Privacy and Bias Analysis of Disclosure Avoidance Systems

Keyu Zhu[1] , **Ferdinando Fioretto**[2] , **Pascal Van Hentenryck**[1] , **Saswat Das**[3] and **Christine Task**[4]
[1]Georgia Institute of Technology
[2]Syracuse University
[3]National Institute of Science Education and Research
[4]Knexus Research Corporation
keyu.zhu@gatech.edu, ffiorett@syr.edu, pvh@isye.gatech.edu, saswat.das@niser.ac.in,
christine.task@knexusresearch.com

Abstract

Disclosure avoidance (DA) systems are used to safeguard the confidentiality of data while allowing it to be analyzed and disseminated for analytic purposes. These methods, e.g., cell suppression, swapping, and k-anonymity, are commonly applied and may have significant societal and economic implications. However, a formal analysis of their privacy and bias guarantees has been lacking. This paper presents a framework that addresses this gap: it proposes differentially private versions of these mechanisms and derives their privacy bounds. In addition, the paper compares their performance with traditional differential privacy mechanisms in terms of accuracy and fairness on US Census data release and classification tasks. The results show that, contrary to popular beliefs, traditional differential privacy techniques may be superior in terms of accuracy and fairness to differential private counterparts of widely used DA mechanisms.

1 Introduction

Disclosure avoidance (DA) systems are methods used to protect confidentiality while still enabling data analyses and dissemination. These techniques are used in various fields, such as economics, public health, social science, and data science, and have a long history in censuses and other data collection efforts. For example, the US Census Bureau has leveraged various traditional DA techniques from the 1930 decennial release on. These include suppressing certain tables based on the number of people or households in a given area and swapping data in records with similar characteristics.

While traditional confidentiality measures, such as suppression [Kelly *et al.*, 1992], swapping [Dalenius and Reiss, 1982], and k-anonymity [Sweeney, 2002] are important for protecting against accidental or intentional disclosure, they lack formal guarantees that quantify the privacy risks that individuals incur upon data releases. This is important as it restricts the ability of participants to assess the impact of these protections on published data.

In contrast, differential privacy (DP) [Dwork *et al.*, 2006] is a *relatively newer* DA that provides a rigorous definition of privacy and allows for quantifiable privacy guarantees. In differential privacy, the privacy of an individual is preserved by adding noise to their data in a controlled way. Such a process ensures that the participation of an individual in a dataset does not significantly affect the results of subsequent queries. Marking a significant shift towards more rigorous privacy protections, the US Census has recently adopted differential privacy for the 2020 Census release. However, it is worth noting that many other data agencies and organizations still rely on traditional disclosure avoidance systems to protect the confidentiality of their data.

While these approaches can be effective at protecting against accidental or intentional disclosures, it is unclear what privacy guarantees they provide when compared to differential privacy. On the other hand, while differential privacy can provide stronger privacy guarantees than traditional disclosure avoidance systems, it may come with a cost in terms of accuracy and fairness [Kuppam *et al.*, 2019; Tran *et al.*, 2021; Fioretto *et al.*, 2022], a topic of considerable debate recently.

Given that these DA are used to release data products that inform decisions with significant societal and economic consequences, it is essential to conduct a rigorous comparison of traditional DA and differential privacy in terms of privacy, bias, and fairness. However, one of the challenges faced in this comparison is the absence of a standardized framework for evaluating privacy protections. Differential privacy offers a rigorous definition of privacy and enables quantifiable privacy guarantees. On the other hand, traditional disclosure avoidance techniques may not have a distinct set of privacy metrics, making it challenging to directly compare the level of privacy protection they offer.

Contributions This paper aims at addressing this challenge: it proposes a framework for comparing traditional DA to differential privacy and makes four distinct contributions. (1) It first proposes *carefully randomized* versions of three widely adopted *traditional* DA: suppression, swapping, and k-anonymity. The resulting randomized mechanisms can then be analyzed rigorously. In particular, the paper derives (ϵ, δ)-differential privacy bounds for these new mechanisms and demonstrates that they are close to their traditional counterparts in terms of accuracy. (2) The paper then derives bounds for the bias of the new DA mechanisms, allowing for a direct comparison with classical differential privacy techniques for which such bounds exist. (3) Next, the paper analyzes the fairness impact induced by the considered DA systems and

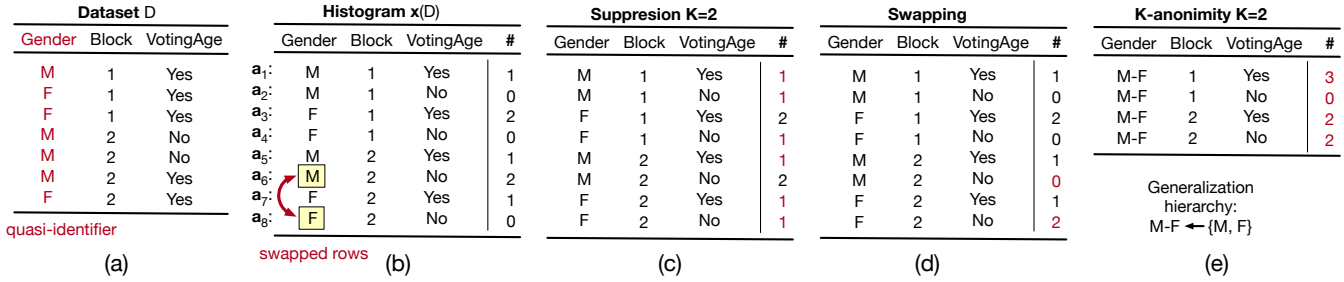

Figure 1: Illustration of the various traditional DA mechanisms.

shows that the fairness violations incurred by the randomized DA algorithms are close to those of their traditional counterparts. (4) Finally, it provides an extensive empirical analysis of the performance of the new DA mechanisms and a comparison with two classical differentially private algorithms on data release and classification tasks.

From a broader perspective, the paper demonstrates that, contrary to popular belief, classical differential privacy mechanisms may be superior to traditional disclosure avoidance systems in important data release and learning tasks in terms of accuracy and fairness for the same privacy levels. As a consequence, the results of this study have the potential to impact the way in which data agencies and organizations approach disclosure avoidance: indeed, it provides the first framework for comparing the relative strengths and limitations of traditional DA and differential privacy.

2 Problem Setting

The paper considers datasets of m records with d attributes, $A_1, \ldots, A_d$. Each record is a d-dimensional tuple of attributes associated with a unique individual from a data universe $\mathcal{X} := \prod_{i=1}^{d} \mathrm{dom}\,(A_i)$, where $\mathrm{dom}\,(A)$ represents the collection of all the possible values for the attribute A. For convenience, assume that $\mathcal{X} = \{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n\}$, with n being the size of the data universe $\mathcal{X}$, and consider the *histogram* $\boldsymbol{x}(D) \in \mathbb{N}_+^n$ of dataset D, whose i^{th} entry $x_i(D)$ represents the count of the individual records with the combination of attributes $\boldsymbol{a}_i$. When there is no ambiguity, the dataset D is omitted in the expression $\boldsymbol{x}(D)$ for simplicity. Additionally, and without loss of generality, the histogram $\boldsymbol{x}$ is assumed to be sorted in some increasing order, i.e., $x_i \leq x_j$, for any $i < j$. Finally, each entry of the histogram $\boldsymbol{x}$ is assumed to be bounded by a value $B > 0$, i.e., $\boldsymbol{x} \in [B]^n$.

Consider, for example, the illustration in Figure 1(a); The dataset D contains records with three attributes: (geographic) "Block", "Gender", and "Voting Age". The associated histogram is illustrated in Figure 1(b). In this instance, the attribute "Gender", when combined with external information like "Zip code", can become personally identifying information and thus is known as a *quasi-identifier* (QI) while the remaining attributes are referred to as *non-quasi-identifiers*. Throughout the paper, the sets of quasi-identifiers and non-quasi-identifiers are denoted by Q and N, respectively. Given a record $\boldsymbol{a}$ and a set S of attributes, $\boldsymbol{a}[S]$ is the vector of values for attributes S in $\boldsymbol{a}$.

The goal of the paper is to analyze the privacy, utility, and fairness properties of traditional disclosure avoidance systems (reviewed in the next section) on the task of releasing a privacy-preserving version $\tilde{\boldsymbol{x}}(D)$ of the histogram $\boldsymbol{x}(D)$. The notion of privacy considered in this paper is that of differential privacy, which is reviewed in the next section. The notions of utility and fairness central to the analysis rely on the concept of (statistical) *bias*. For any entry $i \in [n]$, the bias associated with a mechanism $\mathcal{M}$ is

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

where the expectation is taken over the randomness of the mechanism, and $\boldsymbol{\mathcal{B}}\left(\mathcal{M}\right) = [\mathcal{B}\left(\mathcal{M}\right)_1 \quad \ldots \quad \mathcal{B}\left(\mathcal{M}\right)_n]$. Fairness is defined as the maximal difference in biases across the histogram entries.

Definition 1 (α-fairness [Zhu *et al.*, 2022])**.** *A mechanism $\mathcal{M}$ is said to be α-fair if the maximum difference among the biases is bounded by α, i.e.,*

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

3 DA for Private Data Release

This section provides an overview of the prevalent DA methods utilized by data agencies to safeguard sensitive information within datasets. To comply with space limitations, the paper reports the proofs of all theorems in the appendix.

