Bias in allocations using differentially private census data: An analysis of the 2020 U.S. Census

Author One[a,c,1], **Author Two**[b,1,2], **and Author Three**[a]

[a]Affiliation One; [b]Affiliation Two; [c]Affiliation Three

This manuscript was compiled on June 2, 2023

The decennial census is a primary source of data for the US government to make critical decisions. For example, 132 programs used Census Bureau data to distribute more than $675 billion in funds during fiscal year 2015. In order to ensure the published census data does not reveal individual information, the Census Bureau adopts the differential privacy (DP) technique. In particular, in 2020, the Census Bureau implemented the Top Down algorithm which release statistical data from top (nation) to bottom hierarchical level (block). However, it was recently observed that the DP outcomes can introduce biased outcomes, especially for minority groups. In this paper, we analyze the reasons for these disproportionate impacts and proposes guidelines to mitigate these effects. We focus on two aspects that can produce the fair outcomes: (1) shape of allocation function and (2) impact of post-processing steps.

Keyword 1 | Keyword 2 | Keyword 3 | ...

Introduction

Agencies, such as the U.S. Census Bureau, release data sets and statistics about groups of individuals that are then used as inputs to a number critical decision processes. For example, the census data is used to decide whether a jurisdiction must provide language assistance during elections, Title I fund allocation in education (**?**) and to establish national level COVID-19 vaccination distribution plans for states and jurisdictions (**?**). The resulting decisions can have significant societal, economic, and medical consequences for participating individuals.

In many cases, the released data contain sensitive information and their privacy is strictly regulated. For example, in the U.S., the census data is regulated under Title 13 (**?**), which requires that no individual be identified from any data release by the Census Bureau. In Europe, data release are regulated according to the *General Data Protection Regulation* (**?**), which addresses the control and transfer of personal data.

Statistical agencies thus release *privacy-preserving* data and statistics that conform to privacy and confidentiality requirements. In the U.S., a small number of decisions, such as congressional apportionment, are taken using unprotected true values, but the vast majority of decisions rely on privacy-preserving data. Of particular interest are resource allocation decisions relying on the U.S. Census Bureau data, since the bureau will release several privacy-preserving data products using the framework of *Differential Privacy* (**?**) for their 2020 release. In particular, in 2020 the Census Bureau implemented a new privacy preserving framework to release privately hierarchical statistical data, the Top Down algorithm. The algorithms works by firstly splitting the given privacy budget ϵ to six hierarchical levels (nation, state, county, tract, block, group—block). Then in the second step, a post-processing step is applied to make sure the noisy counts are consistent, e.g., the counts should be non-negative. However, (**?**) empirically showed that differential privacy may have a disparate impact on several resource allocation problems. The noise introduced by the privacy mechanism may result in decisions that impact various groups differently. Unfortunately, the paper did not provide a deep understanding why this behavior happens and any mitigation to resolve the issue. This paper builds on these observations and provides a step towards a deeper understanding of the fairness issues arising when differentially private data is used as input to several resource allocation problems. *One of its main results is to prove that several allotment problems and decision rules with significant societal impact (e.g., the allocation of educational funds, the decision to provide minority language assistance on election ballots, or the distribution of COVID-19 vaccines) exhibit inherent unfairness when applied to a differentially private release of the census data.* To counteract this negative results, the paper examines the conditions under which decision making is fair when using differential privacy, and techniques to bound unfairness. The paper also provides a number of mitigation approaches to alleviate biases introduced by differential privacy on such decision making problems. More specifically, the paper makes the following contributions:

1. It formally defines notions of fairness and bounded fairness for decision making subject to privacy requirements.
2. It examines the roots of the induced unfairness by analyzing the structure of the decision making problems.
3. It proposes several guidelines to mitigate the negative fairness effects of the decision problems studied.

To the best of the authors' knowledge, this is the first study that attempt at characterizing the relation between differential privacy and fairness in decision problems. All proofs are reported in the appendix. [Is this still true?]

Significance Statement

Authors must submit a 120-word maximum statement about the significance of their research paper written at a level understandable to an undergraduate educated scientist outside their field of speciality. The primary goal of the significance statement is to explain the relevance of the work in broad context to a broad readership. The significance statement appears in the paper itself and is required for all research papers.

Please provide details of author contributions here.

Please declare any competing interests here.

[1]A.O.(Author One) contributed equally to this work with A.T. (Author Two) (remove if not applicable).

[2]To whom correspondence should be addressed. E-mail: author.twoemail.com

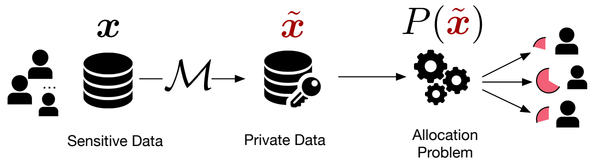

Fig. 1. Diagram of the private allocation problem.

