈值 $k \in \mathbb{Z}_+$和隐私参数 $\epsilon > 0$ 作为输入，并返回原始直方图 $\boldsymbol{x}(D)$的隐私版本 $\tilde{\boldsymbol{x}}(D)$。首先，使用拉普拉斯噪声（尺度为 $2/\epsilon$）~对阈值 k 进行扰动，以获得 k （第1行）。然后，算法遍历直方图中的每条记录；如果 $x_i < k$，则抑制值为 $\lfloor k/2 \rfloor$ 的每个计数 x_i ~ ~，否则释放原始值 x_i（第2–4行）。

算法 2差分隐私单元格抑制

要求： 具有计数向量 $(x_i)^{i\in[n]}$的直方图 $\boldsymbol{x}(D)$，
阈值 $k \in \mathbb{Z}_+$，隐私参数 $\epsilon > 0$。
函数 DPCellSuppress($\boldsymbol{x}(D)$ k, ϵ):
1: $\tilde{k} \leftarrow k +$拉普拉斯$(2/\epsilon)$
2: **对于** $i \in [n]$**执行**
3: $\tilde{x}_i \leftarrow \begin{cases} \lfloor k/2 \rfloor \text{ 如果 } x_i < k \\ x_i \text{ 否则} \end{cases}$
4: **结束循环**
5: **返回**直方图 $\tilde{\boldsymbol{x}}(D)$，计数为 $(\tilde{x}_i)^{i\in[n]}$

关于单元格抑制为何不满足差分隐私。例如，存在一对相邻数据集： D 和 D'。假设 D 比 D' 多一条 $\boldsymbol{a}_n$ 记录，而 D' 比 D多一条 $\boldsymbol{a}_{n-1}$ 记录。假定属性 $\boldsymbol{a}_{n-1}$ 和 $\boldsymbol{a}_n$ 在任一数据集（或）中都是"多数-关系"在任一数据集， D 或 D'，即，

$$\boldsymbol{x}(D')_{n-1} > \boldsymbol{x}(D)_{n-1} \geq k\,,$$
$$\boldsymbol{x}(D)_n > \boldsymbol{x}(D')_n \geq k\,.$$

因此，给定输入数据集 D 和 D'，与 $\boldsymbol{a}_n$ 相关联的原始单元格抑制机制的输出仍分别为 $\boldsymbol{x}(D)_n$ 和 $\boldsymbol{x}(D')_n = \boldsymbol{x}(D)_n - 1$。这意味着，由于该机制具有确定性，因此很难从这两个相邻数据集中推导出相同的输出，这违反了差分隐私的要求。

B.2 差分隐私交换

交换是依据一种度量进行的，该度量用于量化任意两条记录之间的差异。给定特征集合 Λ，将此度量记为 d_{swap}，它定义在直方图中可能记录的取值域上

$$d_{\text{swap}}(\boldsymbol{a}_i, \boldsymbol{a}_j) \triangleq \sum_{\lambda\in\Lambda} \mathbb{1}_{\text{cat}}(\lambda)\rho(\boldsymbol{a}_i[\lambda], \boldsymbol{a}_j[\lambda]) + \mathbb{1}_{\text{num}}(\lambda)\frac{|\boldsymbol{a}_i[\lambda] - \boldsymbol{a}_j[\lambda]|}{\lambda_{\text{range}}}$$

其中，ρ 是离散度量（即 $\rho(a,b) = 0 \iff a = b$，其他情况下为 1），λ_{range} 是数值特征 λ可能取值的范围。$\mathbb{1}_{\text{分类}}$ 和$\mathbb{1}_{\text{数值}}$ 分别是直方图中分类特征集和数值特征集的特征函数。非私有/确定性交换算法的详情请参见算法 3。

算法 3 描述了传统的交换机制。该算法以包含 N 条记录的直方图 $\boldsymbol{x}(D)$、交换参数 $\gamma \in [0,1]$以及准标识符列表作为输入。算法重复执行 $\lfloor (1-\gamma)N/2 \rfloor$ 次：选取直方图中一条此前尚未交换的记录 $\boldsymbol{a}_i$，再选取与 $\boldsymbol{a}_i$ 最接近的、尚未交换的记录 $\boldsymbol{a}_s$（依据度量 d_{swap}），并交换 $\boldsymbol{a}_i$ 和 $\boldsymbol{a}_s$ 的准标识符（第 1–4 行）。