Differential Privacy. Differential privacy (DP) [Dwork *et al.*, 2006] is a strong privacy notion used to quantify and bound the privacy loss of an individual participation to a computation. Informally, it states that the probability of any output does not change much when a record is changed from a dataset, limiting the amount of information that the output reveals about any individual. The action of changing a record from a dataset D, resulting in a new dataset D', defines the notion of *adjacency*, denoted $D \sim D'$.

Definition 2. *A mechanism $\mathcal{M} : \mathcal{D} \to \mathcal{R}$ with domain $\mathcal{D}$ and range $\mathcal{R}$ is (ϵ, δ)-differentially private, if, for any two inputs $D \sim D' \in \mathcal{D}$, and any subset of output responses $R \subseteq \mathcal{R}$:*

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

Parameter $\epsilon > 0$ describes the *privacy loss* of the algorithm, with values close to 0 denoting strong privacy, while parameter $\delta \in [0, 1)$ captures the probability of failure of the algorithm to satisfy ϵ-DP. In particular, the *Laplace mechanism* for histogram data release, defined by $\mathcal{M}_{\text{Lap}}(\boldsymbol{x}) = \boldsymbol{x} +$

$\text{Lap}(2/\epsilon)$, where $\text{Lap}(\eta)$ is the Laplace distribution centered at 0 and with scaling factor η, satisfies $(\epsilon, 0)$-DP. Additionally, the *discrete Gaussian mechanism* [Canonne *et al.*, 2020], defined by $\mathcal{M}_{\text{Gaus}}(\boldsymbol{x}) = \boldsymbol{x} + \mathcal{N}_{\mathbb{Z}}(0, 4/\epsilon^2)$, where $\mathcal{N}_{\mathbb{Z}}(0, \sigma)$ is the discrete Gaussian distribution with 0 mean and standard deviation σ, satisfies $(\frac{1}{2}\epsilon^2 + \epsilon\sqrt{2\log(1/\delta)}, \delta)$-DP.

We next discuss three predominant traditional DA systems which, in contrast to differential privacy, do not provide formal bounds on privacy leakage.

Cell suppression. The cell suppression technique [Kelly *et al.*, 1992], frequently employed by statistical agencies (e.g., [Tatauranga Aotearoa, 2020]), aims at concealing the low-frequency counts in histograms before data dissemination.

Definition 3. *Given a histogram $\boldsymbol{x}$ and a threshold value k, cell suppression returns a private histogram $\tilde{\boldsymbol{x}}$ with entries*

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

Figure 1(c) illustrates the application of cell suppression with threshold value $k = 2$ to the histogram of Figure 1(b). The affected row counts are highlighted in red. A significant limitation of this approach is that it only protects sensitive attributes with a low number of records while neglecting others.

Swapping. Swapping [Dalenius and Reiss, 1982] is a mechanism that swaps the values of a set of sensitive attributes (the quasi-identifiers) in a record with those of another record. Informally speaking, the basic steps of the algorithms can be summarized as follows:

1. Select multiple pairs of records in the histogram with probability proportional to their discrepancies;
2. Swap the values of the quasi-identifiers attributes for each selected pair of records.

Like cell suppression, swapping produces a privacy-preserving histogram $\tilde{\boldsymbol{x}}$. However, contrary to cell suppression (and differential privacy mechanisms), swapping requires a piece of additional information: the quasi-identifier attributes of the dataset. Figure 1(d) illustrates the application of swapping where two rows are swapped, using "Gender" as the quasi-identifier attribute (see figure (a)). The affected row counts are highlighted in red. While swapping has been commonly used, for example by the US Census Bureau, to swap similar individuals within close geographies, it is not immune to reconstruction attacks [Garfinkel *et al.*, 2019].

k-anonymity. Next, k-anonymity protects sensitive data in a dataset by ensuring that each record in the dataset is indistinguishable from at least $k - 1$ other records.

Definition 4 (k-Anonymity [Sweeney, 2002])**.** *A dataset satisfies k-anonymity, relative to a set of the quasi-identifiers, if and only if when the dataset is projected to include only quasi-identifiers, every record appears at least k times.*

The basic idea behind k-anonymity is to generalize certain identifying attributes of individuals in the dataset such that each group of individuals with similar characteristics contains at least k individuals. An outline of the algorithm is provided below (a formal description is given in Appendix B):

1. define a *hierarchy* H for each quasi-identifier;
2. constructs a histogram that lists the number of records for each combination of quasi-identifiers;
3. suppress the combinations in the generalization histogram that have fewer than k instances.
4. release the resulting histogram $\tilde{\boldsymbol{x}}(D, H)$.

An important observation is that, contrary to the previous methods reviewed, k-anonymity produces a privacy-preserving histogram $\tilde{\boldsymbol{x}}(D, H)$ in a different space than $\mathcal{X}$. This important observation will be relevant in the error analysis. It additionally requires access to quasi-identifier attributes as well as a generalization histogram.

Figure 1(e) illustrates the application of k-anonymity with $k = 2$ to the histogram of Figure 1(b), using a generalization hierarchy grouping Males and Females into a single attribute. Despite being widely adopted to publish statistics and medical data, k-anonymity does not prevent re-identification attacks that exploit external public data [Li *et al.*, 2011].

4 DA Analysis Roadmap

This section outlines the methodology followed in the rest of the paper. Section 5 presents DP counterparts to traditional DA systems, including cell suppression, swapping, and k-anonymity. It aims to show that these DP counterparts preserve the main characteristics of the original mechanisms and provide an analysis of their privacy and errors under a unified privacy setting of histogram data release $\tilde{\boldsymbol{x}}(D)$. It is important to note that classical DP algorithms (e.g., Laplace mechanism) and cell suppression, make no assumptions about data attributes. In contrast, swapping relies on the use of quasi-identifiers, and k-anonymity further requires a generalization hierarchy This hierarchy forces k-anonymity to produce a histogram $\tilde{\boldsymbol{x}}(D, H)$ in a different space than that of $\tilde{\boldsymbol{x}}(D)$. While this does not affect the privacy analysis, which allows for a meaningful comparison across all mechanisms, it challenges the evaluation of the performance of these techniques. The paper addresses this challenge by also presenting a unified empirical framework for comparing the errors and biases of the various techniques in terms of the original data space $\mathcal{X}$. This necessitates a reconstruction step for k-anonymity, which is outlined in Appendix B.3. It is important to recognize that, while the DP DA mechanisms share many characteristics with their traditional DA counterparts, *they should not be considered as "noisy" versions of them*. As a result, the analytical and experimental results presented may not necessarily show a decrease in error as the privacy budget increases. In fact, they may even be more precise than the traditional mechanisms for some privacy budgets.

Next, we present the DP versions of the traditional DA systems and their privacy analyses. These analyses specify the value of the δ parameter for a given value of ϵ. Section 6 analyzes the fairness results. Finally, Section 7 presents an experimental evaluation on an extract of the American Community Survey (ACS) data [NIST, 2021].

5 Privacy and Errors Analysis

This section presents the first main contribution of the paper. It introduces differentially private counterparts to the DA presented earlier and analyzes their privacy guarantees and errors. The section starts with a technical lemma that specifies a sufficient condition for (ϵ, δ)-DP. The lemma is a critical tool

to derive the privacy guarantees of the randomized versions of the DA discussed next.

Lemma 1. *Let D, D' be datasets such that $D \sim D'$, let S be defined as*

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

and let $S^{\complement}$ denote the complement set of S. If

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

then mechanism $\mathcal{M}$ is (ϵ, δ)-differentially private.

5.1 Differentially Private Cell Suppression

While cell suppression protects the privacy of the minorities of the dataset, it neglects the privacy protection of the majorities and thus does not satisfy the requirements of differential privacy. Indeed, the deterministic nature of this mechanism prevents it from generating different outputs for two neighboring datasets. An extended discussion is deferred to Appendix B. To address this issue, the paper introduces a randomized version of cell suppression, referred to as *DP cell suppression*. This mechanism, denoted by $\mathcal{M}_{\text{CS}}$, releases a private count for every $i \in [n]$ as follows:

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

where $\eta_i \sim \text{Lap}\left(2/\epsilon\right)$ is an additive noise variable drawn from a 0-centered Laplace distribution with factor $2/\epsilon$ and k is the cell suppression threshold.

Mechanism $\mathcal{M}_{\text{CS}}$ has similarities with the *Sparse Vector Technique* (SVT) [Dwork *et al.*, 2014] which, given a sequence of queries and a real-valued threshold, outputs a vector indicating whether each (noisy) query answer is above or below the corresponding (noisy) threshold. However, there are three fundamental differences: (1) $\mathcal{M}_{\text{CS}}$, does not perturb the threshold value k; (2) it generates numeric outputs in contrast to binary outputs of *SVT*; and (3) it reports true counts rather than noisy counts, as long as the noisy counts are above the threshold (first condition of Equation (2)).

Figure 2(left) reports the empirical errors of $\mathcal{M}_{\text{CS}}$ for several threshold values k (x-axis) and ϵ parameters. The errors are given for the ACS Massachusetts dataset [NIST, 2021] (described in details in Appendix C.1): they report the ℓ_1 distances $\|\tilde{\boldsymbol{x}} - \boldsymbol{x}\|_1$ between the histograms of the cell suppression and its DP counterpart. Notice how close the errors incurred by $\mathcal{M}_{\text{CS}}$ are with respect to the original mechanism. This is important as it enables a meaningful comparison of $\mathcal{M}_{\text{CS}}$ and other DP mechanisms, since $\mathcal{M}_{\text{CS}}$ has a similar bias as the traditional cell suppression that is currently widely adopted by statistical agencies and organizations.