Preliminaries: Differential Privacy

Differential Privacy (**?**) (DP) is a rigorous privacy notion that characterizes the amount of information of an individual's data being disclosed in a computation.

Definition 1 *A randomized algorithm $\mathcal{M} : \mathcal{X} \to \mathcal{R}$ with domain $\mathcal{X}$ and range $\mathcal{R}$ satisfies* ϵ-differential privacy *if for any output $O \subseteq \mathcal{R}$ and data sets $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}$ differing by at most one entry (written $\boldsymbol{x} \sim \boldsymbol{x}'$)*

$$\Pr[\mathcal{M}(\boldsymbol{x}) \in O] \leq \exp(\epsilon) \Pr[\mathcal{M}(\boldsymbol{x}') \in O]. \quad [1]$$

Parameter $\epsilon > 0$ is the *privacy loss*, with values close to 0 denoting strong privacy. Intuitively, DP states that any event occur with similar probability regardless of the participation of any individual data to the data set.

DP satisfies several properties including *immunity to post-processing*, which states that the privacy loss of DP outputs is not affected by arbitrary data-independent post-processing (**?**).

A function f from a data set $\boldsymbol{x} \in \mathcal{X}$ to a result set $R \subseteq \mathbb{R}^n$ can be made differentially private by injecting random noise onto its output. The amount of noise relies on the notion of *global sensitivity* $\Delta_f = \max_{\boldsymbol{x} \sim \boldsymbol{x}'} \|f(\boldsymbol{x}) - f(\boldsymbol{x}')\|_1$. The *Laplace mechanism* (**?**) that outputs $f(\boldsymbol{x}) + \boldsymbol{\eta}$, where $\boldsymbol{\eta} \in \mathbb{R}^n$ is drawn from the i.i.d. Laplace distribution with 0 mean and scale ${}^{\Delta_f}/_{\epsilon}$ over n dimensions, achieves ϵ-DP. Similarly, once can also utilize the Gaussian mechanism to satisfy ϵ-DP by sampling from the Gaussian distribution with mean 0 and scale ${}^{\Delta_f}/_{\epsilon}$.

Problem Setting and Goals

The paper considers a dataset $\boldsymbol{x} \in \mathcal{X} \subseteq \mathbb{R}^k$ of n entities, whose elements $x_i = (x_{i1}, \ldots, x_{1k})$ describe k measurable quantities of entity $i \in [n]$, such as the number of individuals living in a geographical region i and their English proficiency. The paper considers two classes of problems:

- An *allotment problem* $P : \mathcal{X} \times [n] \to \mathbb{R}$ is a function that distributes a finite set of resources to some problem entity. P may represent, for instance, the amount of money allotted to a school district.
- A *decision rule* $P : \mathcal{X} \times [n] \to \{0, 1\}$ determines whether some entity qualifies for some benefits. For instance, P may represent if election ballots should be described in a minority language for an electoral district.

The paper assumes that P has bounded range, and uses the shorthand $P_i(\boldsymbol{x})$ to denote $P(\boldsymbol{x}, i)$ for entity i. The focus of the paper is to study the effects of a DP data-release mechanism $\mathcal{M}$ to the outcomes of problem P. Mechanism $\mathcal{M}$ is applied to the dataset $\boldsymbol{x}$ to produce a privacy-preserving counterpart $\tilde{\boldsymbol{x}}$ and the resulting private outcome $P_i(\tilde{\boldsymbol{x}})$ is used to make some allocation decisions.

Figure 1 provides an illustrative diagram.

Because random noise is added to the original dataset $\boldsymbol{x}$, the output $P_i(\tilde{\boldsymbol{x}})$ incurs some error. *The focus of this paper is to characterize and quantify the disparate impact of this error among the problem entities.* In particular, the paper focuses on two notations of errors.

Definition 2 (Statistical bias) *The statistical bias $B_P^i(\mathcal{M}, \boldsymbol{x})$ of the mechanism $\mathcal{M}$ measures the difference between the expected private outcome with the true outcome:*

$$B_P^i(\mathcal{M}, \boldsymbol{x}) = \mathbb{E}_{\tilde{\boldsymbol{x}} \sim \mathcal{M}(\boldsymbol{x})} \left[P_i(\tilde{\boldsymbol{x}}) \right] - P_i(\boldsymbol{x}), \quad [2]$$

The paper also considers another notation of error which is the normalized version of the above bias.

Definition 3 (Multiplicative error) *The multiplicative error under mechanism $\mathcal{M}$ and problem P for entity i is given by:* ${}^{B_P^i(\mathcal{M}, \boldsymbol{x})}/_{P^i(x)}$

Our notion of fairness will be based on these two notations of errors.

