Decision Making with Differential Privacy under a Fairness Lens

Cuong Tran[1] , **Ferdinando Fioretto**[1] , **Pascal Van Hentenryck**[2] and **Zhiyan Yao**[3*]
[1]Syracuse University
[2]Georgia Institute of Technology
[3]Nanjing University of Science and Technology
{cutran, ffiorett}@syr.edu, pvh@isye.gatech.edu, zyao09@syr.edu

Abstract

Many agencies release datasets and statistics about groups of individuals that are used as input to a number of critical decision processes. To conform with privacy and confidentiality requirements, these agencies are often required to release privacy-preserving versions of the data. This paper studies the release of differentially private datasets and analyzes their impact on some critical resource allocation tasks under a fairness perspective. The paper shows that, when the decisions take as input differentially private data, the noise added to achieve privacy disproportionately impacts some groups over others. The paper analyzes the reasons for these disproportionate impacts and proposes guidelines to mitigate these effects. The proposed approaches are evaluated on critical decision problems that use differentially private census data.

1 Introduction

Statistics about groups of individuals are often used as inputs to critical decision processes. The U.S. Census Bureau, for example, releases data that is then used to allocate funds and distribute critical resources to states and jurisdictions. These decision processes may determine whether a jurisdiction must provide language assistance during elections, establish vaccine distribution plans, and allocate funds to school districts. The resulting decisions may have significant societal, economic, and medical impacts for participating individuals.

In many cases, the released data contain sensitive information whose 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 released by the Census Bureau. In Europe, data release is regulated according to the *General Data Protection Regulation*, which addresses the control and transfer of personal data. As a result, such data releases must necessarily rely on privacy-preserving technologies. Differential Privacy (DP) has become the paradigm of choice for protecting data privacy, and its deployments include data products related with the 2020 release of the U.S. Census Bureau [Abowd, 2018].

Although DP provides strong privacy guarantees, *it may induce biases and fairness issues in downstream decision processes*, as shown empirically in [Pujol *et al.*, 2020]. Since at least $675 billion are being allocated based on U.S. census data, the use of differential privacy without a proper understanding of these biases and fairness issues may adversely affect the health, well-being, and sense of belonging of many individuals. Indeed, the allotment of federal funds, apportionment of congressional seats, and distribution of vaccines should ideally be fair and unbiased. Similar issues arise in several other areas including election, energy, and food policies. The problem is further exacerbated by the recent recognition that *commonly adopted DP mechanisms for data release may introduce unexpected biases on their own, independently of a downstream decision process [Zhu* et al.*, 2021].*

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 decision problems with significant societal impact (e.g., the allocation of educational funds and the decision to provide minority language assistance on election ballots) 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 DP, and techniques to bound unfairness. The paper also provides a number of mitigation approaches to alleviate unfairness on such decision making problems. More specifically, the paper makes the following contributions: **(1)** It formalizes the notion of bounded fairness for decision making subject to privacy requirements. **(2)** It characterizes decision making problems that are fair or admits bounded fairness. In addition, it investigates the composition of decision rules and how they impact bounded fairness. **(3)** It proves that several decision problems with high societal impact induce inherent biases when using a differentially private input. **(4)** It examines the roots of the induced unfairness by analyzing the structure of the decision making problems. **(5)** 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 [Fioretto *et al.*, 2021].

*Work done while the author was visiting Syracuse University.

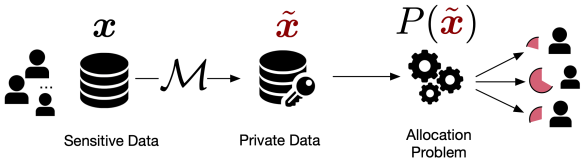

Figure 1: Diagram of the private allocation problem.

2 Preliminaries: Differential Privacy

Differential Privacy [Dwork *et al.*, 2006] (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 datasets $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}$ differing by at most one entry (written $\boldsymbol{x} \sim \boldsymbol{x}'$)*

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

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 dataset. 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 [Dwork and Roth, 2013].