Privacy Analysis. The next theorem reports the privacy guarantee provided by $\mathcal{M}_{\text{CS}}$.

Theorem 1. *Given a value $\epsilon > 0$ and a threshold $k < B$, mechanism $\mathcal{M}_{CS}$ is (ϵ, δ)-differentially private with*

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

where B is a bound on the histogram entries.

Error Analysis. Having examined privacy, the paper shows how close the histograms $\tilde{\boldsymbol{x}}$ returned by $\mathcal{M}_{\text{CS}}$ are to the original histogram $\boldsymbol{x}$. The error analysis focuses on the statistical bias which, for each entry $i \in [n]$, can be expressed as

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

Observe that the error merely takes place when the noisy count is below the threshold k and is quantified as the difference between half of the threshold and the true count. Therefore, the following theorem relates the errors associated with $\mathcal{M}_{\text{CS}}$ with the probabilities of noisy counts being below the threshold, and the differences between half of the threshold and the counts of the original histogram.

Theorem 2. *The statistical bias of the DP cell suppression mechanism $\mathcal{M}_{CS}$ can be bounded as follows,*

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

where $\boldsymbol{p}$ is a shorthand for the vector

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

5.2 Differentially Private Swapping

Despite its randomized nature, the swapping mechanism fails to meet the requirements of differential privacy. To illustrate its failure, let us take a look at an instance of two neighboring datasets D and D'. Suppose that D' has a record, say $\boldsymbol{a}_1$, which does not match any record in D for any attribute $A \in Q$. No matter how swapping is performed, it cannot generate a record $\boldsymbol{a}_1$ from the input dataset D.

To obtain a DP counterpart to swapping, *it is thus critical to reason about the universe of quasi-identifiers, not simply the set of quasi-identifiers present in the database.* Let $\mathcal{X}_Q = \{\boldsymbol{q}_1, \ldots, \boldsymbol{q}_{n_Q}\}$ denote the data universe of quasi-identifiers. Instead of swapping quasi-identifiers, the mechanism will randomly choose some quasi-identifiers from $\mathcal{X}_Q$. The mechanism, referred to as DP swapping and denoted by $\mathcal{M}_{\text{SW}}$, works as follows: for every $\boldsymbol{a}_i \in \mathcal{X}$, consider the pair $(\boldsymbol{a}_i, x_i)$ denoting the tuple and its associated count in the histogram $\boldsymbol{x}$. $\mathcal{M}_{\text{SW}}$ defines $\tilde{\boldsymbol{a}}_i[N] = \boldsymbol{a}_i[N]$ and

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

where $\text{Uniform}(C)$ denotes the uniform probability over the event space C. The result of the step above may create multiple entries $(\tilde{\boldsymbol{a}}_i, x_i)$ and $(\tilde{\boldsymbol{a}}_j, x_j)$ with $\tilde{\boldsymbol{a}}_i = \tilde{\boldsymbol{a}}_j$. The procedure collapses all such tuples by summing the various x_i and x_j. The induced sub-histogram is then extended to a histogram $\tilde{\boldsymbol{x}}$. Notice that $\mathcal{M}_{\text{SW}}$ only modifies quasi-identifiers and produces a private histogram $\tilde{\boldsymbol{x}}(\tilde{D})$, similarly to what done by the original swapping algorithm.

Figure 2 (center) compares the ℓ_1 distances $\|\tilde{\boldsymbol{x}} - \boldsymbol{x}\|_1$ between the histograms generated by $\mathcal{M}_{\text{SW}}$ and its traditional counterpart for various amounts of rows swapped (in %) and parameters ϵ. Once again, observe how close the errors of the two mechanisms are.

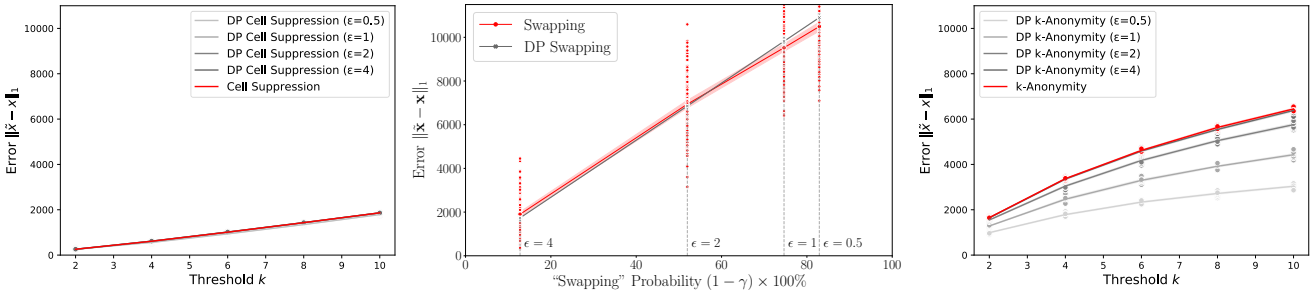

Figure 2: MA ACS dataset: Errors $\|\tilde{\boldsymbol{x}} - \boldsymbol{x}\|_1$ for cell suppression (left), swapping (center) and k-anonymity (right) and their differentially private counterparts (average of 200 repetitions).

Privacy analysis. Recall that differential privacy protects the disclosure of any individual user participating in the dataset. On the other hand, DP swapping operates at the level of a histogram count and it does so in way which impedes an analysis relying on pure $(\epsilon, 0)$-DP. The privacy analysis of $\mathcal{M}_{\text{SW}}$ is reported in the following theorem.

Theorem 3. *For a given $\epsilon > 0$, the $\mathcal{M}_{SW}$ algorithm is (ϵ, δ)-DP with δ given by*

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

with γ defined in Equation (4) and $n_Q = |\mathcal{X}_Q|$,

Error analysis. Next we discuss how close the errors of the histograms returned by swapping and its DP counterparts are.

Proposition 1. *The bias associated with each element $i \in [n]$ of the DP swapping histogram can be expressed as*

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

with the index set $\mathcal{I}_i$ collecting all the elements of the data universe $\mathcal{X} = \{\boldsymbol{a}_j \mid j \in [n]\}$, which share the same non-quasi-identifiers N with $\boldsymbol{a}_i$, i.e.,

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

Theorem 4. *The statistical bias of the DP swapping mechanism $\mathcal{M}_{SW}$ can be expressed as follows,*

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

where $\mathcal{I}_i$ is an index set defined in Equation (5) and $\boldsymbol{x}_{\mathcal{I}_i}$ is the reduced histogram consisting of the count x_j for any $j \in \mathcal{I}_i$. Additionally, $D_{\text{mean}}\left(\boldsymbol{x}_{\mathcal{I}_i}\right)$ is the mean absolute deviation of the histogram $\boldsymbol{x}_{\mathcal{I}_i}$, i.e.,

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

5.3 Differentially Private k-anonymity

Like cell suppression, k-anonymity is a deterministic algorithm, which cannot produce outputs satisfying DP. In recent years, however, there were several attempts to integrate k-anonymity and differential privacy. In particular, Li *et al.* [2011] proposed a mechanism that applies the generalization histogram and cell suppression with parameter k to a subsampled version of the original dataset. This paper proposes a generalization of this mechanism.

DP k-anonymity algorithm. The paper proposes a simple modification to the k-anonymity algorithm presented in the previous section. The mechanism, called DP-k-anonymity and denoted by $\mathcal{M}_{\text{KA}}$, operates in two steps:

1. Produce a subsampled version D_β of the dataset D in which each row is retained with probability $\beta \in (0, 1)$.
2. Apply the classical k-anonymity algorithm on D_β.

Contrary to [Li *et al.*, 2011], $\mathcal{M}_{\text{KA}}$ makes it possible to use a generalization hierarchy for merging cells, like in its deterministic counterpart. Figure 2 (right) reports the empirical errors of $\mathcal{M}_{\text{KA}}$ for several values k (x-axis); The values of β are implied by the choice of the privacy parameter ϵ (see Theorem 5). The figure reports the ℓ_1 distances between the histograms (in the original space $\mathcal{X}$) reconstructed from the generalized DP k-anonymized and the original histogram $\boldsymbol{x}(D)$ as well as those derived via its deterministic counterpart. A description of the reconstruction step adopted is provided in Appendix B. Once again, notice that the errors incurred by the DP and deterministic k-anonymity counterparts are very close to each other: in fact, the DP versions improve upon the deterministic mechanism for small values of ϵ as an artifact of the sampling procedure which considers fewer records.

Privacy Analysis. The next result generalizes [Li *et al.*, 2011] and reports the privacy guarantees provided by $\mathcal{M}_{\text{KA}}$.