Definition 4 (α-fairness (?)) *Given the true data $\boldsymbol{x}$, the mechanism $\mathcal{M}$ is said to be α-fair if, for any $i \in [n]$,*

$$\xi_P^i(\mathcal{M}, \boldsymbol{x}) = \left| B_P^i(\mathcal{M}, \boldsymbol{x}) - B_P^j(\mathcal{M}, \boldsymbol{x}) \right| \leq \alpha \,,$$

where $\xi_P^i(\mathcal{M}, \boldsymbol{x})$ is referred to as the disparity error *associated with district i. The mechanism $\mathcal{M}$ is α'-minimally fair if $\alpha' = \inf \alpha$ such that $\mathcal{M}$ is α-fair. To put it differently, the mechanism $\mathcal{M}$ is α'-minimally fair if*

$$\begin{aligned} \alpha' &= \max_{j \neq i} \left| B_P^i(\mathcal{M}, \boldsymbol{x}) - B_P^j(\mathcal{M}, \boldsymbol{x}) \right| \\ &= \max_{j \in [n]} B_P^j(\mathcal{M}, \boldsymbol{x}) - \min_{j \in [n]} B_P^j(\mathcal{M}, \boldsymbol{x}) \,. \end{aligned}$$

Throughout this report, every time we say that a mechanism is α-fair, we mean that it is α-minimally fair.

Motivating examples

A. Title I school allocation. The *Title I of the Elementary and Secondary Education Act of 1965* (**?**) distributes funds through its basic, concentration, and target grants which account for \$6.2B, \$1.3B, and \$4.2B of allocations, respectively. The federal allotment is divided among nearly 17000 qualifying school districts in proportion to the count x_i of children aged 5 to 17 who live in necessitous families in district i. Each grant is specified by a set of thresholds that determines which districts are eligible for that particular grant. We refer to (**?**) which discusses the specifics of each grant, which we describe briefly here.

The basic allocation grant is formalized by:

$$P_i^F(\boldsymbol{x}) \stackrel{\text{def}}{=} \left(\frac{x_i \cdot a_i}{\sum_{i \in [n]} x_i \cdot a_i} \right) \cdot F_B \quad [3]$$

where $\boldsymbol{x} = (x_i)_{i \in [n]}$ is the vector of all eligible districts counts and a_i is a weight factor reflecting students expenditures (defined as the adjusted state per-pupil expenditure). F_B is the total basic Title I appropriation for the fiscal year. Finally, eligible districts are computed by thresholding; each grant type follows slightly different rules for eligbility. For example, a district is eligible for the basic grant if the population of eligible students is > 10 and the proportion of eligible students to total students in the district is > 0.02. Concentration grant eligibility is determined if the district contains > 6500 individuals or the proportion of eligible students is > 0.15. Lastly, districts are eligible for the targeted grant if they have > 10 eligible students and the proportion of eligible students is > 0.05. The funding is also weighted by population, as outlined in Section A of (**?**)

B. Privacy Budget. The Disclosure Avoidance System for the US Census Bureau defines a global rho that is distributed hierarchically. The global privacy-loss budget for 2020 was defined as $\rho = 2.56$. The DAS suggests the use of this bound to conver the ρ-based privacy-loss budgets to the $\epsilon, \delta)$ equivalents. $= \rho + 2 * \sqrt{-\rho * \log_e \delta}$ where we define $\delta = 10e-10$. In a 1-year estimate of the ACS, there are 1426 different "Detailed tables" available from the US Census Bureau. The rho allocation for the national, state, and county level geographic hierarchies are $\frac{104}{4099}$, $\frac{1440}{4099}$, and $\frac{447}{4099}$ respectively. Assuming that the county level privacy-loss budget is split equally across these tables, each field would have $\rho = 2.56 \cdot \frac{447}{4099} \cdot \frac{1}{1426} = 0.0001957$ at the county level, $\rho = 0.00063$ at the state level, and $\rho = 0.0000455$ at the national level.

We consider applying the Gaussian mechanism to satisfy differential privacy on children population estimate data along with a hierarchical constraint that ensures that the sum of the noisy population estimate of sub-hierarchies are consistent with their parent. For example, the sum of Title I-eligible children in the districts of Illinois should be consistent with the noisy estimate of Title I-eligible children at the state level. We experiment with three privacy budgets $\rho = 2.65$ and $\rho = 1.0$ and $\rho = 0.1$ to release privately the schools' population. These private counts are used to determine amount of money each school district should receive using the allocation mechanisms described above for the basic, concentration, and target grants. The statistical bias in Definition 2 here represents the gain/loss in USD a school can have under the privacy preserving mechanism. 2 summarizes our findings of the bias as a function of district size for all 13,190 districts. We observe that misallocations are most pronounced at thresholds for each grant, and substantially impact smaller school districts more than larger ones.

It can be seen from Figure 2 that schools of low population receive far more money than they actually need as a consequence of the bias in the allocation mechanism, while larger districts may actually receive less funding. This figure made a clear evidence on the disparate impact of the privacy preserving mechanism in practice.

