Bias and Variance of Post-processing in Differential Privacy

Keyu Zhu[1], **Pascal Van Hentenryck**[1], **Ferdinando Fioretto**[2]
[1] Georgia Institute of Technology,
[2] Syracuse University
keyu.zhu@gatech.edu, pvh@isye.gatech.edu, ffiorett@syr.edu

Abstract

Post-processing immunity is a fundamental property of differential privacy: it enables the application of arbitrary data-independent transformations to the results of differentially private outputs without affecting their privacy guarantees. When query outputs must satisfy domain constraints, post-processing can be used to project the privacy-preserving outputs onto the feasible region. Moreover, when the feasible region is convex, a widely adopted class of post-processing steps is also guaranteed to improve accuracy. Post-processing has been applied successfully in many applications including census data-release, energy systems, and mobility. However, its effects on the noise distribution is poorly understood: It is often argued that post-processing may introduce bias and increase variance. This paper takes a first step towards understanding the properties of post-processing. It considers the release of census data and examines, both theoretically and empirically, the behavior of a widely adopted class of post-processing functions.

Introduction

Data sets and statistics about groups of individuals are increasingly collected and released, feeding many optimization and learning algorithms. 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 (U.S. Census Bureau 2020), which requires that no individual be identified from any data release by the Census Bureau. In Europe, data releases are regulated according to the General Data Protection Regulation (EU 2016), which addresses the control and transfer of personal data. Hence statistical agencies release privacy-preserving data and statistics that conform to these requirements.

Differential Privacy (Dwork et al. 2006) is of particular interest to meet this goal. Differential privacy is a formal privacy definition that bounds the disclosure risk of any individual participating in a computation. It is considered the de-facto standard for privacy protection and has been adopted by various corporations (Erlingsson, Pihur, and Korolova 2014; Team 2017) and governmental agencies (Abowd 2018).

On data-release tasks, differentially private algorithms, typically, inject carefully calibrated noise to the data before release. However, whereas this process guarantees privacy, it also affects the fidelity of the released data. In particular, the injected noise often produces data sets that violate consistency constraints of the application domain. For example, in census statistics, the number of people satisfying a property must be consistent in a geographical hierarchy, e.g., at the national, state, and county levels. The injection of independent noise, however, cannot ensure the consistency of these constraints.

To overcome this limitation, the differentially private outputs can be *post-processed* via data-independent functions that transform the noisy data to render it consistent with the domain constraints. The post-processing step is guaranteed to retain differential privacy. Moreover, when the feasible region is convex, a largely adopted class of post-processing functions, called *projections*, is guaranteed to improve accuracy. Post-processing has been applied successfully in many applications, including census data (Abowd 2018), energy systems (Fioretto, Mak, and Van Hentenryck 2020), and mobility (He et al. 2015). However, the effect of post-processing on the noise distribution is poorly understood: It is often argued that it may introduce bias and/or increase variance. Figure 1 illustrates this aspect on a census data-release problem, described later in the paper. It depicts the distribution of the Laplacian residual $\tilde{\boldsymbol{x}} - \boldsymbol{x}$, where $\boldsymbol{x}$ denotes the true data and $\tilde{\boldsymbol{x}}$ the noisy data, obtained by applying Laplacian noise to $\boldsymbol{x}$, as well as the post-processed residual $\hat{\boldsymbol{x}} - \boldsymbol{x}$, where $\hat{\boldsymbol{x}}$ is the projection of $\tilde{\boldsymbol{x}}$ onto the feasible region. The results are shown for two counties, and, as can be seen, the post-processing introduces significant bias on their associated privacy-preserving data.

The key contribution of this paper is to take a first step towards understanding the properties of post-processing. Motivated by census applications, it studies the behavior of two widely adopted classes of post-processing functions, called *projections*, for domains where the feasibility space is specified by linear equations. The two classes differ by the presence of non-negativity constraints. The paper shows that, when non-negativity constraints are absent, the projection does not introduce bias. When projections include non-negativity constraints, the paper presents an upper bound on the bias, which provides some insights on the type of prob-

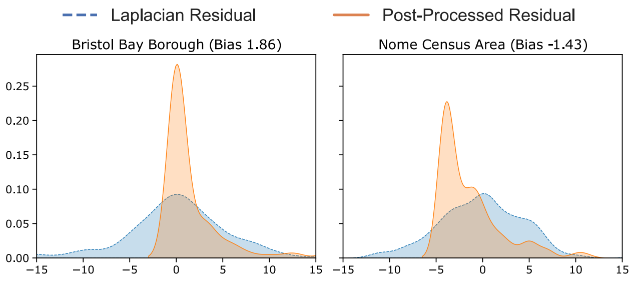

Figure 1: Bias of Post-Processing on the Census Problem.

lems for which the bias will be significant. Finally, the paper provides a detailed analysis of an important sub-problem used to satisfy hierarchical constraints in data-release tasks: It fully characterizes the residual distribution of the post-processed data, shows that it converges towards the Laplace distribution, and shed some interesting light on the effect of projections on the variance of the post-processed data, which may have strong implications with respect to group fairness.

While the paper discusses results focusing on differential privacy mechanisms that use Laplace or Geometric noise, the presented results generalize to other symmetric distributions.

Related Work

The adoption of post-processing to ensure that differentially private output satisfy some property of interest is commonly adopted in the privacy literature. Important contributions include the *hierarchical mechanism* of Hay et al. (2010) and its extensions (Qardaji, Yang, and Li 2013; Cormode et al. 2012a), which uses a post-processing step that enforces additive constraints based on a tree structure of the data universe to answer count queries over ranges. Other methods have incorporated a partitioning scheme to the data-release problem to increase the accuracy of the privacy-preserving data by cleverly splitting the privacy budget in different hierarchical levels (Xiao, Xiong, and Yuan 2010; Cormode et al. 2012b; Zhang, Xiao, and Xie 2016).

These post-processing algorithms have been used to release privacy preserving data sets for a wide array of applications, including transportation (Fioretto, Lee, and Van Hentenryck 2018), location privacy (He et al. 2015), and energy optimization (Fioretto, Mak, and Van Hentenryck 2020). Of particular interest is the TopDown algorithm (Abowd 2018), used by the US Census for the 2018 end-to-end test in preparation for the 2020 release. The algorithm is based on post-processing to satisfy consistency of hierarchical counts.

Preliminaries: Differential Privacy

Differential privacy (DP) (Dwork et al. 2006) is a rigorous privacy notion used to protect disclosures of an individual's data in a computation. Informally, it states that the probability of any differentially private output does not change much when a single individual data is added or removed to the data set, limiting the amount of information that the output reveals about any individual.

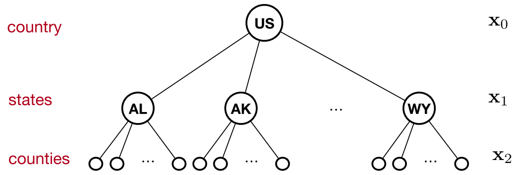

Figure 2: Example of hierarchical data set.