A function f from a dataset $\boldsymbol{x} \in \mathcal{X}$ 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* [Dwork *et al.*, 2006] 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 Δ_f/ϵ over n dimensions, achieves ϵ-DP.

3 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 measuring the bias of problem P_i

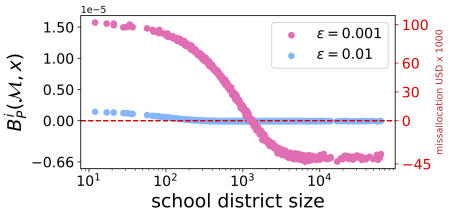

Figure 2: Disproportionate Title 1 Funds Allocation in NY.

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

which characterizes the distance between the expected privacy-preserving allocation and the one based on the ground truth. The paper considers the absolute bias $|B_P^i|$, in place of the bias B_P^i, when P is a decision rule. The distinction will become clear in the next sections.

The results in the paper assume that $\mathcal{M}$, used to release counts, is the Laplace mechanism with an appropriate finite sensitivity Δ. *However, the results are general and apply to any data-release DP mechanism that add unbiased noise*.

4 Motivating Problems

This section reviews two Census-motivated problem classes that grant benefits or privileges to groups of people [Pujol *et al.*, 2020].

Allotment problems The *Title I of the Elementary and Secondary Education Act of 1965* [Sonnenberg, 2016] distributes about $6.5 billion through basic grants. The federal allotment is divided among qualifying school districts in proportion to the count x_i of children aged 5 to 17 who live in necessitous families in district i. The allocation is formalized by

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

where $\boldsymbol{x} = (x_i)_{i \in [n]}$ is the vector of all districts counts and a_i is a weight factor reflecting students expenditures.

Figure 2 illustrates the expected disparity errors arising when using private data as input to problem P^F, for various privacy losses ϵ. These errors are expressed in terms of bias (left y-axis) and USD misallocation (right y-axis) across the different New York school districts, ordered by their size. The allotments for small districts are typically overestimated while those for large districts are underestimated. Translated in economic factors, some school districts may receive up to 42,000 dollars less than warranted.

Decision Rules *Minority language voting right benefits* are granted to qualifying voting jurisdictions. The problem is formalized as

$$P_i^M(\boldsymbol{x}) \stackrel{\text{def}}{=} \left(\frac{x_i^{sp}}{x_i^s} > 0.05 \vee x_i^{sp} > 10^4 \right) \wedge \frac{x_i^{spe}}{x_i^{sp}} > 0.0131.$$

For a jurisdiction i, x_i^s, x_i^{sp}, and x_i^{spe} denote, respectively, the number of people in i speaking the minority language of

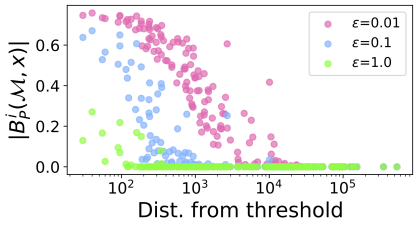

Figure 3: Disproportionate Minority Language Voting Benefits.

interest, those that have also a limited English proficiency, and those that, in addition, have less than a 5^{th} grade education. Jurisdiction i must provide language assistance (including voter registration and ballots) *iff* $P_i^M(\boldsymbol{x})$ is *True*.

Figure 3 illustrates the decision error (y-axis), corresponding to the absolute bias $|B^i_{P^M}(\mathcal{M}, \boldsymbol{x})|$, for sorted x^s_i, considering only true positives[1] for the *Hispanic* language. The figure shows that there are significant disparities in decision errors and that these errors strongly correlate to their distance to the thresholds. Similar issues were also observed in [Pujol *et al.*, 2020].

5 Fair Allotments and Decision Rules

This section analyzes the fairness impact in allotment problems and decision rules. The adopted fairness concept captures the desire of equalizing the allocation errors among entities, which is of paramount importance given the critical societal and economic impact of the motivating applications.