Theorem 5. *For a given $k > 0$ and sampling probability $\beta \in (0, 1)$, $\mathcal{M}_{KA}$ satisfies (ϵ, δ)-DP for $\delta = d(k, \beta, \epsilon)$, where the function d is defined as*

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

ν is a shorthand for $\lfloor (1 - \exp(-\epsilon))w \rfloor$ and f represents the probability mass function of the binomial random variable,

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

Error Analysis. The goal of the analysis is again to show that the DP version of k-anonymity is close, in errors, to its

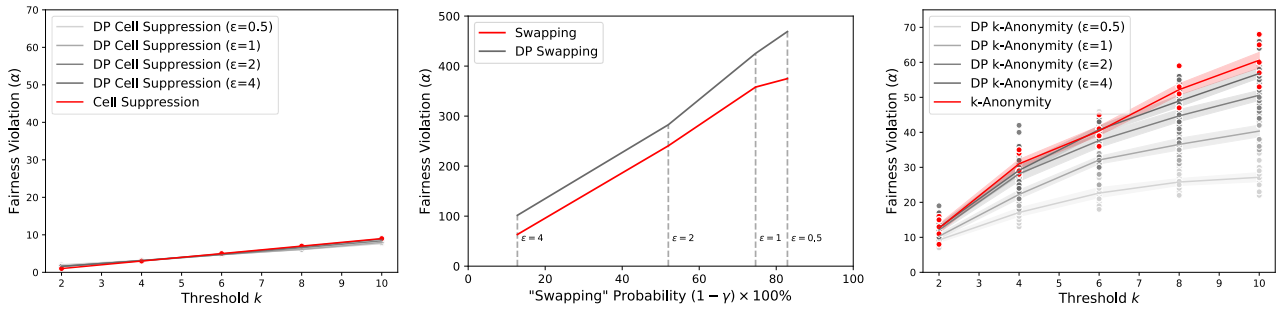

Figure 3: MA ACS dataset: Fairness values α for cell suppression (left), swapping (center) and k-anonymity (right) and their differentially private counterparts (average of 200 repetitions).

deterministic counterpart. Recall that, contrary to cell suppression and swapping, k-anonymity does not return a histogram in the same space of the attribute universe, due to its application of the generalization hierarchy. As a consequence, the output of k-anonymity consists of the counts of individual records with respect to the data universe $\mathcal{X}_H$ of the generalization hierarchy. Because of this critical difference, the rest of this section analyzes the mechanism errors by bounding whether a count of the histogram $\boldsymbol{x}(D_\beta)$ (produced in step 1) is merged by the k-anonymization step (step 2).

Let $\mathcal{M}^{\beta}_{\text{KA}}$ denote the binary vector

$$\mathcal{M}^{\beta}_{\text{KA}}(D_\beta) \coloneqq \left[\mathbf{1}\left\{x_1(D_\beta) < k\right\} \quad \dots \quad \mathbf{1}\left\{x_n(D_\beta) < k\right\}\right].$$

The error analysis focuses on statistical bias regarding whether a count would be merged by the generalization hierarchy, i.e., the difference between $\mathcal{M}^{\beta}_{\text{KA}}(D_\beta)$ and $\mathcal{M}^{\beta}_{\text{KA}}(D)$:

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

Next, we establish the equivalence between the count $x_i(D_\beta)$ and a binomial random variable $B(x_i, \beta)$ and presents the mathematical expressions characterizing the bias.

Theorem 6. *The statistical bias associated of the DP k-anonymity mechanism $\mathcal{M}^{\beta}_{KA}$ ca be expressed as, each $i \in [n]$,*

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

6 Fairness analysis

The second main contribution of this paper is an analysis of the fairness of various differentially private DA algorithms, compared to traditional differential privacy. The definition of fairness used in this paper (Definition 1) is the maximum difference in biases in the privacy-preserving histograms. It should be noted that the bias of the DP k-anonymity algorithm, which utilizes a generalization histogram, is examined in a different context than the one of the other mechanisms. Thus, the paper specifically focuses on an analytical comparison of the fairness of the DP k-anonymity algorithm.

The next result quantifies the unfairness of the DP cell suppression and swapping, along with the Laplace mechanism.

Theorem 7 (α-fairness for $\mathcal{M}_{\text{CS}}$)**.** *The DP cell suppression algorithm is α_{CS}-fair with α_{CS} given by*

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

where p_1 and p_n are the first and last entries of $\mathbf{p}$ defined in Equation (3) *respectively.*

Theorem 8 (α-fairness for $\mathcal{M}_{\text{SW}}$)**.** *The DP swapping algorithm $\mathcal{M}_{SW}$ is α_{SW}-fair with α_{SW} given by*

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

Theorem 9 (α-fairness for $\mathcal{M}_{\text{Lap}}$)**.** *The Laplace mechanism $\mathcal{M}_{Lap}$ is α_{Lap}-fair with α_{Lap} given by*

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

Figure 3 illustrates the fairness violations values, represented by the value of α, for cell suppression, swapping, and k-anonymity, as well as their differentially private counterparts, for various privacy parameters ϵ and values of k (for cell suppression and k-anonymity) or percentage of rows swapped (for swapping). It can be observed that the fairness violations of the differentially private mechanisms are comparable (or better) to those of their traditional counterparts. This is particularly noteworthy as the privacy parameter ϵ increases. As previously mentioned, it is important to remember that the differentially private mechanisms are not "noisy" versions of their traditional counterparts; rather they are conceptually similar mechanisms. Consequently, they may exhibit lower fairness violations compared to their traditional counterparts, as seen for DP k-anonymity.

The following theorem is the third key result of this paper. *It proves the superiority of the Laplace mechanism over DP cell suppression and swapping in terms of fairness errors.*

Theorem 10. *Suppose that the minimum count of the original histogram $\boldsymbol{x}(D)$ is between 2 and the threshold k, i.e., $2 \leq$*

ϵ	Mechanism	δ	Bias (ℓ_1 norm)	α-fairness
0.5	Laplace	**0**	**763.775**	**3.655**
	Discrete Gaussian	0.363	980.81	4.945
	DP Suppression	0.999	935.525	4.345
	DP Swapping	0.868	10906.79	469.015
	DP k-anonymity	0.878	2337.8	22.65
1	Laplace	**0**	**342.885**	**1.845**
	Discrete Gaussian	0.132	659.215	3.065
	DP Suppression	0.999	1003.035	4.5
	DP Swapping	0.874	9859.26	425.335
	DP k-anonymity	0.906	3297.4	32.1
2	Laplace	**0**	**154.78**	**0.905**
	Discrete Gaussian	0.017	436.925	2.2
	DP Suppression	0.999	1018.335	4.72
	DP Swapping	0.899	6841.19	282.73
	DP k-anonymity	0.981	4175.5	37.65
4	Laplace	**0**	**67.34**	**0.465**
	Discrete Gaussian	3E-4	290.715	1.495
	DP Suppression	0.999	1014.63	4.92
	DP Swapping	0.969	1664.63	101.645
	DP k-anonymity	0.999	4590.7	40.75

Table 1: MA dataset data release: Comparison of DP mechanisms in terms of δ, ℓ_1 norm of the empirical bias and α-fairness.

$x_1 \leq k$. *Then, the fairness error associated with the Laplace mechanism is not greater than that of the DP cell suppression or DP swapping mechanism, namely,*

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

It is worth noting that k-anonymity operates in a different space from the original histogram space, thus a theoretical comparison between the Laplace mechanism and DP k-anonymity is not feasible. However, the paper next presents empirical evidence that the Laplace mechanism has a significant advantage over DP k-anonymity as well.

7 Experimental Evaluation

This study assesses the performance of the DP variants of traditional DA mechanisms and compares them with two key DP mechanisms, the Laplace and the Discrete Gaussian Mechanisms, reviewed in Section 3. The experiments use the ACS 2019 IPUMS datasets for Massachusetts, Texas, and Outlier [NIST, 2021]. All the experiments report the average of 200 repetitions. Results for the latter two datasets are included in the appendix as their trends are similar to the former. The appendix also includes a more extensive description of the dataset and experimental settings. This section focuses on evaluating the mechanisms in two settings: data release and classification.

Data Release. The first task compares datasets reconstructed from histograms generated by the various DP mechanisms studied. Readers are referred to Appendix B for details on the reconstruction algorithms. Table 1 assesses the performance of the DP variants of the traditional DA mechanisms and the Laplace and the discrete Gaussian mechanisms in terms of errors and fairness violations. In mechanisms that may produce negative counts, a simple post-processing projection into the non-negative orthant is applied.

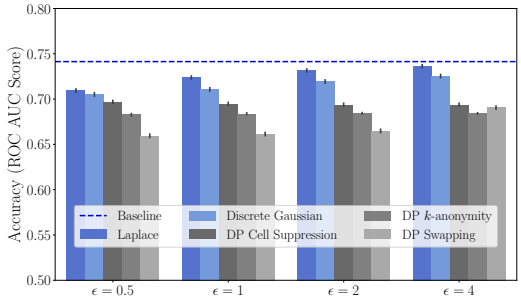

Figure 4: MA dataset: Results for Logistic Regression.