Definition 1 (Differential Privacy)**.** *A randomized mechanism $\mathcal{M} : \mathcal{D} \to \mathcal{R}$ with domain $\mathcal{D}$ and range $\mathcal{R}$ is ϵ-differentially private if, for any output $O \subseteq \mathcal{R}$ and data sets $D, D' \in \mathcal{D}$ differing by at most one entry (written $D \sim D'$),*

$$\Pr[\mathcal{M}(D) \in O] \leq \exp(\epsilon)\, Pr[\mathcal{M}(D') \in O]. \qquad (1)$$

The parameter $\epsilon \geq 0$ is the *privacy loss* of the mechanism, with values close to 0 denoting strong privacy.

An important differential privacy property is its *immunity to post-processing*, stating that a differentially private output can be arbitrarily transformed, using some data-independent function, without impacting its privacy guarantees.

Theorem 2 (Post-Processing (Dwork et al. 2006))**.** *Let $\mathcal{M}$ be an ϵ-differentially private mechanism and g be an arbitrary mapping from the set of possible outputs to an arbitrary set. Then, $g \circ \mathcal{M}$ is ϵ-differentially private.*

A function f (also called *query*) from a data set $D \in \mathcal{D}$ to a result set $R \subseteq \mathbb{R}^n$ can be made differentially private by injecting random noise to its output. The amount of noise depends on the *global sensitivity* of the query, denoted by Δ_f and defined as $\Delta_f = \max_{D \sim D'} \|f(D) - f(D')\|_1$.

The Laplace distribution with 0 mean and scale λ, denoted by $\mathrm{Lap}(\lambda)$, has a probability density function $\mathrm{Lap}(x|\lambda) = \frac{1}{2\lambda}e^{-\frac{|x|}{\lambda}}$. It can be used to obtain an ϵ-differentially private algorithm to answer numeric queries (Dwork et al. 2006). In the following, $\mathrm{Lap}(\lambda)^n$ denotes the i.i.d. Laplace distribution with 0 mean and scale λ over n dimensions.

Theorem 3 (Laplace Mechanism)**.** *Let $f : \mathcal{D} \to \mathbb{R}^n$ be a numeric query. The Laplace mechanism that outputs $f(D) + \boldsymbol{\eta}$, where $\boldsymbol{\eta} \in \mathbb{R}^n$ is drawn from the Laplace distribution $\mathrm{Lap}(^{\Delta_f}/_{\epsilon})^n$, achieves ϵ-differential privacy.*

Settings and Goal

The paper uses the following notation: boldface symbols denote vectors while italic symbols are used to denote scalars or random variables. The paper considers data sets of the form $\boldsymbol{x} \in \mathbb{R}^n$, where each element x_i of $\boldsymbol{x}$ is a real-valued quantity describing, for example, the number of individuals living in a geographical region. To produce ϵ-differentially private outputs, this work adopts the Laplace mechanism which, for an appropriately chosen λ, produces a new privacy-preserving data set $\tilde{\boldsymbol{x}} = \boldsymbol{x} + \mathrm{Lap}(\lambda)^n$. However, all results presented in this paper generalizes to other symmetric distributions as discussed later.

The original data $\boldsymbol{x}$ is assumed to satisfy a set of data independent constraints $\mathcal{K}$. This paper focuses on the case where $\mathcal{K}$ is a set of linear constraints which, as mentioned in the introduction, arise in a widespread number of applications (Fioretto and Van Hentenryck 2019; Abowd 2018; He et al. 2015; Fioretto, Mak, and Van Hentenryck 2020). Of particular relevance to this work are *hierarchical data release problems*, as those faced by the US Census Bureau. Consider the illustration in Figure 2. The tree depicts the hierarchy of the US territories, partitioned in states and counties. Each node is associated with a value representing the number of individuals living in the corresponding territory. The constraint set $\mathcal{K}$ then specifies that the value of a node is the sum of the values of its children and that all values are non-negative.

Due to the use of independent noise, the differentially private version $\tilde{\boldsymbol{x}}$ of $\boldsymbol{x}$, may not satisfy the original constraints. This scenario happens with very high probability in the hierarchical data-release problem considered. The paper, thus, focuses on mechanisms that generates outputs $\hat{\boldsymbol{x}}$ that satisfy two properties: (1) they guarantee ϵ-differential privacy, and (2) $\hat{\boldsymbol{x}}$ satisfy the constraints in $\mathcal{K}$.

Projection Operators

To meet these two objectives, the paper studies an important class of post-processing operators, called *projections*, that transform released data $\tilde{\boldsymbol{x}}$ to satisfy the constraints in $\mathcal{K}$. This paper focuses on the following two projections:

$$\begin{aligned} \hat{\boldsymbol{x}} &:= \operatorname*{arg\,min}_{\boldsymbol{v}\in\mathcal{K}} \|\boldsymbol{v} - \tilde{\boldsymbol{x}}\|_2 \\ \mathcal{K} &= \{\boldsymbol{v} \in \mathbb{R}^n \mid A\boldsymbol{v} = \boldsymbol{b}\} \end{aligned} \tag{P}$$

and,

$$\begin{aligned} \hat{\boldsymbol{x}}_+ &:= \operatorname*{arg\,min}_{\boldsymbol{v}\in\mathcal{K}} \|\boldsymbol{v} - \tilde{\boldsymbol{x}}\|_2 \\ \mathcal{K} &= \{\boldsymbol{v} \in \mathbb{R}^n \mid A\boldsymbol{v} = \boldsymbol{b}; \boldsymbol{v} \geq \boldsymbol{0}\}, \end{aligned} \tag{P_+}$$

where $\tilde{\boldsymbol{x}}$ is the privacy-preserving input to the projection operators, obtained by applying the Laplace mechanism to $\boldsymbol{x}$, A is an $m \times n$ matrix, and $\boldsymbol{b}$ is an m-dimensional vector. A and $\boldsymbol{b}$ are assumed to be public, non-sensitive information in this paper. By the post-processing immunity of differential privacy (Theorem 2) the projection operators (P) and (P_+) satisfy differential privacy. Both optimizations find a feasible solution that minimizes the l_2-distance to the noisy data $\tilde{\boldsymbol{x}}$. The existence and uniqueness of their solutions are guaranteed by convexity. These programs have been adopted by a vast array of applications. In particular, the census hierarchical data-release problem, analyzed in this paper as a case study, restores consistency of the hierarchical constraints using an instance of problem (P_+).

The theoretical results in this paper are illustrated using an empirical analysis from this census case study. For each instance associated with the true counts $\boldsymbol{x}$, the noise $\boldsymbol{\eta}$ is i.i.d. drawn from the double-sided geometric distribution $\boldsymbol{\eta} \sim \text{Geom}({}^{\Delta_f}/_{\epsilon})^n$, i.e., the discrete analog to the Laplace distribution. The results in this paper are generally presented for continuous distributions but they carry over naturally to this geometric distribution. The privacy budget ϵ is set to be 0.5 and the experiments perform 100 independent runs.

Analysis of Bias in Post-Processing

Bias of Program (P)

This section studies the bias induced by program (P), when the noisy data $\tilde{\boldsymbol{x}}$ is obtained by applying noise drawn from a symmetric probability distribution. Recall that a distribution with probability density function f is symmetric if there exists a value x_0 such that $f(x_0 - \delta) = f(x_0 + \delta)$ for all δ. This is the case of the Laplace and the double-sided geometric distributions. This section relies on the concept of a reflection operator.