Further analysis. While there is a some level of unfairness under privacy-preserving mechanism towards different school districts, we observed that level varies differently by state. We report the fairness level per state, i.e the difference between the maximum bias minus the minimum bias per state in Figure 8. We see that under $\rho = 2.65$ California has the largest unfairness level over its schools.

To understand why California has the highest level of unfairness, we provide the scatter plot between level of fairness α per state with its number of schools and its number of students in Figure 3. We observed a Pearson correlation of 0.37 between the number of schools per state and its fairness level.

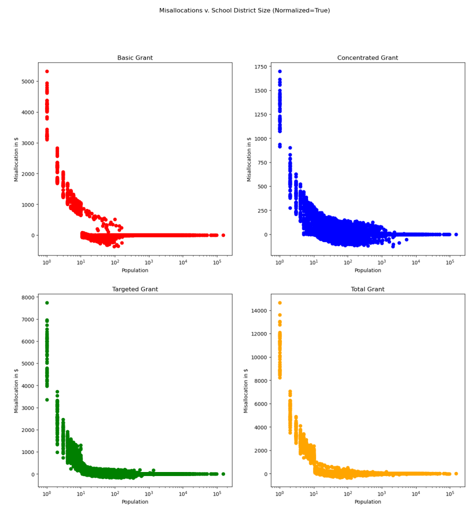

Fig. 2. Disproportionate Title 1 Funds Allocations in the US

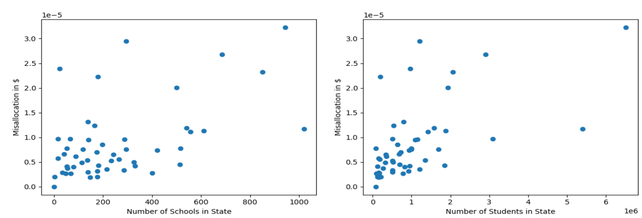

Fig. 3. Correlation between number of schools (left) and number of students (right) per state with its level of α fairness.

Also, a high Pearson correlation of 0.71 between number of state's students and its fairness level . A In other words, the state with more schools and more students tends to be suffered more unfairness. For example, the California state has the highest number of students, more than 6.5 million students overall. It also ranks secondly based on the number of schools with 1022 schools just below Texas with 1222 schools. This partly explains why this state suffers the most unfairness for all schools as showed in Figure 8.

Misallocations noramlized per student. Examining both total misallocations and misallocations per student is crucial when assessing the impact of differential privacy on the allocation of Title I education funds in districts. Total misallocations provide a comprehensive perspective by considering the absolute magnitude of funding disparities across states. It enables us to identify states that are disproportionately affected in terms of the overall amount of misallocated funds. This analysis is important because significant total misallocations could lead to substantial disparities in educational resources, hindering the ability of disadvantaged districts to provide quality education.

On the other hand, evaluating misallocations per student offers a more nuanced understanding of the fairness of funding distribution. It allows us to account for the relative impact on individual students within each state. By examining misallocations per student, we can identify states where the disparity in funding affects a larger proportion of the student population, potentially exacerbating educational inequalities and limiting opportunities for marginalized students.

When funds are misallocated in total, it can particularly affect certain types of funding. For instance, if a state with a higher population of disadvantaged students experiences a significant total misallocation, it may receive insufficient funding to adequately support targeted interventions, such as specialized educational programs, additional resources, or support services for at-risk students. This can further perpetuate existing disparities and hinder efforts to address educational inequities.

Similarly, when funds are seen misallocated per student, it highlights the distributional challenges faced by specific states. If a state with a higher concentration of disadvantaged students experiences larger misallocations per student, it may imply that individual students in that state are being deprived of the financial support they require to succeed academically. This can impact the availability of resources such as instructional materials, qualified teachers, technology, or extracurricular activities, impeding their educational progress and perpetuating systemic inequalities.

In conclusion, analyzing both total misallocations and misallocations per student provides a comprehensive understanding of the impact of differential privacy on the allocation of Title I education funds. This dual perspective helps identify states disproportionately affected in terms of both the absolute amount of misallocated funds and the relative impact on individual students. By examining these two aspects, policymakers can gain insights into the specific challenges faced by different states and devise targeted strategies to ensure fair and equitable distribution of education funds.

Section 203 of the Voting Rights Act

Section 203 of the Voting Rights Act requires certain jurisdictions to provide bilingual election materials and assistance to voters who are not proficient in English. To determine which jurisdictions are covered by this provision, the Census Bureau collects data on the number and percentage of voting-age citizens who are members of a language minority group and have limited English proficiency.

The Census Bureau uses three measures to calculate these numbers: the Limited English Proficient Population Count (LEPPCT), the Voting Age Citizen In-Language Count (VACIT), and the Voting Age Citizen Language Minority Group Citizen Voting Age Population (VACLEP).

The LEPPCT is the percentage of people in a language minority group who have limited English proficiency. The VACIT is the number of voting-age citizens who are members of a language minority group and who have indicated a need for language assistance in voting. The VACLEP is the number of voting-age citizens who are members of a language minority group.