Definition 2. *A data-release mechanism $\mathcal{M}$ is said fair w.r.t. a problem P if, for all datasets $\boldsymbol{x} \in \mathcal{X}$,*

$$B^i_P(\mathcal{M}, \boldsymbol{x}) = B^j_P(\mathcal{M}, \boldsymbol{x}) \quad \forall i, j \in [n].$$

That is, P does not induce disproportionate errors when taking as input a DP dataset generated by $\mathcal{M}$. The paper also introduces a notion to quantify and bound the mechanism unfairness.

Definition 3. *A mechanism $\mathcal{M}$ is said α-fair w.r.t. problem P if, for all datasets $\boldsymbol{x} \in \mathcal{X}$ and all $i \in [n]$,*

$$\xi^i_B(P, \mathcal{M}, \boldsymbol{x}) = \max_{j \in [n]} \left| B^i_P(\mathcal{M}, \boldsymbol{x}) - B^j_P(\mathcal{M}, \boldsymbol{x}) \right| \leq \alpha,$$

where ξ^i_B is referred to as the disparity error *of entity i.*

Parameter α is called the *fairness bound* and captures the fairness violation, with values close to 0 denoting strong fairness. A fair mechanism is also 0-fair.

Note that computing the fairness bound α analytically may involve computing the expectation of complex functions P. Therefore, in the analytical assessments, the paper recurs to a sampling approach to compute the *empirical expectation* $\hat{E}[P_i(\tilde{\boldsymbol{x}})] = \frac{1}{m} \sum_{j \in [m]} P_i(\tilde{\boldsymbol{x}}^j)$ in place of the true expectation in Equation (2). Therein, m is a sufficiently large sample size and $\tilde{\boldsymbol{x}}^j$ is the j-th outcome of the application of mechanism $\mathcal{M}$ on dataset $\boldsymbol{x}$.

[1]This is because misclassification, in this case, implies potentially disenfranchising a group of individuals.

5.1 Fair Allotments: Characterization

The first result characterizes a sufficient condition for the allotment problems to achieve finite fairness violations. The presentation uses $\boldsymbol{H}P_i$ to denote the Hessian of problem P_i, and $\mathrm{Tr}(\cdot)$ to denote the trace of a matrix. In this context, the Hessian entries are functions receiving a dataset as input. The presentation thus uses $(\boldsymbol{H}P_i)_{j,l}(\boldsymbol{x})$ and $\mathrm{Tr}(\boldsymbol{H}P_i)(\boldsymbol{x})$ to denote the application of the second partial derivatives of P_i and of the *Hessian trace function* on dataset $\boldsymbol{x}$.

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 $\boldsymbol{x} \in \mathcal{X}$,*

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

The above shed light on the relationship between fairness and the difference in the local curvatures of problem P on any pairs of entities. As long as this local curvature is constant across all entities, then the difference in the bias induced by the noise onto the decision problem of any two entities can be bounded, and so can the (loss of) fairness.

The following corollaries illustrate the restrictions on the problem structure needed to satisfy fairness.

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 $\boldsymbol{x}$,*

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

5.2 Fair Decision Rules: Characterization

The next results bound the fairness violations of a class of indicator functions, called *thresholding functions*, and discusses the loss of fairness caused by the *composition of boolean predicates*, two recurrent features in decision rules. The fairness definition adopted uses the concept of absolute bias, in place of bias in Definition 3. Indeed, the absolute bias $|B^i_P|$ corresponds to the classification error for (binary) decision rules of P_i, i.e., $\Pr[P_i(\tilde{\boldsymbol{x}}) \neq P_i(\boldsymbol{x})]$. The results also assume $\mathcal{M}$ to be a non-trivial mechanism, i.e., $|B^i_P(\mathcal{M}, \boldsymbol{x})| < 0.5 \; \forall i \in [n]$. Note that this is a non-restrictive condition, since the focus of data-release mechanisms is to preserve the quality of the original inputs, and the mechanisms considered in this paper (and in the DP-literature, in general) all satisfy this assumption.