Definition 4 (Reflection operator)**.** *The operator* $\text{Ref}_{\boldsymbol{v}}(\cdot)$ *is said to be a reflection operator across the vector* $\boldsymbol{v} \in \mathbb{R}^n$ *if, for any* $\boldsymbol{u} \in \mathbb{R}^n$*, the following identity holds:*

$$\text{Ref}_{\boldsymbol{v}}(\boldsymbol{u}) = 2\boldsymbol{v} - \boldsymbol{u}.$$

Lemma 5. *The reflection operator* $\text{Ref}_{\boldsymbol{x}}$ *and* $\hat{\boldsymbol{x}}$ *(as an operator) are commutative, i.e.,*

$$\text{Ref}_{\boldsymbol{x}}\left(\hat{\boldsymbol{x}}(\tilde{\boldsymbol{x}})\right) = \hat{\boldsymbol{x}}\left(\text{Ref}_{\boldsymbol{x}}(\tilde{\boldsymbol{x}})\right). \tag{2}$$

Proof. The right hand side of (2) is given by

$$\begin{aligned} \boldsymbol{x}' := \operatorname*{arg\,min}_{\boldsymbol{v}\in\mathbb{R}^n} &\ \|\boldsymbol{v} - \text{Ref}_{\boldsymbol{x}}(\tilde{\boldsymbol{x}})\|_2 \\ \text{s.t. } &A\boldsymbol{v} = \boldsymbol{b}, \end{aligned}$$

where $\boldsymbol{x}'$ is a shorthand for the solution $\hat{\boldsymbol{x}}(\text{Ref}_{\boldsymbol{x}}(\tilde{\boldsymbol{x}}))$. By reflection, $\text{Ref}_{\boldsymbol{x}}(\boldsymbol{x}')$ is a solution to the optimization problem:

$$\begin{aligned} \text{Ref}_{\boldsymbol{x}}(\boldsymbol{x}') = \operatorname*{arg\,min}_{\boldsymbol{v}\in\mathbb{R}^n} &\ \|\text{Ref}_{\boldsymbol{x}}(\boldsymbol{v}) - \text{Ref}_{\boldsymbol{x}}(\tilde{\boldsymbol{x}})\|_2 \\ \text{s.t. } &A\,\text{Ref}_{\boldsymbol{x}}(\boldsymbol{v}) = \boldsymbol{b}. \end{aligned}$$

By the definition of the reflection operator and the feasibility of the true data, we have that

$$\begin{aligned} \|\text{Ref}_{\boldsymbol{x}}(\boldsymbol{v}) - \text{Ref}_{\boldsymbol{x}}(\tilde{\boldsymbol{x}})\|_2 &= \|\boldsymbol{v} - \tilde{\boldsymbol{x}}\|_2\,, \\ A\boldsymbol{v} = A(2\boldsymbol{x} - \text{Ref}_{\boldsymbol{x}}(\boldsymbol{v})) &= 2\boldsymbol{b} - \boldsymbol{b} = \boldsymbol{b} \end{aligned} \tag{3}$$

and the previous optimization problem is equivalent to (P):

$$\begin{aligned} \text{Ref}_{\boldsymbol{x}}(\boldsymbol{x}') = \operatorname*{arg\,min}_{\boldsymbol{v}\in\mathbb{R}^n} &\ \|\boldsymbol{v} - \tilde{\boldsymbol{x}}\|_2 \\ \text{s.t. } &A\boldsymbol{v} = \boldsymbol{b}, \end{aligned}$$

because since (P) is convex, $\text{Ref}_{\boldsymbol{x}}(\boldsymbol{x}') = \hat{\boldsymbol{x}}(\tilde{\boldsymbol{x}})$. By applying the reflection operator on both sides, $\boldsymbol{x}' = \hat{\boldsymbol{x}}(\text{Ref}_{\boldsymbol{x}}(\tilde{\boldsymbol{x}})) = \text{Ref}_{\boldsymbol{x}}(\hat{\boldsymbol{x}}(\tilde{\boldsymbol{x}}))$. □

Figure 3 illustrates Lemma 5: It shows that the true data $\boldsymbol{x}$ is the midpoint between the post-processed solutions associated with the noisy data $\tilde{\boldsymbol{x}}$ and its reflection.

Let $\text{Err}\,(\boldsymbol{y}) = \boldsymbol{y} - \boldsymbol{x}$ be the entrywise difference between $\boldsymbol{y}$ and the true data.

Corollary 6. *The errors associated with the noisy data* $\tilde{\boldsymbol{x}}$ *and its reflection* $\text{Ref}_{\boldsymbol{x}}(\tilde{\boldsymbol{x}})$ *sums to* $\boldsymbol{0}$*, i.e.,*

$$\text{Err}\,(\hat{\boldsymbol{x}}(\tilde{\boldsymbol{x}})) + \text{Err}\,(\hat{\boldsymbol{x}}(\text{Ref}_{\boldsymbol{x}}(\tilde{\boldsymbol{x}}))) = \boldsymbol{0}.$$

Proof. By Lemma 5, $\text{Err}\,(\hat{\boldsymbol{x}}(\tilde{\boldsymbol{x}})) + \text{Err}\,(\hat{\boldsymbol{x}}(\text{Ref}_{\boldsymbol{x}}(\tilde{\boldsymbol{x}})))$

$$\begin{aligned} &= \text{Err}\,(\hat{\boldsymbol{x}}(\tilde{\boldsymbol{x}})) + \text{Err}\,(\text{Ref}_{\boldsymbol{x}}\,(\hat{\boldsymbol{x}}(\tilde{\boldsymbol{x}}))) \\ &= (\hat{\boldsymbol{x}}(\tilde{\boldsymbol{x}}) - \boldsymbol{x}) + (\text{Ref}_{\boldsymbol{x}}\,(\hat{\boldsymbol{x}}(\tilde{\boldsymbol{x}})) - \boldsymbol{x}) = \boldsymbol{0}. \end{aligned}$$

□

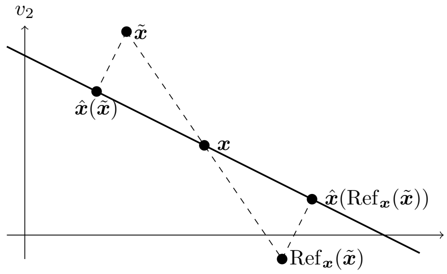
v_2
$\tilde{\boldsymbol{x}}$
$\hat{\boldsymbol{x}}(\tilde{\boldsymbol{x}})$
$\boldsymbol{x}$
$\hat{\boldsymbol{x}}(\mathrm{Ref}_{\boldsymbol{x}}(\tilde{\boldsymbol{x}}))$
v_1
$\mathrm{Ref}_{\boldsymbol{x}}(\tilde{\boldsymbol{x}})$

Figure 3: Illustration of Lemma 5 and Corollary 6.

The following theorem is a positive result: It shows that program (P) does not introduce bias.