By using these three measures, the Census Bureau determines which jurisdictions meet the criteria for coverage under Section 203. This information is then used by election officials to provide bilingual election materials and assistance to voters who need it, ensuring that everyone has an equal opportunity to participate in the electoral process.

Allocation Formula. To explain how coverage is computed based on the given formulas, we can break them down into terms:

Differentially Private mechanisms

To release privately the outcomes P_i^F given a privacy constraint ϵ, there are several mechanisms. These mechanisms can roughly be divided into the two categories, strict and non-strict allocation mechanisms. A strict allocation mechanism requires that its outcome should always lie in the probability simplex $\Delta_n = \{x \,|x \in \mathbb{R}^{+n}, \mathbf{1}^T x = 1\}$ while a non-strict allocation mechanism only asks its output to be non-negative. The rest of this report aims to study the (approximate) optimal (strict allocation) mechanisms under different fairness metrics.

Strict Allocation Mechanism.

Definition 5 (Baseline Mechanism (BL)) *The* baseline mechanism *outputs the allocation for each distribute $i \in [n]$ as follows.*

$$\mathcal{M}_{\mathrm{BL}}\left(\tilde{\boldsymbol{x}}\right)_i = \frac{a_i \cdot \left(\tilde{x}_i\right)_+}{\sum_{j=1}^{n} a_j \cdot \left(\tilde{x}_j\right)_+} .$$

Where $\tilde{x}_i$ is the noisy private population count, while the supscript $x_+ = \max(x, 0)$ takes the non-negative part of the number x.

Definition 6 (Projection onto Simplex Mechanism (PoS)) *The* projection onto simplex mechanism *outputs the allocation for each distribute $i \in [n]$ as follows.*

$$\mathcal{M}_{\mathrm{PoS}}\left(\tilde{\boldsymbol{x}}\right)_i = \underset{\boldsymbol{v}\in\mathbb{R}^n}{\arg\min}\ \left\|\boldsymbol{v} - P^F\left(\tilde{\boldsymbol{x}}\right)\right\|_2 \qquad \text{s.t.} \sum_{i=1}^{n} v_i = 1,\ \boldsymbol{v} \geq \mathbf{0}\,.$$

Non-strict Allocation Mechanism.

Definition 7 (Positive Allocation Mechanism (PA)) *The* positive allocation mechanism *outputs the allocation for each distribute $i \in [n]$ as follows.*

$$\mathcal{M}_{\mathrm{PA}}\left(\tilde{\boldsymbol{x}}\right)_i = \left(P_i^F\left(\tilde{\boldsymbol{x}}\right)\right)_+ = \left(\frac{a_i \cdot \tilde{x}_i}{\sum_{j=1}^{n} a_j \cdot \tilde{x}_j}\right)_+ .$$

Definition 8 (Repair Mechanism (RP) (?)) *The* repair mechanism *outputs the allocation for each distribute $i \in [n]$ as follows.*

$$\mathcal{M}_{\mathrm{RP}}\left(\tilde{\boldsymbol{x}}\right)_i = \frac{a_i \cdot \left(\tilde{x}_i\right)_+ + \Delta}{\sum_{j=1}^{n} a_j \cdot \left(\tilde{x}_j\right)_+ - \Delta'} ,$$

where

$$\Delta = \frac{\ln\left(2n/\delta\right)}{\epsilon}, \qquad \Delta' = \frac{n\ln\left(2n^2/\delta\right)}{\epsilon} .$$

Proposition 1 (No-penalty allocation (?)) *The following inequality holds with probability at least $1-\delta$.*

$$\mathcal{M}_{\mathrm{RP}}\left(\tilde{\boldsymbol{x}}\right)_i \geq P_i^F\left(\boldsymbol{x}\right) , \qquad \forall\ i \in [n] .$$

Source of unfairness

We investigate the two main sources of unfairness highlighted in previous section: (1) shape of allocation function and (2) post-processing steps.

Shape of allocation function.

Theorem 1 *Let P be an allotment problem which is at least twice differentiable. A data-release mechanism $\mathcal{M}$ is α-fair w.r.t. P for some $\alpha < \infty$ if there exist some constant values c^i_{jl} $(i \in [n], j, l \in [k])$ such that, for all datasets $\boldsymbol{x} \in \mathcal{X}$,*

$$(\boldsymbol{H} P_i)_{j,l}(\boldsymbol{x}) = c^i_{j,l} \;\; (i \in [n] \; j, l \in [k]).$$