算法 3交换

要求： $\boldsymbol{x}(D)$ 大小为 N 的直方图，交换参数 $\gamma \in [0,1]$，准标识符列表 Q
函数 交换($\boldsymbol{x}(D)$ γ, Q):
1: **对于** $\lfloor (1-\gamma)N/2 \rfloor$ 次**执行**
2: 随机选择一行未交换行 $\boldsymbol{a}_i$，其来自 $\boldsymbol{x}(D)$
3: $\boldsymbol{a}_s \leftarrow$ 参数 $\min_{\substack{\boldsymbol{a}_j \in \boldsymbol{x}(D)\setminus\boldsymbol{a}_i \\ \boldsymbol{a}_j \text{ 未进行交换}}} d_{\text{swap}}(\boldsymbol{a}_j, \boldsymbol{a}_i)$
4: $\boldsymbol{a}_i[Q]$、$\boldsymbol{a}_s[Q] \leftarrow \boldsymbol{a}_s[Q]$、$\boldsymbol{a}_i[Q]$
5: **结束循环**
6: **返回**已交换的直方图 $\boldsymbol{x}(D)$.

交换的差分隐私对应机制在第 5.2 小节中进行了描述。算法 4 展示了这种交换形式。它以一个包含 N 条记录的直方图 $\boldsymbol{x}(D)$、隐私参数 $\epsilon > 0$ 以及准标识符列表 Q 作为输入，并返回一个私有/修改后的直方图 $\boldsymbol{x}(D)$。对于直方图中的每条记录 $\boldsymbol{a}_i$，该算法以概率 $\gamma \triangleq \frac{\exp(\epsilon)}{\exp(\epsilon)+n_Q-1}$ 保留该记录；否则以概率 $1-\gamma$

从 $\chi_Q \setminus \boldsymbol{a}_i[Q]$ 中均匀随机地选取一组准标识符取值，其中 χ_Q 为准标识符的数据全集，并将其分配给 $\boldsymbol{a}_i[Q]$ （第 1–3 行）。

算法 4差分隐私交换

要求： 大小为 N 的直方图 $\boldsymbol{x}(D)$，隐私参数 $\epsilon > 0$，准标识符列表 Q

函数 DPSwapping($\boldsymbol{x}(D)$ ϵ， Q):

1: **对于**行 $\boldsymbol{a}_i$，遍历 D **执行**

2: $$\boldsymbol{a}_i[Q] \leftarrow \begin{cases} \boldsymbol{a}_i[Q] & \text{w.p. } \frac{\exp(\epsilon)}{\exp(\epsilon)+n_Q-1}, \\ \text{Uniform}(\mathcal{X}_Q \setminus \boldsymbol{a}_i[Q]) & \text{否则} \end{cases}$$

3: **循环结束**

4: **返回**交换后的直方图 $\boldsymbol{x}(D)$，行 $\{\boldsymbol{a}_i\}$

B.3 差分隐私 k-匿名性

在本文中，为了对数据集进行 k-匿名化，我们采用了 Mondrian 算法（LeFevre 等 [2006]）。这是一种自顶向下的贪心算法，以数据集作为输入，并输出其 k-匿名化版本。感兴趣的读者可参阅所引用的论文，以了解该算法的详细信息。用于 Python 的匿名化软件包 anonypy 包含通过 Mondrian 算法实现 k- 匿名化的功能，本文中关于 k- 匿名化的结果均使用了该功能。

算法5 生成合成的 k-匿名化数据集 需要：数据集 D，匿名化参数 $k \in \mathbb{Z}_+$函数 produceKanonymizedDataset(D, k)：1：使用 Mondrian方法（LeFevre等， [2006]）将 D 进行 k-匿名化，以得到 $D_{k\text{-anon}}$ 。2： $\tilde{D} \leftarrow$ reconstructDataset($D_{k\text{-anon}}$)˜返回重构数据集 D。函数 reconstructDataset($D_{k\text{-anon}}$, D)：˜3：初始化一个空数据集 D，其特征集合与 D相同。4：对 $D_{k\text{-anon}}$ 中的每一行 r 执行。5：重复 r[次] 。6：为 D创建新行 $\tilde{r}$ 。7：对每个特征 λ 执行。8：如果 λ 是分类的，则。9：从列表 $r[\lambda]$ 中为 $\tilde{r}$（λ）均匀随机地赋一个值。10：否则。11：为 $\tilde{r}$（λ）赋一个来自高斯分布$\mathcal{N}$（μ， σ）的值（四舍五入到最近的非负整数），其中 $\mu \triangleq \frac{a+b}{2}$ 和 $\sigma \triangleq \frac{b-a}{4}$，且 a、b 是区间 $r[\lambda] \triangleq [a, b]$的端点。12：结束条件判断。13：结束循环。14：结束循环。15：结束循环。˜返回重构数据集 D。