Theorem 7. *Program* (P) *does not introduce bias, i.e.,*

$$\mathrm{Bias}\left[\hat{\boldsymbol{x}}(\tilde{\boldsymbol{x}})\right] := \mathbb{E}_{\tilde{\boldsymbol{x}}}\left[\mathrm{Err}\left(\hat{\boldsymbol{x}}(\tilde{\boldsymbol{x}})\right)\right] = \mathbf{0},$$

where the expectation is taken over the distribution of the noisy data $\tilde{\boldsymbol{x}}$*. In other words, the post-processed solution* $\hat{\boldsymbol{x}}$ *(as a random vector) is* unbiased*.*

Proof. Let $f_{\tilde{\boldsymbol{x}}}$ denote the probability density function of the noisy data $\tilde{\boldsymbol{x}}$, which is symmetric with respect to the true data $\boldsymbol{x}$. Then, the expectation of the resulting error is computed as follows.

$$\begin{aligned}
\mathbb{E}_{\tilde{\boldsymbol{x}}}\left[\mathrm{Err}\left(\hat{\boldsymbol{x}}(\tilde{\boldsymbol{x}})\right)\right] &= \int_{\boldsymbol{y}\in\mathbb{R}^n} \mathrm{Err}\left(\hat{\boldsymbol{x}}(\boldsymbol{y})\right)\cdot f_{\tilde{\boldsymbol{x}}}(\boldsymbol{y})d\boldsymbol{y}\\
&= \frac{1}{2}\int_{\boldsymbol{y}\in\mathbb{R}^n} \mathrm{Err}\left(\hat{\boldsymbol{x}}(\boldsymbol{y})\right)\cdot f_{\tilde{\boldsymbol{x}}}(\boldsymbol{y})d\boldsymbol{y} +\\
&\quad \frac{1}{2}\int_{\boldsymbol{y}\in\mathbb{R}^n} \mathrm{Err}\left(\hat{\boldsymbol{x}}(\mathrm{Ref}_{\boldsymbol{x}}(\boldsymbol{y}))\right)\cdot f_{\tilde{\boldsymbol{x}}}(\mathrm{Ref}_{\boldsymbol{x}}(\boldsymbol{y}))d\boldsymbol{y}\\
&= \frac{1}{2}\int_{\boldsymbol{y}\in\mathbb{R}^n} \left[\mathrm{Err}\left(\hat{\boldsymbol{x}}(\boldsymbol{y})\right) + \mathrm{Err}\left(\hat{\boldsymbol{x}}(\mathrm{Ref}_{\boldsymbol{x}}(\boldsymbol{y}))\right)\right]\cdot f_{\tilde{\boldsymbol{x}}}(\boldsymbol{y})d\boldsymbol{y} \quad (4)\\
&= \frac{1}{2}\int_{\boldsymbol{y}\in\mathbb{R}^n} \mathbf{0}\cdot f_{\tilde{\boldsymbol{x}}}(\boldsymbol{y})d\boldsymbol{y} \quad (5)\\
&= \mathbf{0},
\end{aligned}$$

where Equation (4) comes from the symmetric distribution of the noisy data $\tilde{\boldsymbol{x}}$, i.e., for any $\boldsymbol{y}\in\mathbb{R}^n$,

$$f_{\tilde{\boldsymbol{x}}}(\boldsymbol{y}) = f_{\tilde{\boldsymbol{x}}}(\boldsymbol{x}+(\boldsymbol{y}-\boldsymbol{x})) = f_{\tilde{\boldsymbol{x}}}(\boldsymbol{x}-(\boldsymbol{y}-\boldsymbol{x})) = f_{\tilde{\boldsymbol{x}}}(\mathrm{Ref}_{\boldsymbol{x}}(\boldsymbol{y})),$$

and Equation (5) is due to Corollary 6. □

Bias of Program (P_+)

Theorem 7 indicates that the bias in program (P_+) comes from the non-negativity constraints. The section bounds this bias by leveraging the insights of Theorem 7. It assumes that the feasible region $\mathcal{K}$ is bounded (which holds, in many practical settings, including the census data release case) and that the noisy data $\tilde{\boldsymbol{x}}$ is the output of the Laplace mechanism applied to the true data $\boldsymbol{x}$, i.e., $\tilde{\boldsymbol{x}} = \boldsymbol{x} + \mathrm{Lap}(\lambda)^n$. It will leverage the prior positive results by isolating a subset of the feasible space close under refection. The first lemma computes the probability that the Laplace mechanism produces an output in a ball of radius r. The proof is by induction on the dimension n.

Lemma 8. *Given a random vector* $\boldsymbol{\eta} = [\eta_1,\dots,\eta_n]$*, where* $\{\eta_i\}_{i\in[n]}$ *are i.i.d. random variables drawn from a Laplace distribution* $\mathrm{Lap}(\lambda)$ *(*$\lambda > 0$*), the following identity holds for any* $r \geq 0$*:*

$$\Pr\left(\boldsymbol{\eta}\in B_r\left(\mathbf{0}\right)\right) = 1 - \exp\left(\frac{-r}{\lambda}\right)\cdot\sum_{i=0}^{n-1}\frac{r^i}{i!\cdot\lambda^i}, \qquad (6)$$

where $B_r\left(\mathbf{0}\right)$ *is the* l_1*-ball of radius* r *centered at* $\mathbf{0}$*, i.e.,*

$$B_r\left(\mathbf{0}\right) = \left\{\boldsymbol{v}\in\mathbb{R}^n \mid \left\|\boldsymbol{v}\right\|_1 \leq r\right\}.$$

A similar result can be obtained for the double-sided geometric distribution. If the noisy data follows a $\mathrm{Geom}(a)^n$ distribution, then

$$\Pr\left(\boldsymbol{\eta}\in B_r\left(\mathbf{0}\right)\right) = 1 - \frac{2a^{r+1}}{1+a}\sum_{i=0}^{n-1} h_i(r)\cdot\left(\frac{1-a}{1+a}\right)^i,$$

where $\{h_i\}_{i\in\mathbb{N}}$ is a family of polynomials with $h_0(r) = 1$ and $h_{i+1}(r) = \sum_{v=-r}^{r} h_i(r-|v|)$ for any $i\in\mathbb{N}$. The rest of this section is presented in terms of the Laplace distribution but the results can be generalized to any distribution satisfying a version of Lemma 8.

Corollary 9. *Suppose that the noisy data* $\tilde{\boldsymbol{x}}$ *is the output of the Laplace mechanism, i.e.,* $\tilde{\boldsymbol{x}} = \boldsymbol{x} + \mathrm{Lap}(\lambda)^n$ *with* $\lambda > 0$*. Then, for any* $r \geq 0$*,*

$$\Pr\left(\tilde{\boldsymbol{x}}\in B_r\left(\boldsymbol{x}\right)\right) = 1 - \exp\left(\frac{-r}{\lambda}\right)\cdot\sum_{i=0}^{n-1}\frac{r^i}{i!\cdot\lambda^i}.$$

