OPTIMAL ALLOCATION OF PRIVACY BUDGET ON HIERARCHICAL DATA RELEASE

Joonhyuk Ko University of Virginia tah3af@virginia.edu Juba Ziani Georgia Institute of Technology jziani3@gatech.edu Ferdinando Fioretto University of Virginia fioretto@virginia.edu

ABSTRACT

Releasing useful information from datasets with hierarchical structures while preserving individual privacy presents a significant challenge. Standard privacy-preserving mechanisms, and in particular Differential Privacy, often require careful allocation of a finite privacy budget across different levels and components of the hierarchy. Sub-optimal allocation can lead to either excessive noise, rendering the data useless, or to insufficient protections for sensitive information. This paper *addresses the critical problem of optimal privacy budget allocation for hierarchical data release*. It formulates this challenge as a constrained optimization problem, aiming to maximize data utility subject to a total privacy budget while considering the inherent trade-offs between data granularity and privacy loss. The proposed approach is supported by theoretical analysis and validated through comprehensive experiments on real hierarchical datasets. These experiments demonstrate that optimal privacy budget allocation significantly enhances the utility of the released data and improves the performance of downstream tasks.

1 Introduction

Hierarchical data structures are ubiquitous, appearing in domains such as geographical information systems, organizational charts, biological taxonomies, and product catalogs. While the analysis of such data impacts society and the economy as it touches disciplines from scientific discovery to policy-making, it is often regulated by privacy laws or ethical considerations. For instance, the U.S. Census Bureau's data is a prime example of hierarchical data, where population statistics are organized at various levels, such as national, state, and county. This data is crucial for a wide range of applications, including resource allocation, migration studies, and public health decisions. Its release is also governed by strict privacy regulations, including the Title 13 of the U.S. Code, which prohibits the disclosure of personally identifiable information. As a result, privacy-preserving techniques have been developed to ensure that sensitive information remains confidential while still allowing for meaningful analysis. Among these, Differential Privacy (DP) [1] has emerged as the gold standard, primarily due to its strong, quantifiable guarantees against re-identification and inference attacks.

However, hierarchical data presents unique challenges for differential privacy, as its privacy budget must be allocated across different levels of the hierarchy. For instance, one might need to release aggregate statistics at a coarse geographical level (e.g., state) and more fine-grained statistics at a lower level (e.g., county or city). A naive or uniform allocation of the privacy budget across these levels can lead to suboptimal outcomes: either too much noise is added to fine-grained data, destroying its utility, or coarse-grained data is overly protected at the expense of detail elsewhere. This necessitates a principled approach to privacy budget allocation that explicitly considers the structure of the data and the utility requirements at different granularities.

This observation raises a fundamental question: *How can we optimally allocate a given privacy budget across a hierarchical data structure to maximize the utility of the released information?* The answer to this question is not straightforward, as it involves balancing the trade-off between privacy and utility across multiple levels of the hierarchy. As the paper will show, commonly adopted heuristic approaches often achieve suboptimal results, leading to inaccurate downstream analyses. This paper precisely addresses this problem, and provides the following contributions:



Level	Region	Population
Level 3	Block 1 (100)	120
Level 3	Block 2 (100)	80
Level 3	Block 3 (100)	100
Level 2	Tract 100	300
Level 3	Block 1 (200)	90
Level 3	Block 2 (200)	60
Level 2	Tract 200	150
Level 1	VA	450

Figure 1: A hierarchical structure of geographic regions, where blocks are nested within tracts, and tracts within the state of Virginia.

Table 1: Population counts corresponding to each region in the hierarchy shown in Figure 1.

- C_1 It defines and characterizes the problem of optimal privacy budget allocation for hierarchical data release under differential privacy. This includes defining appropriate utility metrics that capture the usefulness of data at different levels of the hierarchy.
- C_2 It provides an analysis of the properties of the optimal allocation strategy (as detailed in Section 3), in particular the relationship between bias and variance of the released data, and how these are affected by the privacy budget allocation.
- C_3 Crucially, the analysis shows that the optimal allocation strategy is not necessarily uniform across levels, and utility can be significantly improved by accounting for the hierarchy's structure. To address this, it proposes an optimization-based approach that minimizes an analytical expression for mean squared error, combining bias and variance introduced by differential privacy under non-negativity post-processing, commonly used in histogram and contingency table releases. Importantly, by minimizing this objective, the method improves the utility of the released data even after the post-processing, which is known to distort bias and variance [2].
- C_4 Finally, the paper demonstrates the practical utility of the proposed method through extensive experiments on real-world datasets derived from the U.S. Census. The results show a substantial reduction of **both** bias and variance compared to uniform allocation under the same privacy budget. Furthermore, motivated by U.S. Census budget allocation, the framework is extended to a downstream resource allocation task using privatized data. This extension examines how different classes of preference functions influence bias and variance, with results consistently showing higher utility for the optimal allocation strategy over uniform allocation.