Corollary 1 *If P is a linear function, then $\mathcal{M}$ is fair w.r.t. P.*

Corollary 2 *$\mathcal{M}$ is fair w.r.t. P if there exists a constant c such that, for all dataset $\boldsymbol{x}$,*

$$Tr(\boldsymbol{H} P_i)(\boldsymbol{x}) = c \;\; (i \in [n]).$$

Corollary 3 *Consider an allocation problem P. Mechanism $\mathcal{M}$ is not fair w.r.t. P if there exist two entries $i, j \in [n]$ such that $Tr(\boldsymbol{H} P_i)(\boldsymbol{x}) \neq Tr(\boldsymbol{H} P_j)(\boldsymbol{x})$ for some dataset $\boldsymbol{x}$.*

The above implies that fairness cannot be achieved if P is *a non-convex function*, as is the case for *all* the allocation problems considered in this paper. *A fundamental consequence of this result is the recognition that adding Laplacian noise to the inputs of the motivating example will necessarily introduce fairness issues.* For instance, consider P^F and notice that the trace of its Hessian

$$\text{Tr}(\boldsymbol{H} P^F_i) = 2a_i \left[\frac{x_i \sum_{j\in[n]} a_j^2 - a_i \left(\sum_{j\in[n]} x_j a_j \right)}{\left(\sum_{j\in[n]} x_j a_j \right)^3} \right],$$

is not constant with respect to its inputs. Thus, any two entries i, j whose $x_i \neq x_j$ imply $\text{Tr}(\boldsymbol{H} P^F_i) \neq \text{Tr}(\boldsymbol{H} P^F_j)$. As illustrated in Figure **??**, Problem P^F can introduce significant disparity errors. For $\epsilon = 0.001, 0.01$, and 0.1 the estimated fairness bounds are 0.003, 3×10^{-5}, and 1.2×10^{-6} respectively, which amount to an average misallocation of \$43,281, \$4,328, and \$865.6 respectively. The estimated fairness bounds were obtained by performing a linear search over all n school districts and selecting the maximal $\text{Tr}(\boldsymbol{H} P^F_i)$.

Impact of post-processing. The post-processing steps can be applied at the inputs x, over the outcome $P^F_i(x)$ or at both. We investigate the impact of post-processing over input and outcome separately in this section.

Post-processing over the inputs. This step is performed to make sure the released private counts satisfies consistency constraints (**?**). For example, the released private counts should be non-negative integer numbers, or sum of counts at all cities' in a state should be equal to that state's count.

Non-negative truncation $\tilde{x} = \max(0, \tilde{x})$

We have the following result which state that non-negative truncation introduces positive bias, and the closer to zero the true count is, the higher the bias.

Theorem 2 *Let $\tilde{x} = x + Lap(\lambda)$, with scale $\lambda > 0$, and $\hat{x} = PP^{\geq \ell}(\tilde{x})$, with $\ell < x$, be its post-processed value. Then,*

$$\mathbb{E}[\hat{x}] = x + \frac{\lambda}{2} \exp(\frac{\ell - x}{\lambda}).$$

Integral transform

The integral transform $\text{PP}^{\mathbb{N}}(z)$ is used when the released data should be of integral quantities. To make sure that this processing step does not introduce additional bias, we can rely on the stochastic rounding technique:

$$\text{PP}^{\mathbb{N}}(z) = \begin{cases} \lfloor z \rfloor & \text{w.p.: } 1 - (z - \lfloor z \rfloor) \\ \lfloor z \rfloor + 1 & \text{w.p.: } z - \lfloor z \rfloor \end{cases} \qquad [4]$$

The stochastic rounding guarantees that $\mathbb{E}[\text{PP}^{\mathbb{N}}(\tilde{x})] = \tilde{x}$ so no additional bias will introduce to $\text{PP}^{\mathbb{N}}(\tilde{x})$

Post-processing over the outcomes.

Mitigating Solutions

Mechanisms. Different kind of post-processing mechanisms are considered in this section. These mechanisms require that their outcomes should always lie in the probability simplex Δ_n. The rest of this work aims to study the (approximate) optimal mechanisms under different fairness metrics.

Definition 9 (Baseline Mechanism (BL)) *The* baseline mechanism *outputs the allocation for each entity $i \in [n]$ as follows.*

$$\mathcal{M}_{\text{BL}}(\tilde{\boldsymbol{x}})_i = \frac{a_i \cdot (\tilde{x}_i)_+}{\sum_{j=1}^n a_j \cdot (\tilde{x}_j)_+}.$$

Definition 10 (Projection onto Simplex Mechanism (PoS)) *The* projection onto simplex mechanism *outputs the allocation for each entity $i \in [n]$ as follows.*

$$\mathcal{M}_{\text{PoS}}(\tilde{\boldsymbol{x}})_i = \operatorname*{arg\,min}_{\boldsymbol{v} \in \mathbb{R}^n} \left\| \boldsymbol{v} - P^F(\tilde{\boldsymbol{x}}) \right\|_2 \qquad \text{s.t.} \sum_{i=1}^n v_i = 1, \; \boldsymbol{v} \geq \boldsymbol{0}.$$