These results are particularly significant: contrary to commonly held beliefs, they demonstrate that classical DP algorithms not only provide strong privacy guarantees (see the δ values), but also produce histograms that improve over traditional DA mechanisms in terms of both accuracy (see the bias column) and fairness metrics (see α-fairness column). *As a consequence, when agencies desire to release data sets, it is advisable they consider traditional DP mechanisms.*

Classification. It is also important to compare the performance of all the DP mechanisms in a classification task. The setting employs the private datasets obtained through a data-release query in order to train a logistic regression classifier. The task is to predict whether an individual earns more than $50,000 per year, and the results in Figure 4 are presented in terms of accuracy on the original, non-private dataset. Observe how the Laplace and discrete Gaussian mechanisms lead to classifiers with much higher accuracies than classifiers trained over data produced by other traditional DA mechanisms. Notably, the classification accuracy of Laplace and discrete Gaussian is much closer to that of the baseline method (trained on non-private datasets) than any other method. Again, this is significant: *despite their simplicity, these tasks are the basis for numerous statistical analyses performed routinely by data agencies and organizations.*

8 Conclusion

This paper presented a framework for comparing traditional disclosure avoidance systems (DA) to differential privacy. It proposed carefully randomized versions of three widely adopted traditional DA methods, i.e., suppression, swapping, and k-anonymity, and derived (ϵ, δ)-DP bounds for these mechanisms. The paper also analyzed these DP algorithms empirically and showed that they are close to their traditional counterparts both in terms of accuracy and fairness. The DP DA mechanisms were then compared experimentally with traditional DP mechanisms (i.e., the Laplace or the discrete Gaussian mechanisms) on data release and classification tasks. Contrary to popular belief, the experimental evaluation showed that classical DP mechanisms may be superior to traditional DA in terms of accuracy and fairness for the same privacy levels. This study has the potential to impact the way in which data agencies and organizations approach disclosure avoidance in the future as it provides a framework that enables a comparison of the strengths and limitations of traditional DA and differential privacy.

Ethical Statement

From an ethical standpoint, the study's purpose is not to condone the release of data by agencies that have not fully considered the privacy implications of their actions. The study should not be taken as a means to discredit traditional DA methods. Furthermore, the empirical analysis presented should be understood as specific to the mechanisms and datasets discussed in the study.

It is also important to consider the potential benefits of the study, such as improved accuracy and fairness in data release which may be gained with the adoption of traditional differentially private tools. Additionally, the study has the potential to advance the development of more effective privacy-preserving technologies.

Acknowledgments

This research is partially supported by NSF grant 2133169, AI institute 2112533, and NSF CAREER Award 2143706. Fioretto is also supported by a Google Research Scholar Award and an Amazon Research Award. Its views and conclusions are those of the authors only. The authors would also like to thank Matt Williams for the helpful discussions on swapping and other traditional DA techniques.

References

Clément L Canonne, Gautam Kamath, and Thomas Steinke. The discrete gaussian for differential privacy. *Advances in Neural Information Processing Systems*, 33:15676–15688, 2020.

FH Clarke and Yu S Ledyaev. Mean value inequalities. *Proceedings of the American Mathematical Society*, pages 1075–1083, 1994.

Tore Dalenius and Steven P Reiss. Data-swapping: A technique for disclosure control. *Journal of statistical planning and inference*, 6(1):73–85, 1982.

Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In *Theory of cryptography conference*, pages 265–284. Springer, 2006.

Cynthia Dwork, Aaron Roth, et al. The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science*, 9(3–4):211–407, 2014.

Ferdinando Fioretto, Cuong Tran, Pascal Van Hentenryck, and Keyu Zhu. Differential privacy and fairness in decisions and learning tasks: A survey. In *In Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, pages 5470–5477, 2022.

Simson Garfinkel, John M Abowd, and Christian Martindale. Understanding database reconstruction attacks on public data. *Communications of the ACM*, 62(3):46–53, 2019.

James P Kelly, Bruce L Golden, and Arjang A Assad. Cell suppression: Disclosure protection for sensitive tabular data. *Networks*, 22(4):397–417, 1992.

Satya Kuppam, Ryan McKenna, David Pujol, Michael Hay, Ashwin Machanavajjhala, and Gerome Miklau. Fair decision making using privacy-protected data. *arXiv preprint arXiv:1905.12744*, 2019.

K. LeFevre, D.J. DeWitt, and R. Ramakrishnan. Mondrian multidimensional k-anonymity. In *22nd International Conference on Data Engineering (ICDE'06)*, pages 25–25, 2006.

Ninghui Li, Wahbeh H Qardaji, and Dong Su. Provably private data anonymization: Or, k-anonymity meets differential privacy. *CoRR, abs/1101.2604*, 49:55, 2011.

NIST. Sdnist v1.4 beta: Synthetic data report tool. national institute of standards and technology, 2021.

Latanya Sweeney. k-anonymity: A model for protecting privacy. *International journal of uncertainty, fuzziness and knowledge-based systems*, 10(05):557–570, 2002.

Tatauranga Aotearoa. Microdata output guide. https://www.stats.govt.nz/assets/Methods/Microdata-Output-Guide-2020-v5-Sept22update.pdf, 2020.

Cuong Tran, Ferdinando Fioretto, Pascal Van Hentenryck, and Zhiyan Yao. Decision making with differential privacy under the fairness lens. In *International Joinmogt Conference on Artificial Intelligence (IJCAI)*, pages 560–566, 2021.

Keyu Zhu, Ferdinando Fioretto, and Pascal Van Hentenryck. Post-processing of differentially private data: A fairness perspective. In Lud De Raedt, editor, *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI-22*, pages 4029–4035. International Joint Conferences on Artificial Intelligence Organization, 7 2022. Main Track.

A Missing Proofs

Proof of Lemma 1. For any output $O \subseteq \mathcal{R}$ and neighboring datasets $D, D' \in \mathcal{X}^m$,

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

□

Proof of Theorem 1. Without loss of generality, the neighboring datasets, $D, D' \in \mathcal{X}^m$, are assumed to have different last records with $\boldsymbol{a}_n$ in D and $\boldsymbol{a}_{n-1}$ in D'. It implies that the histograms of D and D' differ in the last two entries, i.e.,

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

Consider the following set

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

For any element $\boldsymbol{o} \in \underline{S}$, it follows that

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

where, in Equation (8), η_{n-1}, η'_{n-1}, η_n, and η'_n are all i.i.d. Laplacian random variables with the parameter $2/\epsilon$. Besides, Equation (9) comes from the triangle inequality. This inequality implies that the set $\underline{S}$ is a subset of S in Lemma 1. As a result, the probability $\Pr\left(\mathcal{M}_{\text{CS}}(D) \in S^{\complement}\right)$ can then be evaluated as follows

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

where Equation (10) comes from the fact that the function $x \mapsto \Pr\left(x + \eta < k\right)$ with $\eta \sim \text{Lap}\left(2/\epsilon\right)$ is decreasing and the histogram $\boldsymbol{x}(D)$ is assumed to be bounded by the constant B entrywise. Additionally, Equation (11) is due to the assumption that $k < B$. By Lemma 1, the privacy guarantee is established for the mechanism $\mathcal{M}_{\text{CS}}$, which completes the proof here.

□

Proof of Theorem 2. By the Cauchy-Schwarz inequality,

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

□

Proof of Theorem 3. Without loss of generality, the neighboring datasets, $D, D' \in \mathcal{X}^m$, are assumed to have different last records with $\boldsymbol{a}_n$ in D and $\boldsymbol{a}_{n-1}$ in D', where $\boldsymbol{a}_n[Q] \neq \boldsymbol{a}_{n-1}[Q]$ and $\boldsymbol{a}_n[N] = \boldsymbol{a}_{n-1}[N]$. Consider the following set

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

where $\tilde{\boldsymbol{x}}$ is the resulting histogram associated with $\{\tilde{\boldsymbol{a}}_i \mid i \in [n]\}$ generated by the DP swapping mechanism. It is straightforward to see that, for any $\tilde{\boldsymbol{x}} \in S$,

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

Then, it follows that

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

By Lemma 1, it provides the privacy guarantee for the DP swapping mechanism. □

Proof of Theorem 5. In order to derive the privacy guarantee for the DP k-anonymity mechanism, it suffices to show that the sample dataset D_β satisfies (ϵ, δ)-differential privacy with δ defined in Equation (6) because of the significant property, known as post-processing immunity [Dwork *et al.*, 2014].

Prior to analysis, let $\mathcal{M}_\beta$ denote the randomized mechanism which generates the sample dataset. Without loss of generality, assume that the neighboring datasets, $D, D' \in \mathcal{X}^m$, have different last records with $\boldsymbol{a}_n$ in D and $\boldsymbol{a}_{n-1}$ in D'. It implies that the histograms of D and D' differ in the last two entries, i.e.,

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

Consider the set

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

For any sample dataset $D_\beta \in S_{D,D'}$,

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

Notice that the count $x(D_\beta)_n$ is non-negative, which implies the following

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

Likewise, for the attribute $\boldsymbol{a}_{n-1}$, the following inequalities hold

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

Therefore,

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

Then, by Equation (12) and (13),

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

Thus, for any neighboring datasets D and D', $S_{D,D'}$ turns out to be a set, each element of which $\mathcal{M}_\beta(D)$ and $\mathcal{M}_\beta(D')$ generate with similar probability. By Lemma 1, the error probability δ can then be computed as

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

where ν is a shorthand for $\lfloor (1-\exp(-\epsilon))w \rfloor$ and f is a probability mass function defined in Equation (7). In this way, the privacy guarantee has been established for the DP k-anonymity mechanism. □

Proof of Theorem 7. Observe that, for any $i < j$,

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

It follows that the fairness error α_{CS} can be computed as

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

where Equation (15) is derived from Equation (14) and Equation (16) comes from the triangle inequality. Besides, Equation (17) is due to the fact that the histogram $\boldsymbol{x}$ is sorted in an increasing order, i.e., $x_1 \le \cdots \le x_n$ and, as a consequence, the probabilities $\boldsymbol{p}$ in Equation (3) appear in a decreasing order, i.e., $p_1 \ge \cdots \ge p_n$. Except for the trivial case that, the the counts are all the same, i.e., $x_1 = \cdots = x_n$, this inequality is also tight when the maximum count of the original histogram is exactly half of the threshold, i.e., $x_n = k/2$.