Proof. Let $\boldsymbol{\eta}$ denote the n-dimensional random vector of the Laplacian noise added to the true data $\boldsymbol{x}$, i.e., $\boldsymbol{\eta} = \tilde{\boldsymbol{x}} - \boldsymbol{x}$. By the definition of the Laplace mechanism, $\boldsymbol{\eta} = [\eta_1,\dots,\eta_n]$ consists of n i.i.d. components, each of which is drawn from the Laplace distribution $\mathrm{Lap}(\lambda)$. Then, by Lemma 8,

$$\begin{aligned}
\Pr(\boldsymbol{\eta}\in B_r\left(\mathbf{0}\right)) &= \Pr\left(\tilde{\boldsymbol{x}} - \boldsymbol{x}\in B_r\left(\mathbf{0}\right)\right)\\
&= 1 - \exp\left(\frac{-r}{\lambda}\right)\cdot\sum_{i=0}^{n-1}\frac{r^i}{i!\cdot\lambda^i}, \qquad \forall\, r \geq 0.
\end{aligned}$$

Since $\tilde{\boldsymbol{x}} - \boldsymbol{x}\in B_r\left(\mathbf{0}\right)$ iff $\tilde{\boldsymbol{x}}\in B_r\left(\boldsymbol{x}\right)$, for any $r \geq 0$,

$$\Pr\left(\tilde{\boldsymbol{x}}\in B_r\left(\boldsymbol{x}\right)\right) = 1 - \exp\left(\frac{-r}{\lambda}\right)\cdot\sum_{i=0}^{n-1}\frac{r^i}{i!\cdot\lambda^i}.$$

□

Let r_m be $\min_{i\in[n]}\boldsymbol{x}_i$. The next lemma states that $B_{r_m}\left(\boldsymbol{x}\right)$ is a feasible subspace where there is no bias.

Lemma 10. *For any noisy data* $\tilde{\boldsymbol{x}}\in B_{r_m}\left(\boldsymbol{x}\right)$*, the post-processed solution* $\hat{\boldsymbol{x}}$ *of program* (P) *is non-negative and equal to solution* $\hat{\boldsymbol{x}}_+$ *of program* (P_+)*.*

Proof. Since $\tilde{\boldsymbol{x}}$ belongs to the l_1-ball $B_{r_m}\left(\boldsymbol{x}\right)$, $\left\|\tilde{\boldsymbol{x}} - \boldsymbol{x}\right\|_2$ is also bounded from the above by r_m since

$$\left\|\tilde{\boldsymbol{x}} - \boldsymbol{x}\right\|_2 \leq \left\|\tilde{\boldsymbol{x}} - \boldsymbol{x}\right\|_1 \leq r_m.$$

By convexity of $\mathcal{K}$, $\left\|\hat{\boldsymbol{x}} - \boldsymbol{x}\right\|_\infty \leq \left\|\hat{\boldsymbol{x}} - \boldsymbol{x}\right\|_2 \leq r_m$. Moreover, $\hat{\boldsymbol{x}}$ is non-negative since its l_∞-distance to $\boldsymbol{x}$ is bounded by r_m and the result follows by optimality of $\hat{\boldsymbol{x}}_+$. □

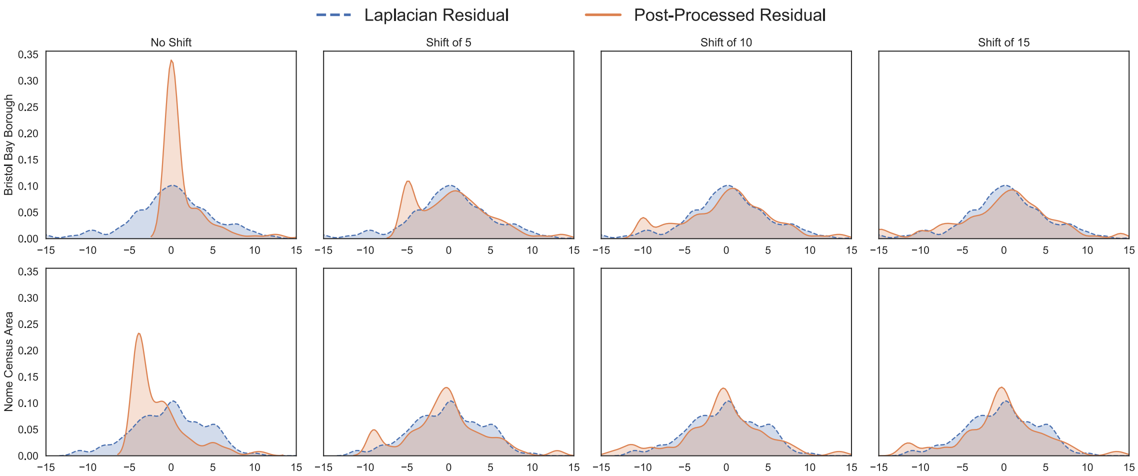

Figure 4: Bias of Post-Processing on the Census Problem as r_m Increases.

The next theorem is the main result of this section and it bounds the bias of program (P_+).

Theorem 11. *Suppose that the noisy data $\tilde{\boldsymbol{x}}$ is the output of the Laplace mechanism with scale λ. The bias of the post-processed solution $\hat{\boldsymbol{x}}_+$ of program (P_+) is bounded, in l_∞ norm, by*

$$\|\mathrm{Bias}\left[\hat{\boldsymbol{x}}_+(\tilde{\boldsymbol{x}})\right]\|_\infty = \|\mathbb{E}_{\tilde{\boldsymbol{x}}}\left[\mathrm{Err}\left(\hat{\boldsymbol{x}}_+(\tilde{\boldsymbol{x}})\right)\right]\|_\infty$$
$$\leq C' \cdot \exp\left(\frac{-r_m}{\lambda}\right) \cdot \sum_{i=0}^{n-1} \frac{(r_m)^i}{i! \cdot \lambda^i},$$

where C' represents the value $\sup_{\boldsymbol{v}\in\mathcal{K}} \|\boldsymbol{v}-\boldsymbol{x}\|_\infty$, which is finite due to the boundedness of the feasible region $\mathcal{K}$.

Proof.

$$\begin{aligned} &\mathrm{Bias}\left[\hat{\boldsymbol{x}}_+(\tilde{\boldsymbol{x}})\right] = \mathbb{E}_{\tilde{\boldsymbol{x}}}\left[\mathrm{Err}\left(\hat{\boldsymbol{x}}_+(\tilde{\boldsymbol{x}})\right)\right] \\ =\, &\mathbb{E}_{\tilde{\boldsymbol{x}}}\left[\mathrm{Err}\left(\hat{\boldsymbol{x}}_+(\tilde{\boldsymbol{x}})\right) \mid \tilde{\boldsymbol{x}} \in B_{r_m}(\boldsymbol{x})\right] \cdot \Pr\left(\tilde{\boldsymbol{x}} \in B_{r_m}(\boldsymbol{x})\right) + \\ &\mathbb{E}_{\tilde{\boldsymbol{x}}}\left[\mathrm{Err}\left(\hat{\boldsymbol{x}}_+(\tilde{\boldsymbol{x}})\right) \mid \tilde{\boldsymbol{x}} \notin B_{r_m}(\boldsymbol{x})\right] \cdot \Pr\left(\tilde{\boldsymbol{x}} \notin B_{r_m}(\boldsymbol{x})\right). \end{aligned}$$