Theorem 2. *Consider a decision rule $P_i(\boldsymbol{x}) = \mathbf{1}\{x_i \geq \ell\}$ for some real value ℓ. Then, mechanism $\mathcal{M}$ is 0.5-fair w.r.t. P_i.*

This is a worst-case result and the mechanism may enjoy a better bound for specific datasets and decision rules. It is however significant since thresholding functions are ubiquitous in decision making over census data.

The next results focus on the composition of Boolean predicates under logical operators. The results are given under the assumption that mechanism $\mathcal{M}$ adds independent noise to the inputs of the predicates P_1 and P_2 to be composed, which is often the case. This assumption for P_1 and P_2 is denoted by $P^1 \perp\!\!\!\perp P^2$. Future work will aim at generalizing this results to broader assumptions.

Theorem 3. *Consider predicates P^1 and P^2 such that $P^1 \perp\!\!\!\perp P^2$ and assume that mechanism $\mathcal{M}$ is α_k-fair w.r.t. P^k ($k \in \{1,2\}$). Then $\mathcal{M}$ is α-fair for predicates $P^1 \vee P^2$ and $P^1 \wedge P^2$ with*

$$\alpha = \left(\alpha_1 + \underline{B}^1 + \alpha_2 + \underline{B}^2 - (\alpha_1 + \underline{B}^1)(\alpha_2 + \underline{B}^2) - \underline{B}^1\,\underline{B}^2\right),$$

where $\overline{B}^k$ and $\underline{B}^k$ are the maximum and minimum absolute biases for $\mathcal{M}$ w.r.t. P^k (for $k = \{1,2\}$).

The result above bounds the fairness violation derived by the composition of Boolean predicates under logical operators.

The extended version of this work also include a surprising, positive compositional fairness result regarding predicates composed under a XOR operator [Fioretto *et al.*, 2021].

6 The Nature of Bias

The previous section characterized conditions bounding fairness violations. In contrast, this section analyzes the reasons for disparity errors arising in the motivating problems.

6.1 The Problem Structure

The first result is an important corollary of Theorem 1. It studies which restrictions on the structure of problem P are needed to satisfy fairness. Once again, P is assumed to be at least twice differentiable.

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 $\mathrm{Tr}(\boldsymbol{H}P_i)(\boldsymbol{x}) \neq \mathrm{Tr}(\boldsymbol{H}P_j)(\boldsymbol{x})$ 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

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

is not constant with respect to its inputs. Thus, any two entries i, j whose $x_i \neq x_j$ imply $\mathrm{Tr}(\boldsymbol{H}P^F_i) \neq \mathrm{Tr}(\boldsymbol{H}P^F_j)$. As illustrated in Figure 2, 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 $\mathrm{Tr}(\boldsymbol{H}P^F_i)$.

Ratio Functions The next result considers *ratio functions* of the form $P_i(\langle x, y\rangle) = x/y$ with $x, y \in \mathbb{R}$ and $x \leq y$, which occur in the Minority language voting right benefits problem P^M_i. In the following $\mathcal{M}$ is the Laplace mechanism.

Corollary 4. *Mechanism $\mathcal{M}$ is not fair w.r.t. $P_i(\langle x, y\rangle) = x/y$ and inputs x, y.*

Figure 4 (left) provides an illustration linked to problem P^M. It shows the original values x^{sp}/x^s (blue circles) and the expected values of the privacy-preserving counterparts (red crosses) of three counties; from left to right: *Loving county, TX*, where $x^{sp}/x^s = 4/80 = 0.05$, *Terrell county, TX*, where $x^{sp}/x^s = 30/600 = 0.05$, and *Union county, NM*, where $x^{sp}/x^s = 160/3305 = 0.0484$. The length of the gray vertical line represents the absolute bias and the dotted line marks a threshold value (0.05) associated with the formula P^M_i. While the three counties have (almost) identical ratios values, they induce significant differences in absolute bias. This is due to the difference in scale of the numerator (and denominator), with smaller numerators inducing higher bias.