[t]0.49

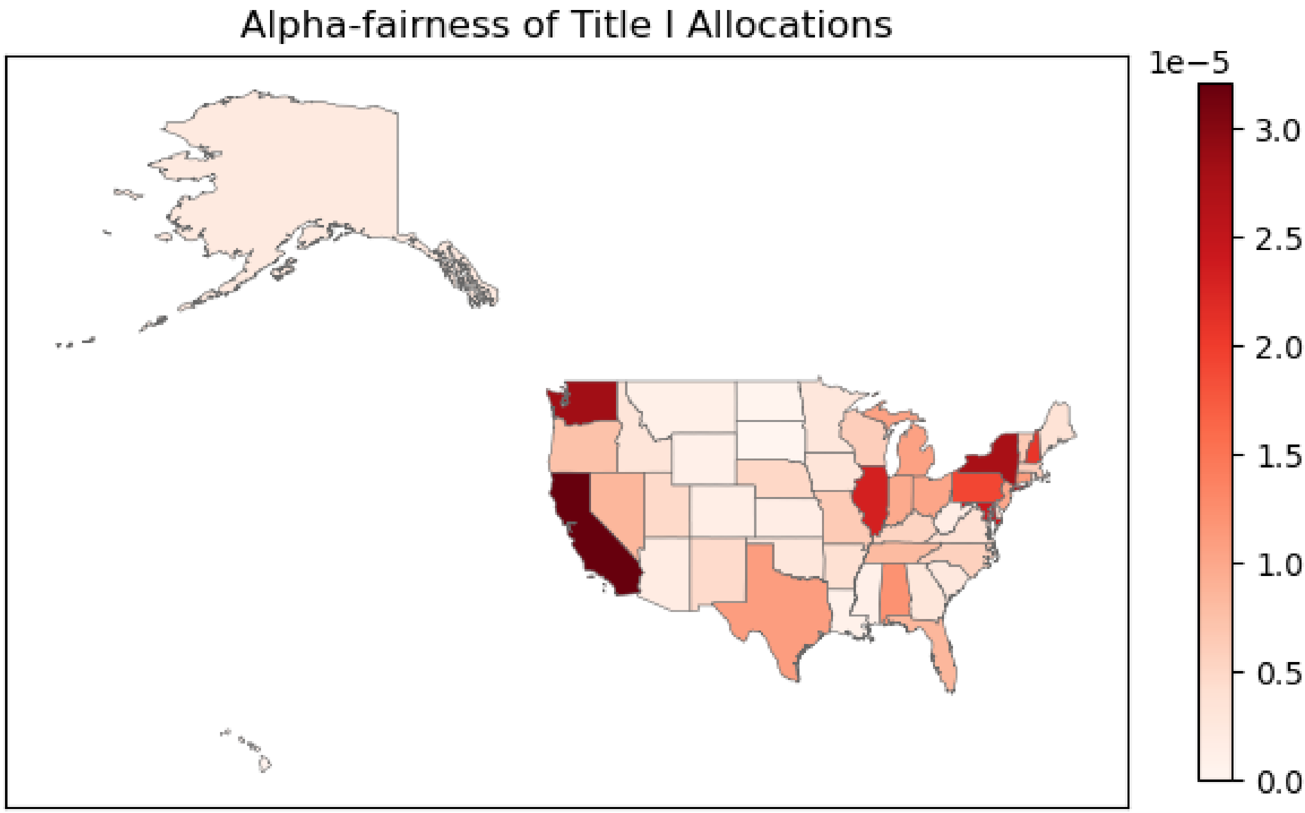

Fig. 4. Level of α fairness per state under Baseline mechanism (BL) at $\rho = 2.56$