By Lemma 10 and Theorem 7, the left-hand side of the sum is zero. As a result,

$$\begin{aligned} &\|\mathrm{Bias}\left[\hat{\boldsymbol{x}}_+(\tilde{\boldsymbol{x}})\right]\|_\infty \\ =\, &\|\mathbb{E}_{\tilde{\boldsymbol{x}}}\left[\mathrm{Err}\left(\hat{\boldsymbol{x}}_+(\tilde{\boldsymbol{x}})\right) \mid \tilde{\boldsymbol{x}} \notin B_{r_m}(\boldsymbol{x})\right]\|_\infty \cdot \Pr\left(\tilde{\boldsymbol{x}} \notin B_{r_m}(\boldsymbol{x})\right) \\ \leq\, &\mathbb{E}_{\tilde{\boldsymbol{x}}}\left[\|\mathrm{Err}\left(\hat{\boldsymbol{x}}_+(\tilde{\boldsymbol{x}})\right)\|_\infty \mid \tilde{\boldsymbol{x}} \notin B_{r_m}(\boldsymbol{x})\right] \cdot \Pr\left(\tilde{\boldsymbol{x}} \notin B_{r_m}(\boldsymbol{x})\right) \\ \leq\, &C' \cdot \Pr\left(\tilde{\boldsymbol{x}} \notin B_{r_m}(\boldsymbol{x})\right) \qquad (7) \\ =\, &C' \cdot \exp\left(\frac{-r_m}{\lambda}\right) \cdot \sum_{i=0}^{n-1} \frac{(r_m)^i}{i! \cdot \lambda^i}, \qquad (8) \end{aligned}$$

where (7) follows from the feasibility of $\hat{\boldsymbol{x}}_+(\tilde{\boldsymbol{x}})$ and

$$\begin{aligned} \|\mathrm{Err}\left(\hat{\boldsymbol{x}}_+(\tilde{\boldsymbol{x}})\right)\|_\infty &= \|\hat{\boldsymbol{x}}_+(\tilde{\boldsymbol{x}}) - \boldsymbol{x}\|_\infty \\ &\leq \sup_{\boldsymbol{v}\in\mathcal{K}} \|\boldsymbol{v}-\boldsymbol{x}\|_\infty = C', \end{aligned}$$

since the feasible region is bounded by hypothesis. Equation (8) follows from Corollary 9. □

Figure 4 illustrates Theorem 11. It reports the same residuals as in Figure 1 but with the true county counts increased by a positive shift factor. This increases the value of r_m and the bias progressively disappears as r_m grows. This observation can give insights to statistical agencies about what can be reported without introducing significant bias, informing their decisions on the granularity of the data releases.

To complement these results, the theoretical bound is also compared on the post-processing of New Mexico and its counties. The state has a population of 7,289,112, 33 counties, $r_m = 348$, and the experiment uses $\lambda = 5$. The theoretical bound is 0.29, while the empirical bias is 0.06. The results may not be as tight for larger states, since the bound depends on C', the maximum distance between the real data and a point in the feasible space.

Analysis of Fairness in Projections

This section provides a detailed analysis of the distribution of the post-processed noise for a special case of program (P) defined as follows:

$$\begin{aligned} \hat{\boldsymbol{x}}_S &:= \underset{\boldsymbol{v}\in\mathcal{K}}{\arg\min}\, \|\boldsymbol{v}-\tilde{\boldsymbol{x}}\|_2 \\ \mathcal{K} &= \left\{ \boldsymbol{v} \in \mathbb{R}^n \;\middle|\; \sum_{i=1}^{n} v_i = b \right\}, \end{aligned} \qquad (P_{\mathrm{S}})$$

where $b \in \mathbb{R}$ is a constant. This specific post-processing step (P_{S}) requires that the components of its output should be summed up to the constant b, which makes it broadly applicable. For instance, in the census context, program (P_{S}) makes sure that the state populations are compatible with the overall US population, which is viewed as public information. Similar post-processing steps take place at the state level. The section will reveal an interesting connection between the post-processing step and the census model itself.

The next theorem is a key result: it characterizes the marginal distribution of the post-processed noise $\hat{\boldsymbol{x}}_S - \boldsymbol{x}$, i.e., the distribution of $\mathrm{Err}\left(\hat{\boldsymbol{x}}_S\right)_i = \hat{\boldsymbol{x}}_{Si} - \boldsymbol{x}_i$ for any $i \in [n]$.

It is expressed in terms of the Laplace distribution but again generalizes to other distributions.

Theorem 12. *Let $\{\eta_i\}_{i\in[n]}$ be n i.i.d. random variables drawn from a Laplace distribution* $\mathrm{Lap}(\lambda)$. *The marginal error of the post-processed solution $\hat{\boldsymbol{x}}_S$ of program* (P_{S}) *follows the distribution:*

$$\mathrm{Err}\left(\hat{\boldsymbol{x}}_S\right)_i = \hat{\boldsymbol{x}}_{Si} - \boldsymbol{x}_i \sim \frac{(n-1)\eta_i - \sum_{j\neq i}\eta_j}{n} \quad \forall\, i \in [n],$$

with variance

$$\begin{aligned}\mathrm{Var}\left(\mathrm{Err}\left(\hat{\boldsymbol{x}}_S\right)_i\right) &= \left(1-\frac{1}{n}\right)\cdot \mathrm{Var}\left(\mathrm{Lap}(\lambda)\right)\\ &= 2\lambda^2\left(1-\frac{1}{n}\right) \qquad\qquad \forall\, i\in[n].\end{aligned}$$

Proof. (Sketch) Without the loss of generality, the proof considers $\mathrm{Err}\left(\hat{\boldsymbol{x}}_S\right)_1$. Let $\{\boldsymbol{e}_i\}_{i\in[n]}$ be the standard basis of the n-dimensional space $\mathbb{R}^n$ such that the noise $\boldsymbol{\eta}$ added to the true data $\boldsymbol{x}$ can be represented as $\boldsymbol{\eta} = \sum_{i=1}^{n}\eta_i\cdot\boldsymbol{e}_i$ where $\{\eta_i\}_{i\in[n]}$ are n i.i.d. random variables drawn from a Laplace distribution $\mathrm{Lap}(\lambda)$. Consider the probability density of the marginal error $\mathrm{Err}\left(\hat{\boldsymbol{x}}_S\right)_1$ at v, i.e., the integration of the original Laplacian noise over a set S_v as follows.

$$f_{\mathrm{Err}(\hat{\boldsymbol{x}}_S)_1}(v) = \int_{\boldsymbol{y}\in S_v}\frac{1}{(2\lambda)^n}\exp\left(-\frac{\|\boldsymbol{y}-\boldsymbol{x}\|_1}{\lambda}\right)d\boldsymbol{y},$$

where $S_v = \{\boldsymbol{u} \mid \hat{\boldsymbol{x}}_S(\boldsymbol{u})_1 - x_1 = v\}$. To compute this integration, it is easier to exploit the definition of the projection operator and express S_v differently through a basis transformation. Given the new basis $\{\boldsymbol{e}'_i\}_{i\in[n]}$, the set S_v can be expressed in the following form

$$S_v = \left\{\boldsymbol{u} = \boldsymbol{x} + \sum_{i=1}^{n} c'_i\cdot\boldsymbol{e}'_i \;\middle|\; c'_n = v\cdot\sqrt{\frac{n}{n-1}}\right\}$$