Thresholding Functions As discussed in Theorem 2, discontinuities caused by indicator functions, including thresholding, may induce unfairness. This is showcased in Figure 4 (center) which describes the same setting depicted in Figure 4 (left) but with the red line indicating the variance of the noisy ratios. Notice the significant differences in error variances, with Loving county exhibiting the largest variance. This aspect is also shown in Figure 3 where the counties with ratios lying near the threshold value have higher decisions errors than those whose ratios lies far from it.

6.2 Predicates Composition

The next result highlights the negative impact coming from the composition of Boolean predicates. The following important result is corollary of Theorem 3 and provides a lower bound on the fairness bound.

Corollary 5. *Let mechanism $\mathcal{M}$ be α_k-fair w.r.t. to problem P^k ($k \in \{1,2\}$). Then $\mathcal{M}$ is α-fair w.r.t. problems $P = P^1 \vee P^2$ and $P = P^1 \wedge P^2$, with $\alpha > \max(\alpha_1, \alpha_2)$.*

Figure 4 (right) illustrates Corollary 5. It once again uses the minority language problem P^M. In the figure, each dot represents the absolute bias $|B^i_{P^M}(\mathcal{M}, \boldsymbol{x})|$ associated with a selected county. Red and blue circles illustrate the absolute bias introduced by mechanism $\mathcal{M}$ for problem $P^1(x^{sp}) = \mathbf{1}\{x^{sp} \geq 10^4\}$ and $P^2(x^{sp}, x^{spe}) = \mathbf{1}\{\frac{x^{spe}}{x^{sp}} > 0.0131\}$ respectively. The selected counties have all similar and small absolute bias on the two predicates P^1 and P^2. However, when they are combined using logical connector $\vee$, the resulting absolute bias increases substantially, as illustrated by the associated green circles.

The extended version [Fioretto *et al.*, 2021] also analyzes an interesting difference in errors based on the Truth values of the composing predicates P^1 and P^2, and shows that the highest error is achieved when they both are True for $\wedge$ and when they both are False for $\vee$ connectors. This may have strong implications in classification tasks.

6.3 Post-Processing

The final analysis of bias relates to the effect of post-processing the output of the differentially private data release. In particular, the section focuses on ensuring non-negativity of the released data. The discussion focuses on problem P^F but the results are, once again, general.

The section first presents a *negative result*: the application of post-processing operator $\pi_{\geq \ell}(z) \stackrel{\text{def}}{=} \max(\ell, z)$ to ensure that the result is at least ℓ induces a positive bias which, in turn, can exacerbate the disparity error of the allotment problem.

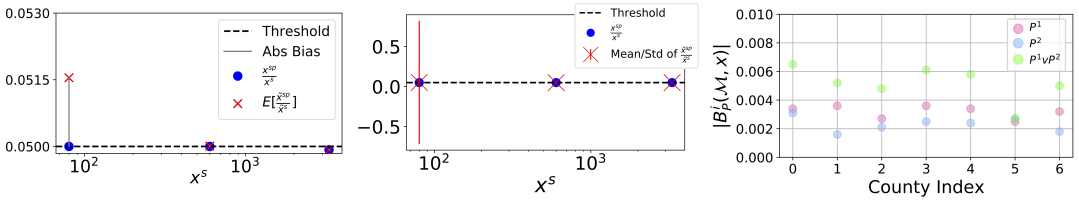

Figure 4: Unfairness effect in *ratios* (left), *thresholding* (middle) and predicates disjunction (right)

Theorem 4. *Let $\tilde{x} = x + Lap(\lambda)$, with scale $\lambda > 0$, and $\hat{x} = \pi_{\geq \ell}(\tilde{x})$, with $\ell < x$, be its post-processed value. Then,*

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

Lemma 4 indicates the presence of positive bias of post-processed Laplace random variable when ensuring non-negativity, and that such bias is $B^i(\mathcal{M}, \boldsymbol{x}) = \mathbb{E}[\hat{x}_i] - x_i = \exp(\frac{\ell - x_i}{\lambda}) \leq \lambda/2$ for $\ell \leq x_i$. As shown in Figure 2 the effect of this bias has a negative impact on the final disparity of the allotment problem, where smaller entities have the largest bias (in the Figure $\ell = 0$).