—— ——

在 k-匿名化版本中，分类的属性值被归为列表，数值的属性值被归为区间，并且每一行都会被分配一个计数属性

算法6差分隐私 k-匿名

要求： 数据集 D，匿名化参数 $k \in \mathbb{Z}_+$，隐私参数 $\epsilon > 0$

函数 produceDPKanonymizedDataset(D， k， ϵ):

1: $\beta \leftarrow 1 - (\exp(-\epsilon))$

2: 通过从数据集中抽样，创建数据集 D 的子集 D' 以概率 β均匀随机地选择 D 行。

3: $\tilde{D} \leftarrow$生成匿名化数据集(D'， k) ▷ 使用算法5中的produceKanonymizedDataset

返回重构数据集 D。

该属性对应于原始数据集中该行在 k-匿名化版本中所代表的行数。

重构步骤。然而，这使得难以在同一空间中使用原始数据集分析匿名化输出。因此，有必要从 k匿名化版本重构出一个与原始数据集处于相同空间的合成数据集。在我们的实验中，我们为 k匿名性加入了一个重构步骤。

算法5描述了如何使用传统的 k-匿名算法，获得数据集的重构隐私化版本。该算法包含两个组成部分： k-匿名化，以及在原始数据集的空间中重构一个输出的隐私化数据集。它以数据集 D 和 $k \in \mathbb{Z}_+$ 为输入，并输出一个重构后的˜数据集 D。

首先，使用 Mondrian方法（LeFevre 等 [2006]）对原始数据集 D 进行 k-匿名化，以获得 $D_{k\text{-anon}}$（第1行）。由于 $D_{k\text{-anon}}$ 与 D不在同一空间中，算法会使用重构步骤（第2行）。

为执行重构，算法取 $D_{k\text{-anon}}$ ，并在 D 的空间中创建一个新的˜空数据集 D （第3行）。算法遍历 k-匿名化数据集中的每一行 r ，对 r[计数] 次（即，对于原始数据集中由 r 在 $D_{k\text{-anon}}$中表示的每一行，执行一次），并为 D创建一个新行$\tilde{r}$ ；对于每个特征 λ，如果 λ 是分类的，则算法从 $r[\lambda]$ 中 λ 的合并值里为 $\tilde{r}[\lambda]$均匀随机地选择一个值；如果 λ 是数值的，则从 $\mathcal{N}$（μ， σ）中随机选择一个值，并四舍五入到最接近的非负整数，其中 μ 是区间$r[\lambda]$ 的中点，而 σ 是区间 $r[\lambda]$ 长度的 $1/4$ 倍（第4至15行）。

算法6描述了上述 k-匿名化过程的差分隐私对应版本。它以数据集 D、$k \in \mathbb{Z}_+$和隐私参数 $\epsilon > 0$ 为输入，并输出一个重构的˜、 k-匿名化数据集 D。该算法计算采样概率 $\beta \triangleq 1 - (\exp(-\epsilon))$，并以概率 β 从D 中均匀随机地抽取行，以获得 D' （第1–2行）。随后，将 D' 和 k 作为输入传递给算法5，以生成重构数据集 D （第3行）。

C 扩展结果

C.1 采用的数据集和设置

数据集 我们实验研究所采用的数据是由美国国家标准与技术研究院提供的 Diverse Community Excerpts 基准数据，且

可在 GitHub 上的 SDNist 合成数据评估库中获取。这些摘录是从美国社区调查（ACS）中精选出的地理区域和特征。每个数据集又进一步划分为称为 PUMA（公共用途微观数据区）的地理区域。具体而言，我们使用了马萨诸塞州、得克萨斯州和 Outlier PUMAs 提供的数据；这三个数据集分别包含约 5 个、6 个和 20 个 PUMA 的信息。

具体而言，完整的 ACS 数据包含约 200 个特征，而 Diverse Community Excerpts 基准数据使用了约 20 个特征。在这些特征中，我们从数据集中选取了一部分用于实验，对应的特征为 ［“RACE”、“SEX”、“OWNERSHP”、“AGE”、“INCTOT”］，，分别对应个人的种族、性别、住房产权状态、年龄和年总收入。在必要时，我们分别将数值特征 AGE 和 INCTOT 离散化或二值化为若干组。例如，除非另有说明，我们将 INCTOT 二值化为一个人每年的收入是否超过 50000 美元：超过则为（1），否则为（0）。对于年龄属性，在使用时可根据实验需要将其离散化为若干组或年龄组。例如，我们的分类实验会将年龄等分为 5 个年龄组。