where $\{\boldsymbol{e}'_i\}_{i\in[n]}$ is an orthonormal basis of $\mathbb{R}^n$ with $\boldsymbol{e}'_1 = \frac{1}{\sqrt{n}}[1,\dots,1]^T$ and $\boldsymbol{e}'_n$ has $\sqrt{(n-1)/n}$ as its first component and $-\sqrt{1/(n(n-1))}$ as the remaining ones. The marginal probability density distribution $\mathrm{Err}\left(\hat{\boldsymbol{x}}_S\right)_1$ is then given by

$$\int_{\boldsymbol{y}'_{-n}\in\mathbb{R}^{n-1}} f'_{\boldsymbol{\eta}}\left(\boldsymbol{y}'_{-n}, v\cdot\sqrt{\frac{n}{n-1}}\right)d\boldsymbol{y}'_{-n}$$

where $\boldsymbol{y} = [y_1,\dots,y_n]^\top$, $\boldsymbol{y}'_{-n} = [y'_1,\dots,y'_{n-1}]^\top$, and $f'_{\boldsymbol{\eta}}$ represents the Laplace probability density function distribution under the new basis $\{\boldsymbol{e}'_i\}_{i\in[n]}$. It comes that the random variable η'_n, i.e., the noise η_n in the new basis, shares the same distribution with

$$\langle\boldsymbol{e}'_n,\boldsymbol{\eta}\rangle = \sum_{i=1}^{n}\eta_i\cdot\langle\boldsymbol{e}'_n,\boldsymbol{e}_i\rangle = \frac{(n-1)\eta_1 - \sum_{i=2}^{n}\eta_i}{\sqrt{n(n-1)}}.$$

Since, for any $v\in\mathbb{R}$, $f_{\mathrm{Err}(\hat{\boldsymbol{x}}_S)_1}(v) = f_{\eta'_n}\left(v\cdot\sqrt{\frac{n}{n-1}}\right)$,

$$\mathrm{Err}\left(\hat{\boldsymbol{x}}_S\right)_1 \sim \frac{(n-1)\eta_1 - \sum_{i=2}^{n}\eta_i}{n}.$$

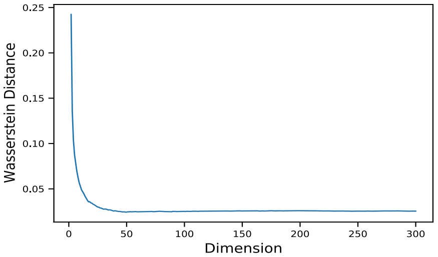

Figure 5: Wasserstein Distance between the Laplacian Distribution and the Marginal Error Distribution.

By independence of $\{\eta_i\}_{i\in[n]}$, it follows that

$$\mathrm{Var}\left(\mathrm{Err}\left(\hat{\boldsymbol{x}}_S\right)_1\right) = 2\lambda^2\left(1-\frac{1}{n}\right).$$

□

Figure 5 highlights Theorem 12. It shows how the Wasserstein distance between the distributions of the Laplace residuals and the post-processed residuals. As the figure indicates, the Wasserstein distance decreases quickly as the problem dimension increases.

Theorem 12 also reveals some fundamental insights about post-processing. First, it shows that post-processing reduces the variance of the noise, while preserving differential privacy. In other words, post-processing in this setting does not introduce bias and leverages the public information (i.e., b) to reduce the variance. This is again a positive result as reducing the variance may reduce fairness issues when using the data in decision-making processes. Second, it shows that different aggregation sizes (i.e., different values of n) may lead to disparate impacts and fairness issues. Indeed, consider two counties a and b with approximately the same sizes which are aggregated differently: a is aggregated with n_a other counties, b is aggregated with n_b, with $n_a \gg n_b$, and the aggregated data is public. Then the variance of the post-processed value for a will be substantially larger than the the variance of the post-processed value of b, potentially creating situations where counties a and b will be treated fundamentally differently in decision-making processes. Hence, although post-processing reduces variance, its application should take into account fairness issues. Once again, the key to ensure fairness is to make sure that quantities being released are of the same order of magnitude.

On the census data-release problem, when comparing the solutions returned by program (P_{S}) for the states of Arizona—which has 15 counties—and Texas—which has 254 counties—it is found that both the theoretical and empirical difference in their variance to be roughly 6%. This result highlights the importance of the finding.

The following results show that the marginal error converges in distribution to the Laplace distribution.

Theorem 13. *The variance of the resulting marginal error of program* (P_{S}) *is increasing in the dimension n and converges to that of the marginal Laplacian noise added to the true data $\boldsymbol{x}$, as the dimension n tends to infinity, i.e.,*

$$\lim_{n\to\infty}\mathrm{Var}\left(\mathrm{Err}\left(\hat{\boldsymbol{x}}_S\right)\right) = \mathrm{Var}\left(\mathrm{Lap}(\lambda)\right) = 2\lambda^2.$$

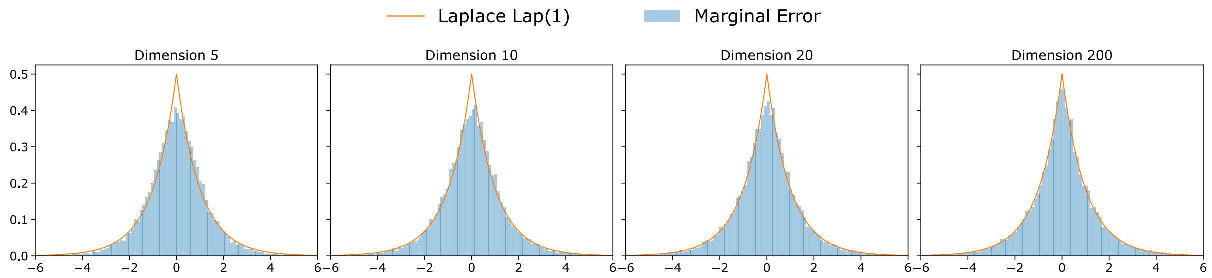

Figure 6: Illustrating the Convergence Results of Theorem 14.

Theorem 14. *As the dimension n goes to infinity, the marginal error of program (P_S) converges in distribution to the marginal Laplacian noise , i.e.,*

$$\mathrm{Err}\left(\hat{\boldsymbol{x}}_S\right) \xrightarrow{d} \mathrm{Lap}(\lambda), \qquad \text{as } n \to \infty,$$

Proof. By Lemma 12, the marginal error $\mathrm{Err}\left(\hat{\boldsymbol{x}}_S\right)$ follows

$$\frac{(n-1)\eta_1 - \sum_{i=2}^{n} \eta_i}{n},$$

where $\{\eta_i\}_{i\in[n]}$ are the n i.i.d. random variables drawn from a Laplace distribution $\mathrm{Lap}(\lambda)$. Let η be a Laplacian random variable $\mathrm{Lap}(\lambda)$: Its cumulative distribution function is

$$\Pr\left(\eta \le v\right) = \begin{cases} \frac{1}{2}\exp\left(\frac{v}{\lambda}\right), & v \le 0, \\ 1 - \frac{1}{2}\exp\left(\frac{-v}{\lambda}\right), & v > 0. \end{cases}$$