Discussion The results highlighted in this section are both surprising and significant. They show that *the motivating allotment problems and decision rules induce inherent unfairness when given as input differentially private data*. This is remarkable since the resulting decisions have significant societal, economic, and political impact on the involved individuals: federal funds, vaccines, and therapeutics may be unfairly allocated, minority language voters may be disenfranchised, and congressional apportionment may not be fairly reflected. The next section identifies a set of guidelines to mitigate these negative effects.

7 Mitigating Solutions

7.1 The Output Perturbation Approach

This section proposes three guidelines that may be adopted to mitigate the unfairness effects presented in the paper, with focus on the motivating allotments problems and decision rules.

A simple approach to mitigate the fairness issues discussed is to recur to *output perturbation* to randomize the outputs of problem P_i, rather than its inputs, using an unbiased mechanism. Injecting noise directly after the computation of the outputs $P_i(\boldsymbol{x})$, ensures that the result will be unbiased. However, this approach has two shortcomings. First, it is not applicable to the context studied in this paper, where a data agency desires to release a privacy-preserving dataset $\tilde{\boldsymbol{x}}$ that may be used for various decision problems. Second, computing the sensitivity of the problem P_i may be hard, it may require to use a conservative estimate, or may even be impossible, if the problem has unbounded range. A conservative sensitivity implies the introduction of significant loss in accuracy, which may render the decisions unusable in practice.

7.2 Linearization by Redundant Releases

A different approach considers modifying on the decision problem P_i itself. Many decision rules and allotment problems are designed in an ad-hoc manner to satisfy some property on the original data, e.g., about the percentage of population required to have a certain level of education. Motivated by Corollaries 1 and 2, this section proposes guidelines to modify the original problem P_i with the goal of reducing the unfairness effects introduced by differential privacy.

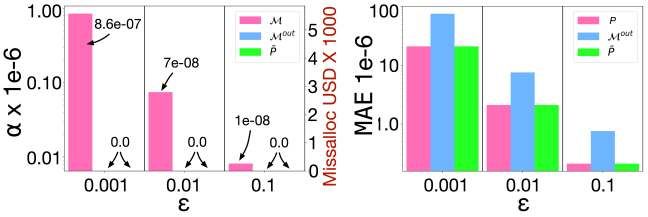

Figure 5: Linearization by redundant release: Fairness and errors.

The idea is to use a linearized version $\bar{P}_i$ of problem P_i. While many linearizion techniques exists [Rebennack and Krasko, 2020], and are often problem specific, the section focuses on a linear proxy $\bar{P}_i^F$ to problem P_i^F that can be obtained by enforcing a redundant data release. While the discussion focuses on problem P_i^F, the guideline is general and applies to any allotment problem with similar structure.

Let $Z = \sum_i a_i x_i$. Problem $P_i^F(\boldsymbol{x}) = a_i x_i / Z$ is linear w.r.t. the inputs x_i but non-linear w.r.t. Z. However, releasing Z, in addition to releasing the privacy-preserving values $\tilde{\boldsymbol{x}}$, would render Z a constant rather than a problem input to P^F. To do so, Z can either be released publicly, at cost of a (typically small) privacy leakage or by perturbing it with fixed noise. The resulting linear proxy allocation problem $\bar{P}_i^F$ is thus linear in the inputs $\boldsymbol{x}$.

Figure 5 illustrates this approach in practice. The left plot shows the fairness bound α and the right plot shows the empirical mean absolute error $\frac{1}{m}\sum_{k=1}^{m} |P_i(\boldsymbol{x}^k) - P_i(\tilde{\boldsymbol{x}}^m)|$, obtained using $m = 10^4$ repetitions, when the DP data $\tilde{\boldsymbol{x}}$ is applied to (1) the original problem P, (2) its linear proxy $\bar{P}$, and (3) when output perturbation (denoted $\mathcal{M}^{\text{out}}$) is adopted. The number on top of each bar reports the fairness bounds, and emphasize that the proposed remedy solutions achieve perfect-fairness. Notice that the proposed linear proxy solution can reduce the fairness violation dramatically while retaining similar errors. While the output perturbation method reduces the disparity error, it also incurs significant errors that make the approach rarely usable in practice. The extended version [Fioretto *et al.*, 2021] also discuss a solution based on a piecewise linear proxy function for the more complex decision rule P^M.