需要注意的是，这样处理后，所考虑的数据集切片中的所有特征现在都属于分类的特征；对于 k-匿名算法的重构步骤而言，这尤其方便：现在可以从匿名化版本中合并属性的列表里均匀随机地选择每一行的特征，而不必从以区间中点为中心的高斯分布中进行采样，因为后者存在采样到区间端点所定义区域之外的值的轻微风险。

设置 这些实验使用 Python 3.9 及以上版本编写并运行。一些涉及大量计算的任务是在配备 AMD EPYC 7452 32 核处理器（@ 1.5 GHz）和 8GB 内存的计算集群上完成的。

C.2 数据发布

此处提供了与前文针对马萨诸塞州数据集所做的分析相同的结果，即得克萨斯州数据集和 Outlier数据集通过不同机制进行数据发布的实证结果。

表 2 和表 3 分别给出了 δ、（相对于 ℓ_1 范数的）偏差以及公平性违背界限 α 的取值（适用时，阈值 k 设为 6）。

此处呈现的趋势与正文表 1 中的趋势相似：差分隐私机制（拉普拉斯和离散高斯机制）几乎总能提供比其他机制更低的 δ、偏差和 α。这再次表明，与其他传统 DA 机制相比，差分隐私方法确实能提供更好的隐私保护、更高的数据发布准确性和更好的公平性保证。

图 6 和图 7 展示了各 DA 方法及其差分隐私变体在数据发布中产生的误差（$\|\tilde{\boldsymbol{x}}-\boldsymbol{x}\|_1$）。图 8 和图 9 展示了与之对应的公平性值（$\alpha$）。

对于正文中的马萨诸塞州数据集，这里再次可以看到，对于得克萨斯州和Outlier数据集，随着 ϵ 增加，每种披露避免机制的差分隐私对应机制在误差和公平性违背方面都趋近于原始披露避免机制。这进一步印证了这一观察：这些满足差分隐私的机制在概念上相似，表现也相近。

ϵ	机制	δ	偏差（ℓ_1 norm）	α-公平性
0.5	拉普拉斯	**0**	**901.21**	**3.645**
	离散高斯机制	0.363	1156.63	4.62
	差分隐私抑制	0.999	1138.62	4.53
	差分隐私交换	0.868	12988.58	437.105
	差分隐私 k-匿名	0.878	2963.3	24.7
1	拉普拉斯	**0**	**409.03**	**1.815**
	离散高斯机制	0.132	777.455	2.96
	差分隐私抑制	0.999	1205.315	4.47
	差分隐私交换	0.874	11624.64	394.28
	差分隐私 k-匿名	0.906	4296.8	35.4
2	拉普拉斯	**0**	**187.59**	**0.905**
	离散高斯机制	0.017	523.825	1.995
	差分隐私抑制	0.999	1219.19	4.71
	差分隐私交换	0.899	8212.7	266.78
	差分隐私 k-匿名性	0.981	5406.7	43.4
4	拉普拉斯	**0**	**81.63**	**0.46**
	离散高斯机制	3E-4	353.99	1.545
	差分隐私抑制	0.999	1217.115	4.91
	差分隐私交换	0.969	2117.62	75.48
	差分隐私 k-匿名性	0.999	5992.0	48.9

表 2：TX 数据集数据发布：从经验偏差的 δ、ℓ_1 范数和 α-公平性方面比较差分隐私机制。

ϵ	机制	δ	偏差（ℓ_1 范数）	α-公平性
0.5	拉普拉斯	**0**	**2992.88**	**4.02**
	离散高斯机制	0.363	3798.96	4.885
	差分隐私抑制	0.999	3687.445	4.39
	差分隐私交换	0.868	35385.63	580.015
	差分隐私 k-匿名	0.878	10260.6	34.5
1	拉普拉斯	**0**	**1372.36**	**2**
	离散高斯机制	0.132	2570.38	3.235
	差分隐私抑制	0.999	3911.82	4.56
	差分隐私交换	0.874	32134.71	553.66
	差分隐私 k-匿名	0.906	14668.5	46.1
2	拉普拉斯	**0**	**628.335**	**0.975**
	离散高斯机制	0.017	1728.1	2.62
	差分隐私抑制	0.999	3969.165	4.76
	差分隐私交换	0.899	22242.12	339.09
	差分隐私 k-匿名性	0.981	18529.7	54.4
4	拉普拉斯	**0**	**274.36**	**0.48**
	离散高斯机制	3E-4	1162	1.71
	差分隐私抑制	0.999	3962.755	4.93
	差分隐私交换	0.969	5634.95	108.565
	差分隐私 k-匿名性	0.999	20456.9	61.5