The cumulative distribution function of $(n-1)\eta_1/n$ is

$$\Pr\left(\frac{n-1}{n}\eta_1 \le v\right) = \begin{cases} \frac{1}{2}\exp\left(\frac{nv}{(n-1)\lambda}\right), & v \le 0, \\ 1 - \frac{1}{2}\exp\left(\frac{-nv}{(n-1)\lambda}\right), & v > 0. \end{cases}$$

Note that, for any $v \in \mathbb{R}$,

$$\lim_{n\to\infty} \Pr\left(\frac{n-1}{n}\eta_1 \le v\right) = \Pr\left(\eta \le v\right),$$

which implies that the random variable $(n-1)\eta_1/n$ converges to η in distribution. By the Weak Law of Large Numbers, the sample mean among $\{\eta_i\}_{i\in\{2,\ldots,n\}}$ converges in probability to their common expectation 0, i.e., for $\xi > 0$,

$$\lim_{n\to\infty} \Pr\left(\left|\frac{\sum_{i=2}^{n}\eta_i}{n-1}\right| \ge \xi\right) = 0.$$

Additionally, for any $\xi > 0$ and $n \ge 2$,

$$\left|\frac{\sum_{i=2}^{n}\eta_i}{n}\right| \ge \xi \implies \left|\frac{\sum_{i=2}^{n}\eta_i}{n-1}\right| \ge \xi.$$

which implies that

$$\Pr\left(\left|\frac{\sum_{i=2}^{n}\eta_i}{n}\right| \ge \xi\right) \le \Pr\left(\left|\frac{\sum_{i=2}^{n}\eta_i}{n-1}\right| \ge \xi\right).$$

By the squeeze theorem, the random variable $\sum_{i=2}^{n}\eta_i/n$ converges to 0 in probability, as the dimension n goes to infinity. Since

$$\frac{n-1}{n}\eta_1 \xrightarrow{d} \eta, \qquad \frac{\sum_{i=2}^{n}\eta_i}{n} \xrightarrow{p} 0,$$

by Slutsky's Theorem (Billingsley 2013), it follows that

$$\frac{(n-1)\eta_1 - \sum_{i=2}^{n}\eta_i}{n} \xrightarrow{d} \eta \qquad \text{as } n \to \infty.$$

□

Figure 6 illustrates Theorem 14. It depicts the convergence to the Laplace distribution as the dimension increases. Finally, it is also interesting to report some experimental results on census data and, in particular, the states of Arizona (population of 2,371,715 and 15 counties) and Texas (population of 8,887,839 and 254 counties). For $\lambda = 10$, the distribution variances are 186.67 and 199.21 for Arizona and Texas respectively. Over 80,000 experiments, the empirical variances were 186.88 and 199.32 respectively. These results clearly highlight the influence of the problem dimension (i.e., the number of counties) on the variance.

Conclusion

This paper was motivated by the recognition that the effect of post-processing on the noise distribution is poorly understood: It took a first step towards understanding the theoretical and empirical properties of post-processing. Motivated by census applications, it studied the behavior of *projections* for domains where the feasibility space is specified by linear equations. The paper showed that, when non-negativity constraints are absent, the projection does not introduce bias. With non-negativity constraints, the paper presented an upper bound on the bias, providing insights on the type of problems for which the bias will be significant. The paper also provided a detailed analysis of the important sub-problem with one linear equation arising in hierarchical data-release problems. It fully characterized the residual distribution of the post-processed noise, showing that it converges towards the selected noise distribution when the dimension of the feasible space increases. This last result shed an interesting light on the effect of post-processing on the variance of the post-processed data. Indeed, in this case, post-processing reduces the variance by exploiting the public information available. These results may have strong implications with respect to group fairness and should inform statistical agencies about the trade-off between the granularity of the released data, the bias, and the variance.

Ethics Statement

The authors believe that this research may inform census agencies around the world and help them release data which will improve group fairness. By avoiding bias and differences in variance, census agencies may ensure that the data release do not lead to situations where downstream decision-making processes negatively affect some groups more than others because of the privacy requirements.

References

Abowd, J. M. 2018. The US Census Bureau adopts differential privacy. In *ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2867–2867. ACM.

Billingsley, P. 2013. *Convergence of probability measures*. John Wiley & Sons.

Cormode, G.; Procopiuc, C.; Srivastava, D.; Shen, E.; and Yu, T. 2012a. Differentially private spatial decompositions. In *2012 IEEE 28th International Conference on Data Engineering*, 20–31. IEEE.

Cormode, G.; Procopiuc, C.; Srivastava, D.; Shen, E.; and Yu, T. 2012b. Differentially private spatial decompositions. In *2012 IEEE 28th International Conference on Data Engineering*, 20–31. IEEE.

Dwork, C.; McSherry, F.; Nissim, K.; and Smith, A. 2006. Calibrating noise to sensitivity in private data analysis. In *Theory of cryptography conference*, 265–284. Springer.

Erlingsson, Ú.; Pihur, V.; and Korolova, A. 2014. Rappor: Randomized aggregatable privacy-preserving ordinal response. In *Proceedings of the 2014 ACM SIGSAC conference on computer and communications security*, 1054–1067. ACM.

EU. 2016. European Parliament and Council of European Union, General Data Protection Regulations (GDPR). URL https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679&from=EN.

Fioretto, F.; Lee, C.; and Van Hentenryck, P. 2018. Constrained-based Differential Privacy for Private Mobility. In *Proceedings of the International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 1405–1413.

Fioretto, F.; Mak, T. W. K.; and Van Hentenryck, P. 2020. Differential Privacy for Power Grid Obfuscation. *IEEE Transactions on Smart Grid* 11(2): 1356–1366. ISSN 1949-3061. doi:10.1109/TSG.2019.2936712.

Fioretto, F.; and Van Hentenryck, P. 2019. Differential privacy of hierarchical census data: An optimization approach. In *International Conference on Principles and Practice of Constraint Programming*, 639–655. Springer.

Hay, M.; Rastogi, V.; Miklau, G.; and Suciu, D. 2010. Boosting the accuracy of differentially private histograms through consistency. *Proceedings of the VLDB Endowment* 3(1-2): 1021–1032.

He, X.; Cormode, G.; Machanavajjhala, A.; Procopiuc, C. M.; and Srivastava, D. 2015. DPT: Differentially Private Trajectory Synthesis Using Hierarchical Reference Systems. *Proc. VLDB Endow.* 8(11): 1154–1165. ISSN 2150-8097.

Qardaji, W.; Yang, W.; and Li, N. 2013. Understanding hierarchical methods for differentially private histograms. *Proceedings of the VLDB Endowment* 6(14): 1954–1965.

Team, A. D. P. 2017. Learning with privacy at scale. *Apple Machine Learning Journal* 1(8).

U.S. Census Bureau. 2020. Title 13: Protection of confidential information. URL https://www.census.gov/about/policies/privacy/data_stewardship/title_13_-_protection_of_confidential_information.html.

Xiao, Y.; Xiong, L.; and Yuan, C. 2010. Differentially private data release through multidimensional partitioning. In *Workshop on Secure Data Management*, 150–168. Springer.

Zhang, J.; Xiao, X.; and Xie, X. 2016. Privtree: A differentially private algorithm for hierarchical decompositions. In *Proceedings of the 2016 International Conference on Management of Data*, 155–170.