□

Proof of Proposition 1. For any $\boldsymbol{a}_i \in \mathcal{X}$, it follows that

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

where Equation (18) just plugs in the value γ defined in Equation (4). Thus, it manages to establish the mathematical expression of the bias of the DP swapping mechanism $\mathcal{M}_{\text{SW}}$. □

Proof of Theorem 8. In the first place, notice that, for any $i \in [n]$,

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

which implies the following

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

Suppose that $\overline{g}$ and $\underline{h}$ are the indices associated with maximum and minimum biases respectively, i.e.,

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

and $\underline{g}$ (or $\overline{h}$) represents the index associated with the minimum (or maximum) bias over the index set $\mathcal{I}_{\overline{g}}$ (or $\mathcal{I}_{\underline{h}}$), i.e.,

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

By Equation (19), the following inequalities hold

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

Then, it follows that

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

where Equation (21) is a direct consequence of Equation (20). □

Proof of Theorem 9. By Definition 1 of α-fairness, the fairness violation coefficient α can be computed as

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

where Equation (22) comes from the fact that the biases decrease, as the counts increase, i.e., $\mathcal{B}\left(\mathcal{M}_{\text{Lap}}\right)_1 \geq \cdots \geq \mathcal{B}\left(\mathcal{M}_{\text{Lap}}\right)_n \geq 0$. Besides, Equation (23) is due to the mean value inequalities [Clarke and Ledyaev, 1994]. It completes the proof here. □

Proof of Theorem 10. First of all, note that

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

where the inequality in Equation (24) comes from the fact that the function $x \mapsto \Pr\left(x + \eta \leq k\right)$ with $\eta \sim \text{Lap}\left(2/\epsilon\right)$ is decreasing and x_1 is below the threshold k, which implies the following

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

Besides, n_Q is the cardinality of the restricted data universe $\mathcal{X}_Q$, which is assumed to be non-empty and thus n_Q is at least 1. Then, it follows that

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

where Equation (25) is based on the assumption that x_1 is no less than 2. □

B Traditional DA Algorithms

This section presents more formal specifications of the traditional DA algorithms adopted in the paper.

B.1 DP Cell Suppression

Algorithms 1 and 2 provide the pseudocode for the traditional cell suppression mechanism and its differentially private counterpart, respectively.

To further elaborate, algorithm 1, which describes the traditional cell suppression mechanism, takes as input a histogram $\boldsymbol{x}(D)$ and a threshold $k \in \mathbb{Z}_+$ and returns a private version $\tilde{\boldsymbol{x}}(D)$ of $\boldsymbol{x}(D)$. The algorithm iterates through each record of the histogram and suppresses each value x_i with value $\lfloor k/2 \rfloor$ if $x_i < k$ or releases the original value x_i otherwise (lines 1–3).

Algorithm 1 Cell Suppression

Require: Histogram $\boldsymbol{x}(D)$ with vector of counts $(x_i)_{i\in[n]}$, threshold $k \in \mathbb{Z}_+$.
function cellSuppress($\boldsymbol{x}(D), k$):
1: **for** $i \in [n]$ **do**
2: $\quad \tilde{x}_i \leftarrow \begin{cases} \lfloor k/2 \rfloor \text{ if } x_i < k \\ x_i \text{ otherwise} \end{cases}$
3: **end for**
4: **return** Histogram $\tilde{\boldsymbol{x}}(D)$ with counts $\tilde{\boldsymbol{x}} = (\tilde{x}_i)_{i\in[n]}$

Algorithm 2 describes the differentially private counterpart of cell suppression. It takes as input a histogram $\boldsymbol{x}(D)$, a threshold $k \in \mathbb{Z}_+$, and a privacy parameter $\epsilon > 0$ and returns a private version $\tilde{\boldsymbol{x}}(D)$ of the original histogram $\boldsymbol{x}(D)$. First the threshold k is perturbed with Laplace noise (of scale $2/\epsilon$) to obtain $\tilde{k}$ (line 1). Then the algorithm iterates over every record of the histogram and suppresses each count x_i with value $\lfloor \tilde{k}/2 \rfloor$ if $x_i < \tilde{k}$ or releases the original value x_i otherwise (lines 2–4).

Algorithm 2 DP Cell Suppression

Require: Histogram $\boldsymbol{x}(D)$ with vector of counts $(x_i)_{i\in[n]}$, threshold $k \in \mathbb{Z}_+$, privacy parameter $\epsilon > 0$.
function DPCellSuppress($\boldsymbol{x}(D), k, \epsilon$):
1: $\tilde{k} \leftarrow k + \text{Laplace}(2/\epsilon)$
2: **for** $i \in [n]$ **do**
3: $\quad \tilde{x}_i \leftarrow \begin{cases} \lfloor \tilde{k}/2 \rfloor \text{ if } x_i < \tilde{k} \\ x_i \text{ otherwise} \end{cases}$
4: **end for**
5: **return** Histogram $\tilde{\boldsymbol{x}}(D)$ with counts $(\tilde{x}_i)_{i\in[n]}$

On why cell suppression does not satisfy differential privacy. For instance, there exists a pair of neighboring datasets, D and D'. Suppose that D has one more record of $\boldsymbol{a}_n$ than D' while D' has one more record of $\boldsymbol{a}_{n-1}$ than D. The attributes $\boldsymbol{a}_{n-1}$ and $\boldsymbol{a}_n$ are assumed to be the "majorities" in whichever dataset, D or D', i.e.,

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

Thus, given the input datasets D and D', the outputs of the original cell suppression mechanism associated with $\boldsymbol{a}_n$ are still $\boldsymbol{x}(D)_n$ and $\boldsymbol{x}(D')_n = \boldsymbol{x}(D)_n - 1$ respectively. It means that this mechanism, due to the nature that it is deterministic, can hardly derive the same output from these two neighboring datasets, which violates the requirements of differential privacy.

B.2 DP Swapping

Swapping is done with respect to a metric that quantifies the discrepancies between any two records. Given the set of features Λ, this metric, let us denote it by d_{swap}, is defined over the domain of possible records of a histogram as

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

Where ρ is the discrete metric (i.e. $\rho(a,b) = 0 \iff a = b$ and 1 otherwise) and λ_{range} is the range of the possible values taken by a numerical feature λ. $\mathbb{1}_{\text{cat}}$ and $\mathbb{1}_{\text{num}}$ are characteristic functions of the sets of categorical and numerical features of the histogram respectively. Refer to algorithm 3 for details on the non-private/deterministic swapping algorithm.

Algorithm 3 describes the traditional swapping mechanism. This takes as input a histogram $\boldsymbol{x}(D)$ with N records, a swapping parameter $\gamma \in [0,1]$, and a list of quasi-identifiers. For $\lfloor (1-\gamma)N/2 \rfloor$ times, the algorithm picks a hitherto unswapped record $\boldsymbol{a}_i$ of the histogram, picks the closest unswapped record $\boldsymbol{a}_s$ to $\boldsymbol{a}_i$ (w.r.t. the metric d_{swap}) and swaps the quasi identifiers of $\boldsymbol{a}_i$ and $\boldsymbol{a}_s$ (lines 1–4).

Algorithm 3 Swapping

Require: Histogram $\boldsymbol{x}(D)$ of size N, Swapping Parameter $\gamma \in [0,1]$, list of quasi-identifiers Q
function swapping($\boldsymbol{x}(D), \gamma, Q$):
1: **for** $\lfloor (1-\gamma)N/2 \rfloor$ times **do**
2: $\quad$ Randomly pick an unswapped row $\boldsymbol{a}_i$ of $\boldsymbol{x}(D)$
3: $\quad \boldsymbol{a}_s \leftarrow \arg \min_{\substack{\boldsymbol{a}_j \in \boldsymbol{x}(D)\setminus \boldsymbol{a}_i \\ \boldsymbol{a}_j \text{ is unswapped}}} d_{\text{swap}}(\boldsymbol{a}_j, \boldsymbol{a}_i)$
4: $\quad \boldsymbol{a}_i[Q], \boldsymbol{a}_s[Q] \leftarrow \boldsymbol{a}_s[Q], \boldsymbol{a}_i[Q]$
5: **end for**
6: **return** Swapped histogram $\boldsymbol{x}(D)$.

The differentially private counterpart of swapping was described in subsection 5.2. Algorithm 4 presents this form of swapping. It takes as input a histogram $\boldsymbol{x}(D)$ with N records, a privacy parameter $\epsilon > 0$, and a list of quasi-identifiers Q and returns a private/modified histogram $\boldsymbol{x}(D)$. For each record $\boldsymbol{a}_i$ of the histogram, the algorithm preserves it with probability $\gamma \triangleq \frac{\exp(\epsilon)}{\exp(\epsilon)+n_Q-1}$; else with probability $1-\gamma$

picks a set of values of quasi-identifiers from $\chi_Q \setminus \boldsymbol{a}_i[Q]$ uniformly at random, where χ_Q is the data universe of quasi-identifiers, and assigns it to $\boldsymbol{a}_i[Q]$ (lines 1–3).