It is important to note that the experiments above use a

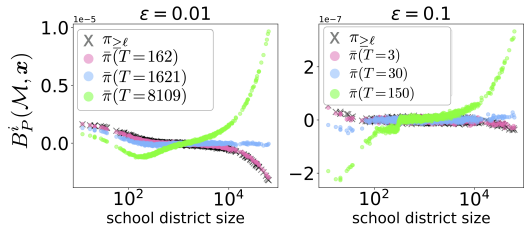

Figure 6: Modified post-processing: Unfairness reduction.

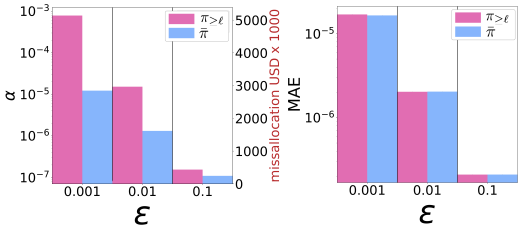

Figure 7: Modified post-processing on problem P^F.

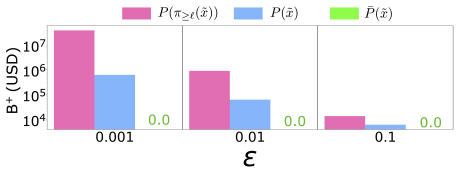

Figure 8: Cost of privacy on problem P^F.

data release mechanism $\mathcal{M}$ that applies no post-processing. A discussion about the mitigating solutions for the bias effects caused by post-processing is presented next.

7.3 Modified Post-Processing

This section introduces a simple, yet effective, solution to mitigate the negative fairness impact of the non-negative post-processing. The proposed solution operates in 3 steps: It first (1) performs a non-negative post-processing of the privacy-preserving input $\tilde{x}$ to obtain value $\bar{x} = \pi_{\geq\ell}(\tilde{x})$. Next, (2) it computes $\bar{x}_T = \bar{x} - \frac{T}{\bar{x}+1-\ell}$. Its goal is to correct the error introduced by the post-processing operator, which is especially large for quantities near the boundary ℓ. Here T is a *temperature* parameter that controls the strengths of the correction. This step reduces the value $\bar{x}$ by quantity $\frac{T}{\bar{x}+1-\ell}$. The effect of this operation is to reduce the expected value $\mathbb{E}[\bar{x}]$ by larger (smaller) amounts as x get closer (farther) to the boundary value ℓ. Finally, (3) it ensures that the final estimate is indeed lower bounded by ℓ, by computing $\hat{x} = \max(\bar{x}_T, \ell)$.

The benefits of this approach, called $\bar{\pi}$, are illustrated in Figure 6, which show the absolute bias $|B^i_{P^F}|$ for the Title 1 fund allocation problem that is induced by the original mechanism $\mathcal{M}$ with standard post-processing $\pi_{\geq 0}$ and by the proposed modified post-processing for different temperature values T. The figure illustrates the role of the temperature T in the disparity errors. Small values T may have small impacts in reducing the disparity errors, while large T values can introduce errors, thus may exacerbate unfairness. The optimal choice for T can be found by solving the following:

$$T^* = \operatorname*{argmin}_T \left(\max_{\boldsymbol{x}\geq\ell} |\mathbb{E}[\hat{\boldsymbol{x}}_T] - \boldsymbol{x}| - \min_{\boldsymbol{x}\geq\ell} |\mathbb{E}[\hat{\boldsymbol{x}}_T] - \boldsymbol{x}| \right), \quad (3)$$

where $\hat{\boldsymbol{x}}_T$ is a random variable obtained by the proposed 3 step solution, with temperature T. The expected value of $\hat{\boldsymbol{x}}$ can be approximated via sampling. Note that naively finding the optimal T may require access to the true data. Solving the problem above in a privacy-preserving way is beyond the scope of the paper and the subject of future work. The reductions in the fairness bound α for problem P^F are reported in Figure 7 (left), while Figure 7 (right) shows that this method has no perceptible impact on the mean absolute error.