[t]0.49

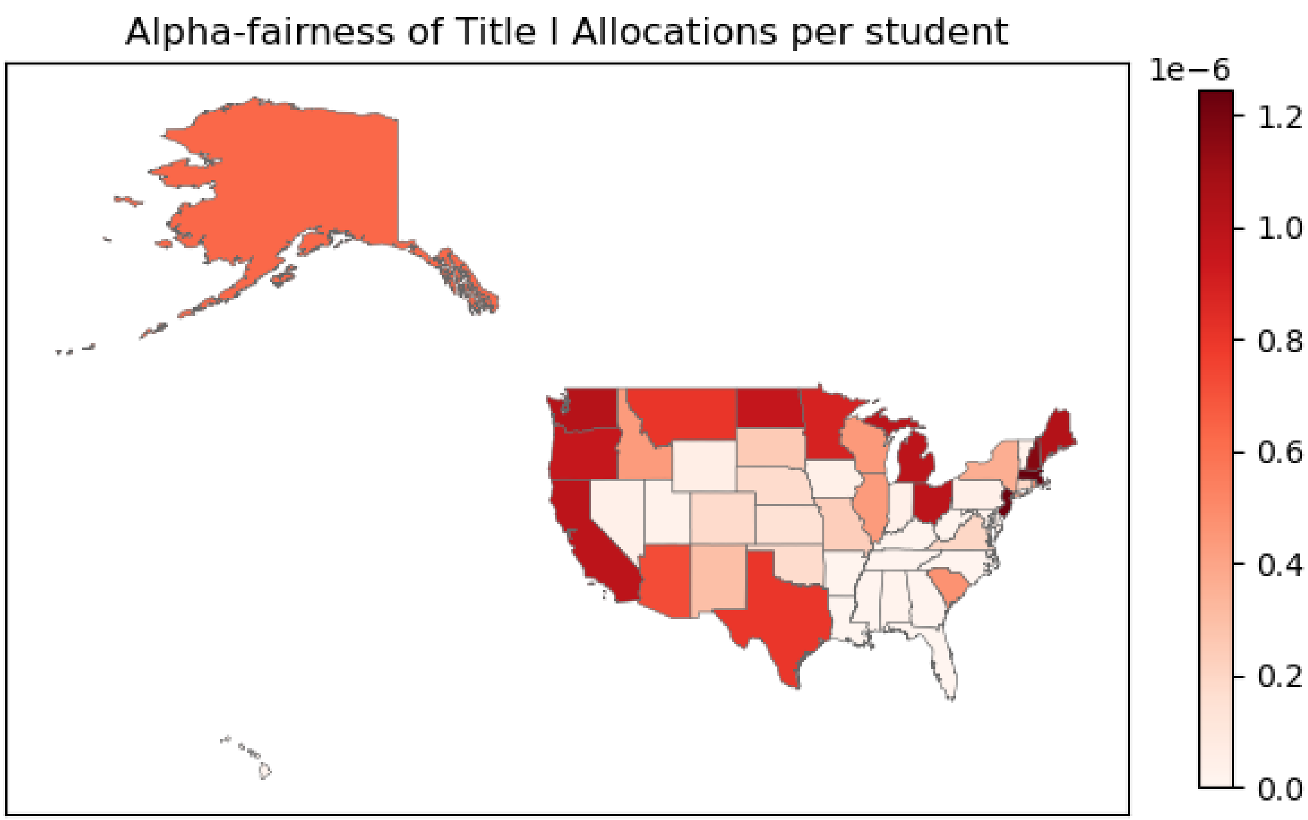

Fig. 5. Level of α fairness per student under Baseline mechanism (BL) at $\rho = 2.56$

Fig. 6. Fairness comparison

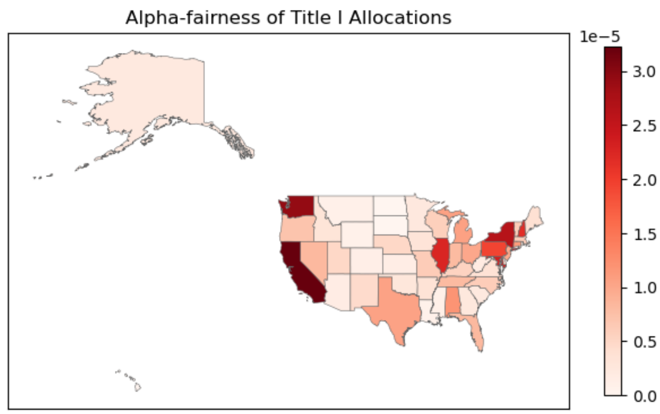

Fig. 7. Level of α fairness per state under Baseline mechanism (BL) at $\rho = 2.56$

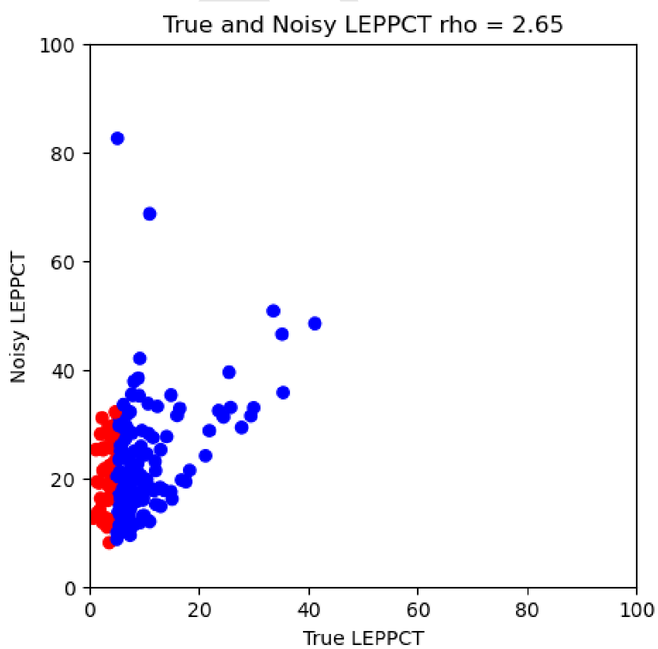

Fig. 8. LEPPCT vs. LEPPCT noisy values after DP mechanism at $\rho = 2.65$