Algorithm 4 DP Swapping

Require: Histogram $\boldsymbol{x}(D)$ of size N, privacy parameter $\epsilon > 0$, list of quasi-identifiers Q

function DPSwapping($\boldsymbol{x}(D), \epsilon, Q$) :

1: **for** row $\boldsymbol{a}_i$ in D **do**

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

3: **end for**

4: **return** Swapped histogram $\boldsymbol{x}(D)$ with rows $\{\boldsymbol{a}_i\}$

B.3 DP k-anonymity

In this paper, to k-anonymize a dataset we utilize the Mondrian algorithm (LeFevre *et al.* [2006]). This is a top-down greedy algorithm that takes a dataset as input and outputs a k-anonymized version of it. Interested readers may refer to the cited paper for details about this algorithm. `anonypy`, an anonymization package for python, includes an implementation for k-anonymity via the Mondrian algorithm, which has been used for the results on k-anonymity in this paper.

Algorithm 5 Producing Synthetic k-Anonymized Dataset

Require: Dataset D, anonymization parameter $k \in \mathbb{Z}_+$

function produceKanonymizedDataset(D, k) :

1: k-anonymize D to get $D_{k\text{-anon}}$ using the Mondrian method (LeFevre *et al.* [2006]).

2: $\tilde{D} \leftarrow$reconstructDataset($D_{k\text{-anon}}$)

return Reconstructed dataset $\tilde{D}$.

function reconstructDataset($D_{k\text{-anon}}, D$) :

3: Initialise an empty dataset $\tilde{D}$ with the same set of features as D

4: **for** every row r in $D_{k\text{-anon}}$ **do**

5: **for** r[count] many times **do**

6: Create new row $\tilde{r}$ for $\tilde{D}$

7: **for** each feature λ **do**

8: **if** λ is categorical **then**

9: Assign $\tilde{r}(\lambda)$ a value from the list $r[\lambda]$ uniformly at random.

10: **else**

11: Assign $\tilde{r}(\lambda)$ a value from the Gaussian $\mathcal{N}(\mu, \sigma)$ (rounded off to the nearest non-negative integer), where $\mu \triangleq \frac{a+b}{2}$ and $\sigma \triangleq \frac{b-a}{4}$, where a, b are the endpoints of the interval $r[\lambda] \triangleq [a, b]$.

12: **end if**

13: **end for**

14: **end for**

15: **end for**

return Reconstructed dataset $\tilde{D}$.

In the k-anonymized version, categorical attribute values are grouped together as lists and numerical ones are grouped together as intervals and each row is assigned a count attribute corresponding to how many rows of the original dataset the said row in the k-anonymized version represents.

Algorithm 6 DP k-Anonymity

Require: Dataset D, anonymization parameter $k \in \mathbb{Z}_+$, privacy parameter $\epsilon > 0$

function produceDPKanonymizedDataset(D, k, ϵ) :

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

2: Create a subset D' of the dataset D by sampling from the rows of D uniformly at random with probability β.

3: $\tilde{D} \leftarrow$produceKanonymizedDataset(D', k) ▷ Using produceKanonymizedDataset from algorithm 5

return Reconstructed dataset $\tilde{D}$.

Reconstruction step. However, this makes it difficult to analyze the anonymized output with the original dataset in the same space. Thus it is necessary to reconstruct a synthetic dataset from the k-anonymized version that is in the same space as that of the original dataset. In our experiments, we include a reconstruction step for k-anonymity.

Algorithm 5 describes how we obtain a reconstructed, privatized version of the dataset using the traditional k-anonymity algorithm. This algorithm involves two components: k-anonymization and reconstructing an output, privatized dataset in the space of the original dataset. It takes as input a dataset D and $k \in \mathbb{Z}_+$ and outputs a reconstructed dataset $\tilde{D}$.

First, the original dataset D is k-anonymized using the Mondrian method (LeFevre *et al.* [2006]) to obtain $D_{k\text{-anon}}$ (line 1). As $D_{k\text{-anon}}$ is not in the same space as D, the algorithm uses a reconstruction step (line 2).

To perform the reconstruction, $D_{k\text{-anon}}$ is taken and a new empty dataset $\tilde{D}$ in the space of D is created (line 3). The algorithm iterates over every row r in the k-anonymized dataset for r[count] times (i.e. once for every row in the original dataset that is represented by r in $D_{k\text{-anon}}$), creates a new row $\tilde{r}$ for D; for each feature λ, if λ is categorical, then the algorithm chooses one of the merged values of λ in $r[\lambda]$ uniformly at random for $\tilde{r}[\lambda]$, or if λ is numeric, then a random value is chosen from $\mathcal{N}(\mu, \sigma)$ and rounded off to the nearest non-negative integer, where μ is the midpoint of the interval $r[\lambda]$ and σ is $1/4$ times the length of the interval $r[\lambda]$ (lines 4-15).

Algorithm 6 describes the DP counterpart of the aforementioned k-anonymity process. It takes as input a dataset D, $k \in \mathbb{Z}_+$, and a privacy parameter $\epsilon > 0$ and outputs a reconstructed, k-anonymized dataset $\tilde{D}$. The algorithm computes a sampling probability $\beta \triangleq 1-(\exp(-\epsilon))$ and samples rows of D uniformly at random with probability β to obtain D' (lines 1–2). Then D' and k are passed as input to algorithm 5 to produce the reconstructed dataset $\tilde{D}$ (line 3).

C Extended Results

C.1 Datasets and Settings adopted

Datasets The data adopted in our experimental studies was the Diverse Community Excerpts Benchmark Data, provided by the National Institute of Standards and Technology and

available on the SDNist synthetic data evaluation library on GitHub. The excerpts are a curated selection of geography and features derived from the American Community Survey (ACS). Each of these datasets is further divided into geographical regions known as PUMAs (Public Use Microdata Areas). In particular, we use data provided for Massachusetts, Texas, and Outlier PUMAs; these three datasets contain information about 5, 6, and 20 PUMAs respectively.

In particular, while the full ACS data has about 200 features, the Diverse Community Excerpts Benchmark Data uses about 20 features. Out of these features, we use a slice of the dataset for our experiments corresponding to the features `['RACE', 'SEX', 'OWNERSHP', 'AGE','INCTOT']`, which correspond to the race, sex, house ownership status, age and the total annual income of an individual respectively. Wherever necessary, the numeric features `AGE` and `INCTOT` are discretized/binarized respectively into groups. For instance, unless stated otherwise, we binarize `INCTOT` into whether a person earns more than $\$\,50000$ per annum (1) or not (0). The age attribute, wherever used, may be discretized into groups/age brackets, depending upon the experiment. For example, our classification experiments involve dividing ages equitably into 5 age brackets.

Note that in doing so, all features in the dataset slice being considered are now categorical, and this is especially convenient when it comes to the reconstruction step of the k-anonymity algorithm: now all the rows' features can be chosen uniformly at random from the list of merged attributes in the anonymized version rather than sampling from a Gaussian centered around the midpoint of an interval, which carries a slight risk of sampling values outside of the region defined by the endpoints of the interval.

Settings These experiments have been coded and run using Python 3.9 and above. Some tasks involving heavy computation were performed using a cluster equipped with AMD EPYC 7452 32-Core CPUs (@ 1.5 GHz) and 8GB of RAM.

C.2 Data Release

Here empirical results on the data release via different mechanisms are provided for the Texas and Outlier datasets as done earlier for the Massachusetts dataset.

Tables 2 and 3 provide the values of δ, biases (w.r.t. the ℓ_1 norm), and the fairness violation bound α respectively (wherever applicable, the threshold k is set to be 6).

Here, a similar trend is seen as for table 1 in the main text and the DP mechanisms (Laplace and discrete Gaussian) almost always offer lower values of δ, biases, and α than the rest. This again demonstrates that DP methods do indeed offer better privacy protection, higher accuracy of data release, and better fairness guarantees than the other traditional DA mechanisms.

Figures 6 and 7 provide plots showing the errors ($\|\tilde{\boldsymbol{x}}-\boldsymbol{x}\|_1$) associated with the data release of each DA method and its DP variants. Figures 8 and 9 provide plots showing the fairness values (α) associated with the same.

As for the Massachusetts dataset in the main text, it is again seen here for the Texas and Outlier datasets that as ϵ increases, the DP counterpart of each DA mechanism approaches the original DA mechanism in terms of errors and fairness violations. This further reinforces the observation that these differentially private mechanisms are conceptually similar and perform similarly.

ϵ	Mechanism	δ	Bias (ℓ_1 norm)	α-fairness
0.5	Laplace	**0**	**901.21**	**3.645**
	Discrete Gaussian	0.363	1156.63	4.62
	DP Suppression	0.999	1138.62	4.53
	DP Swapping	0.868	12988.58	437.105
	DP k-anonymity	0.878	2963.3	24.7
1	Laplace	**0**	**409.03**	**1.815**
	Discrete Gaussian	0.132	777.455	2.96
	DP Suppression	0.999	1205.315	4.47
	DP Swapping	0.874	11624.64	394.28
	DP k-anonymity	0.906	4296.8	35.4
2	Laplace	**0**	**187.59**	**0.905**
	Discrete Gaussian	0.017	523.825	1.995
	DP Suppression	0.999	1219.19	4.71
	DP Swapping	0.899	8212.7	266.78
	DP k-anonymity	0.981	5406.7	43.4
4	Laplace	**0**	**81.63**	**0.46**
	Discrete Gaussian	3E-4	353.99	1.545
	DP Suppression	0.999	1217.115	4.91
	DP Swapping	0.969	2117.62	75.48
	DP k-anonymity	0.999	5992.0	48.9

Table 2: TX dataset data release: Comparison of DP mechanisms in terms of δ, ℓ_1 norm of the empirical bias and α-fairness.