7.4 Fairness Payment

Finally, this section focuses on allotment problems, like P^F, that distribute a budget B among n entities, and where the allotment for entity i represents the fraction of budget B it expects. Differential privacy typically implements a postprocessing step to renormalize the fractions so that they sum to 1. This normalization, together with nonnegativity constraints, introduces a bias and hence more unfairness. One way to alleviate this problem is to increase the total budget B, and avoiding the normalization. This section quantifies the cost of doing so: it defines the *cost of privacy*, which is the increase in budget B^+ required to achieve this goal.

Definition 4 (Cost of Privacy)**.** *Given problem P, that distributes budget B among n entities, data release mechanism $\mathcal{M}$, and dataset $\boldsymbol{x}$, the cost of privacy is:*

$$B^+ = \textstyle\sum_{i\in I^-} |B^i_P(\mathcal{M}, \boldsymbol{x})| \times B$$

with $I^- = \{i : B^i_P(\mathcal{M}, \boldsymbol{x}) < 0\}$.

Figure 8 illustrates the cost of privacy, in USD, required to render each county in the state of New York not negatively penalized by the effects of differential privacy. The figure shows, in decreasing order, the different costs associated with a mechanism $P^F(\pi_{\geq 0}(\tilde{\boldsymbol{x}}))$ that applies a post-processing step, one $P^F(\tilde{\boldsymbol{x}})$ that does not apply post-processing, and one that uses a linear proxy problem $\bar{P}^F(\tilde{\boldsymbol{x}})$.

8 Conclusions

This paper analyzed the disparity arising in decisions granting benefits or privileges to groups of people when these decisions are made adopting differentially private statistics about these groups. It first characterized the conditions for which allotment problems achieve finite fairness violations and bound the fairness violations induced by important components of decision rules, including reasoning about the composition of Boolean predicates under logical operators. Then, the paper analyzed the reasons for disparity errors arising in the motivating problems and recognized the problem structure, the predicate composition, and the mechanism post-processing, as paramount to the bias and unfairness contribution. Finally, it suggested guidelines to act on the decision problems or on the mechanism (i.e., via modified post-processing steps) to mitigate the unfairness issues. The analysis provided in this paper may provide useful guidelines for policy-makers and data agencies for testing the fairness and bias impacts of privacy-preserving decision making.

References

[Abowd, 2018] John M Abowd. The us census bureau adopts differential privacy. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 2867–2867, 2018.

[Dwork and Roth, 2013] Cynthia Dwork and Aaron Roth. The algorithmic foundations of differential privacy. *Theoretical Computer Science*, 9(3-4):211–407, 2013.

[Dwork *et al.*, 2006] 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.

[Fioretto *et al.*, 2021] Ferdinando Fioretto, Cuong Tran, and Pascal Van Hentenryck. Decision making with differential privacy under the fairness lens. *CoRR*, abs/2105.07513, 2021.

[Pujol *et al.*, 2020] David Pujol, Ryan McKenna, Satya Kuppam, Michael Hay, Ashwin Machanavajjhala, and Gerome Miklau. Fair decision making using privacy-protected data. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, FAT* '20, page 189–199, New York, NY, USA, 2020. Association for Computing Machinery.

[Rebennack and Krasko, 2020] Steffen Rebennack and Vitaliy Krasko. Piecewise linear function fitting via mixed-integer linear programming. *INFORMS Journal on Computing*, 32(2):507–530, 2020.

[Sonnenberg, 2016] W. Sonnenberg. Allocating grants for title i. *U.S. Department of Education, Institute for Education Science*, 2016.

[Zhu *et al.*, 2021] Keyu Zhu, Pascal Van Hentenryck, and Ferdinando Fioretto. Bias and variance of post-processing in differential privacy. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, page (to appear), 2021.