表 3：Outlier数据集的数据发布：比较差分隐私机制在经验偏差的 δ, ℓ_1 范数和 α-公平性方面的表现。

C.3 分类

本小节展示了论文中所述逻辑回归任务在得克萨斯州数据集和Outlier数据集上采用多种披露避免方法时的准确率图（图 5）。对于这些

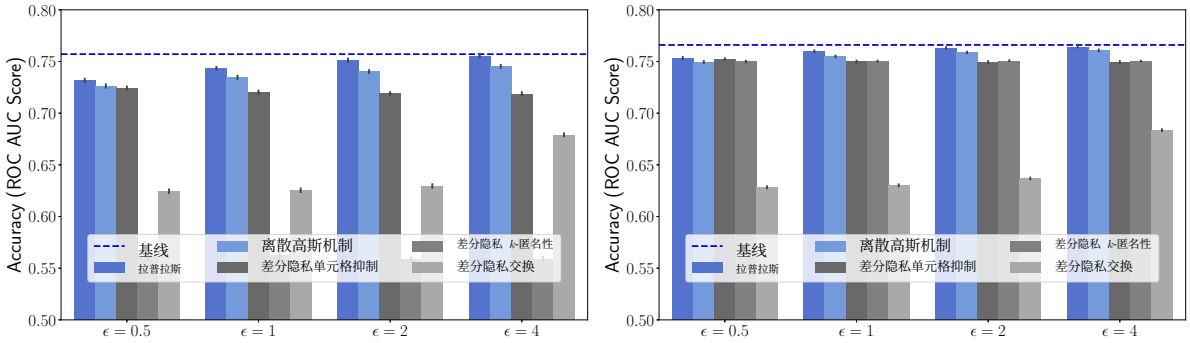


图 5：结果　　逻辑回归在得克萨斯州（左）和 Outlier　　（右）数据集

数据集，同样可以看出，使用差分隐私机制（拉普拉斯机制和离散高斯机制）生成的数据训练逻辑回归分类器，其分类准确率接近基线。此外还可以看出，与任何一种差分隐私机制相比，使用传统数据匿名化机制生成的数据所取得的准确率更低，且与基线准确率相差更远。

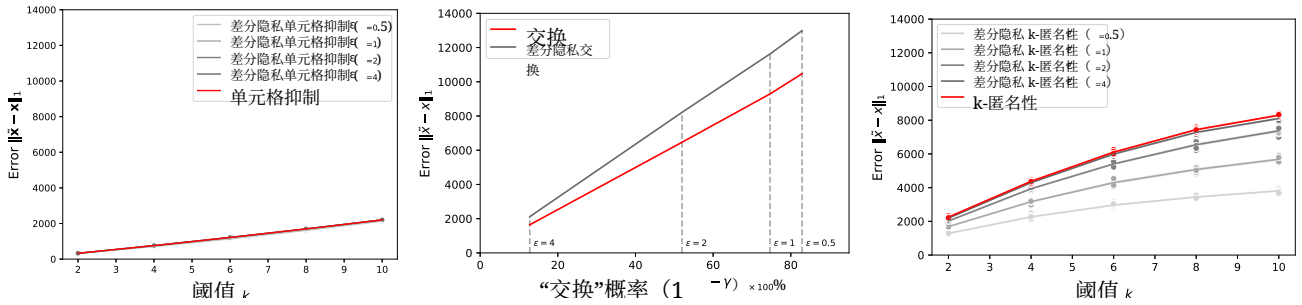


图6：德克萨斯州ACS数据集：单元格抑制（左）、交换（中）和 k-匿名性（右）的误差 $\|\tilde{\boldsymbol{x}}-\boldsymbol{x}\|_1$，以及它们的差分隐私对应方法（200次重复的平均值）。