ϵ	Mechanism	δ	Bias (ℓ_1 norm)	α-fairness
0.5	Laplace	**0**	**2992.88**	**4.02**
	Discrete Gaussian	0.363	3798.96	4.885
	DP Suppression	0.999	3687.445	4.39
	DP Swapping	0.868	35385.63	580.015
	DP k-anonymity	0.878	10260.6	34.5
1	Laplace	**0**	**1372.36**	**2**
	Discrete Gaussian	0.132	2570.38	3.235
	DP Suppression	0.999	3911.82	4.56
	DP Swapping	0.874	32134.71	553.66
	DP k-anonymity	0.906	14668.5	46.1
2	Laplace	**0**	**628.335**	**0.975**
	Discrete Gaussian	0.017	1728.1	2.62
	DP Suppression	0.999	3969.165	4.76
	DP Swapping	0.899	22242.12	339.09
	DP k-anonymity	0.981	18529.7	54.4
4	Laplace	**0**	**274.36**	**0.48**
	Discrete Gaussian	3E-4	1162	1.71
	DP Suppression	0.999	3962.755	4.93
	DP Swapping	0.969	5634.95	108.565
	DP k-anonymity	0.999	20456.9	61.5

Table 3: Outlier dataset data release: Comparison of DP mechanisms in terms of δ, ℓ_1 norm of the empirical bias and α-fairness.

C.3 Classification

This subsection provides plots for accuracies of the logistic regression task described in the paper using various DA methods over the Texas and Outlier dataset (Figure 5). For these

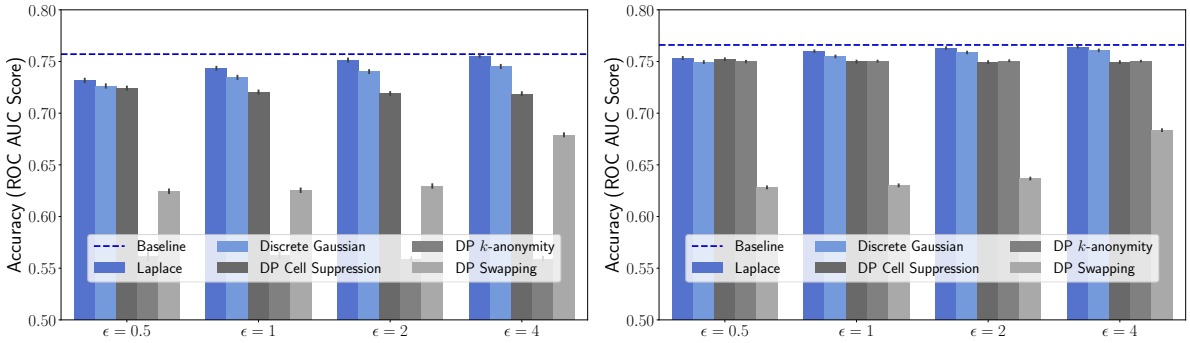

Figure 5: Results for Logistic Regression for Texas (left) and Outlier (right) datasets

datasets as well, it is seen that training logistic regression classifiers with data produced by DP mechanisms (Laplace and discrete Gaussian) yields close-to-baseline classification accuracy. Also, it is seen that using data produced by traditional DA mechanisms yields accuracies that are lower and further away from the baseline accuracy than for any of the DP mechanisms.

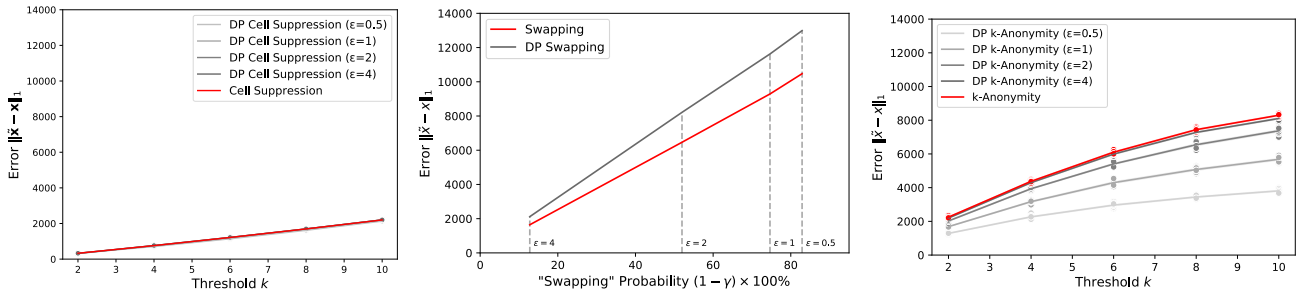

Figure 6: TX ACS dataset: Errors $\|\tilde{\boldsymbol{x}} - \boldsymbol{x}\|_1$ for cell suppression (left), swapping (center) and k-anonymity (right) and their differentially private counterparts (average of 200 repetitions).

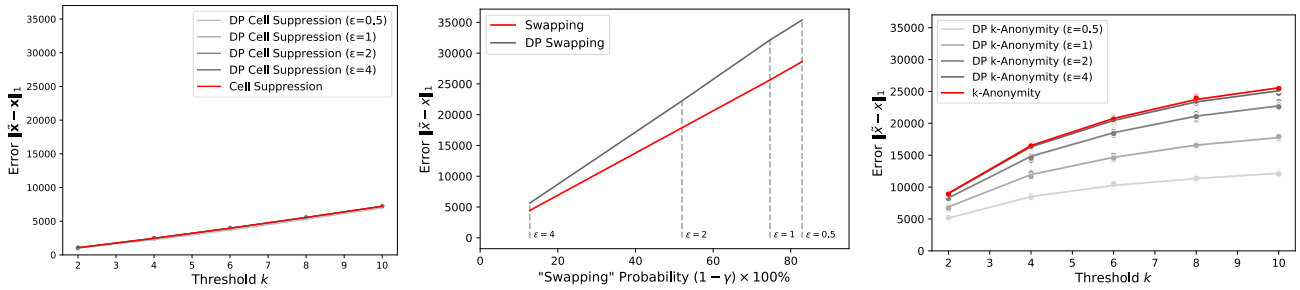

Figure 7: Outlier ACS dataset: Errors $\|\tilde{\boldsymbol{x}} - \boldsymbol{x}\|_1$ for cell suppression (left), swapping (center) and k-anonymity (right) and their differentially private counterparts (average of 200 repetitions).

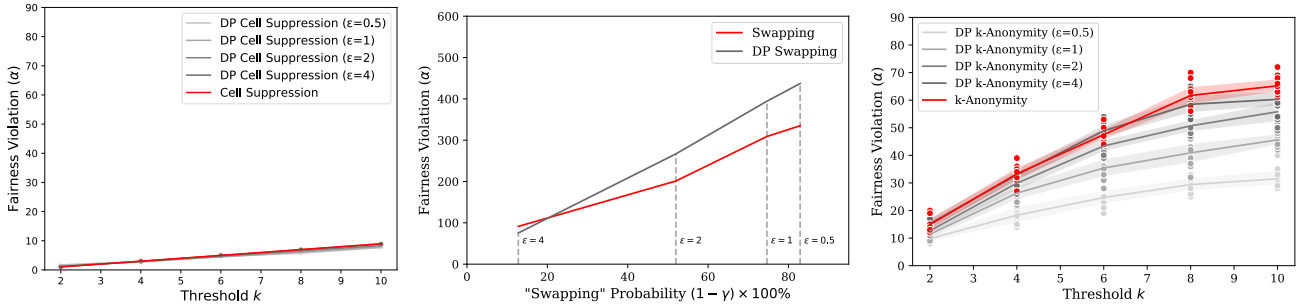

Figure 8: TX ACS dataset: Fairness values α for cell suppression (left), swapping (center) and k-anonymity (right) and their differentially private counterparts (average of 200 repetitions).

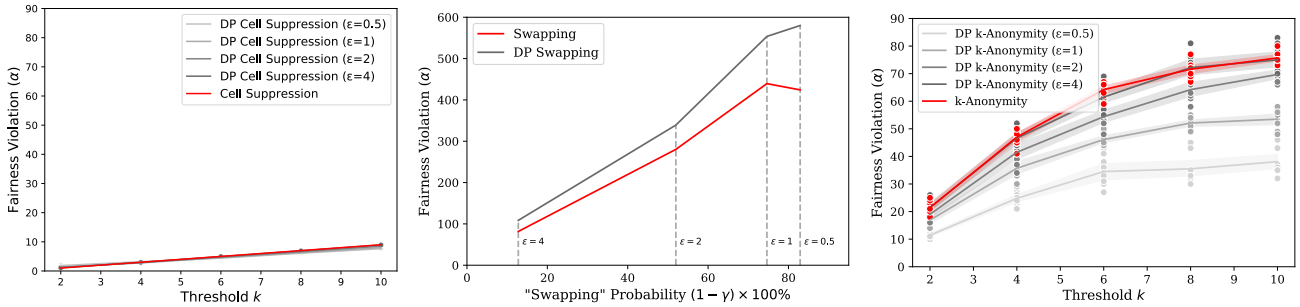

Figure 9: Outlier ACS dataset: Fairness values α for cell suppression (left), swapping (center) and k-anonymity (right) and their differentially private counterparts (average of 200 repetitions).