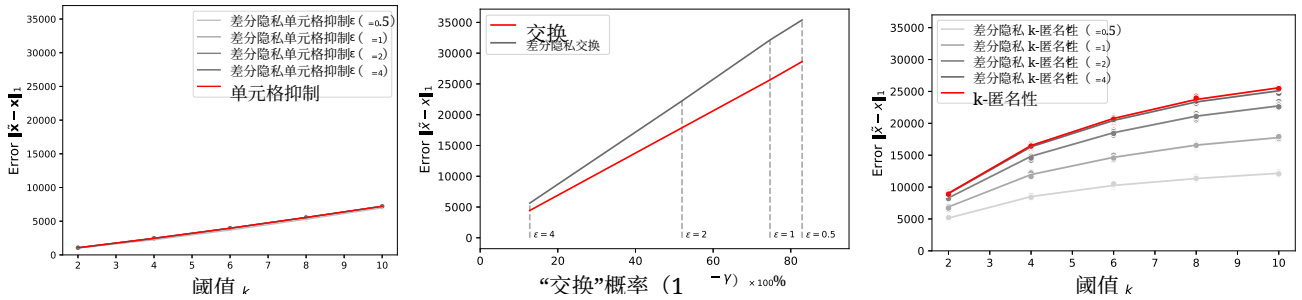


图7：异常值ACS数据集：单元格抑制（左）、交换（中）和 k-匿名性（右）的误差 $\|\tilde{\boldsymbol{x}}-\boldsymbol{x}\|_1$，以及它们的差分隐私对应方法（200次重复的平均值）。

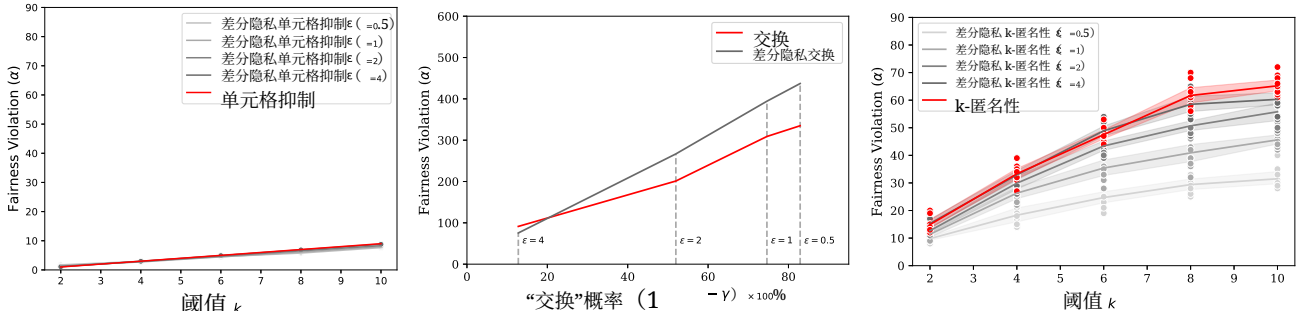


图8：德克萨斯州ACS数据集：公平性值 用于单元格抑制（左）、交换（中）和 k-匿名化（右）及其差分隐私 lly 对应方法（平均进行2 00次重复）。

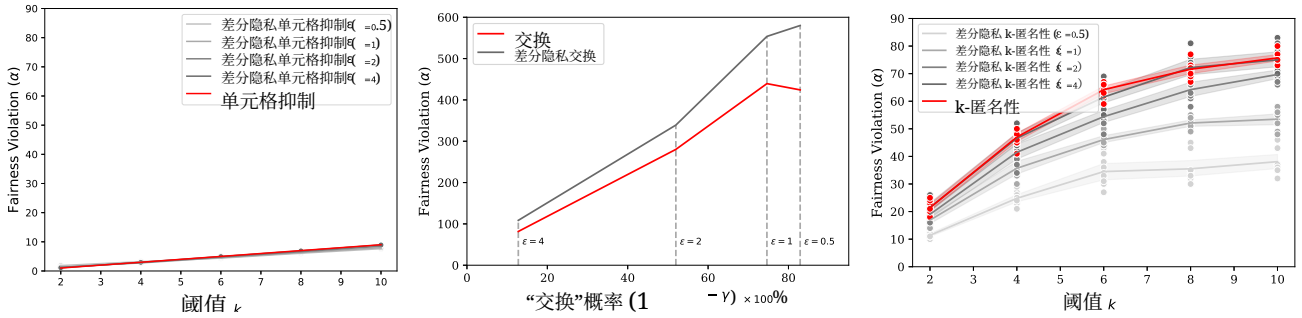


图9：异常值ACS数据集：用于单元格抑制（左）、交换（中）和 k-匿名化（右）的公平性值 α 及其差分隐私对应方法（平均进行200次重复）。