The remainder of this paper is organized as follows. Section 2 introduces essential preliminaries on differential privacy and hierarchical data, and sets up the problem. Section 3 details our proposed allocation methodology. Section 4 presents the experimental results, focusing on Wyoming as a running example, with additional state experiments in Appendix D. Section 5 explores the impact on downstream tasks. Finally, Section 6 concludes the paper and outlines future research directions.

Related Work. Since the adoption of the U.S. Census Bureau's TopDown Algorithm (TDA) [3], there has been growing interest in exploring hierarchical data release under differential privacy. This has motivated theoretical and practical approaches for allocating privacy budgets across hierarchical levels [4, 5, 6, 7]. Accurately privatizing hierarchical data releases, such as those in the U.S. Census, is crucial, as it directly impacts federal resource allocation and has significant implications for fairness in downstream decisions [8, 9, 10]. Beyond population statistics, hierarchical data plays a key role in other domains. Liu et al [11] explored generating private synthetic hierarchical data, while hierarchical structures have also been used in federated learning to enhance privacy and scalability [12]. Fioretto et al. [5] proposed an optimization-based method for post-processing privatized hierarchical data to ensure consistency and improve utility. Cohen et al. [6] examined how privacy budget allocation affects redistricting accuracy in the Census TopDown algorithm, finding that bottom-heavy allocations improve accuracy for small geographic units. Similarly, Dawson et al. [13] proposed a greedy iterative approach using noisy priors and demonstrated that unequal budget allocations can achieve higher accuracy than uniform splits.

However, there remains a gap in understanding how to optimally allocate privacy budgets across different levels of the hierarchy, especially when common post-processing alters the bias and variance of differentially private estimates.

2 Preliminaries and Problem Setting

2.1 Hierarchical Data Release

This paper considers the hierarchical release of population statistics. Let R_1, R_2, \ldots, R_L denote the sets of regions at each level of a hierarchy, where each region in R_ℓ is indexed by a tuple (i_1, \ldots, i_ℓ) indicating its position in the hierarchy. Specifically, R_1 contains highest-level regions (e.g., states in Fig 1), and each R_ℓ contains subregions nested within regions from level $\ell - 1$. Let N_{i_1,\ldots,i_ℓ} denote the population count associated with region $(i_1,\ldots,i_\ell) \in R_\ell$. These counts satisfy the consistency constraints across levels:

$$\sum_{(i_1,\dots,i_L)\in R_L} N_{i_1,\dots,i_L} = \dots = \sum_{i_1\in R_1} N_{i_1} = N_{\text{total}}.$$

Example. In Figure 1, $R_1 = \{(VA)\}$ with $N_{total} = 450$. At level 2, we have $R_2 = \{(VA, 100), (VA, 200)\}$ with populations $N_{VA,100} = 300$ and $N_{VA,200} = 150$. At level 3, R_3 contains five blocks: (VA, 100, 1), (VA, 100, 2), (VA, 100, 3), (VA, 200, 1), (VA, 200, 2) with respective populations 120, 80, 100, 90, and 60. These satisfy the consistency condition:

$$\sum_{(i_1,i_2,i_3)\in R_3} N_{i_1,i_2,i_3} = \sum_{(i_1,i_2)\in R_2} N_{i_1,i_2} = N_{\text{total}}.$$

2.2 Differential Privacy

This paper considers the problem of releasing privatized hierarchical data under the framework of differential privacy (DP) [1, 14]. Differential privacy provides a formal framework for quantifying the privacy loss incurred when releasing information derived from sensitive data. It ensures that the output of a computation is nearly indistinguishable whether or not any single individual's data is included in the input dataset. This is typically achieved by adding controlled noise to the output, calibrated to a privacy budget that quantifies the level of privacy protection.

Definition 1 (Differential Privacy). A randomized algorithm \mathcal{M} satisfies ε -differential privacy if, for all $\mathcal{S} \subseteq \text{Range}(\mathcal{M})$ and any two neighboring databases $D_1 \sim D_2$ (i.e., differing by one entry),

$$\Pr[\mathcal{M}(D_1) \in \mathcal{S}] \le \exp(\varepsilon) \Pr[\mathcal{M}(D_2) \in \mathcal{S}].$$

Here, $\varepsilon > 0$ is the privacy parameter that quantifies the privacy loss: smaller values ($\varepsilon \to 0$) imply strong privacy, while larger values ($\varepsilon \to \infty$) correspond to weak privacy.

Differential privacy provides two key properties essential to this work: sequential composition and post-processing immunity. Sequential composition quantifies total privacy loss across hierarchical levels, while post-processing immunity guarantees that adjustments such as enforcing consistency or non-negativity do not affect the privacy guarantee.

Theorem 1 (Sequential Composition). Let \mathcal{M}_1 be an ε_1 -differentially private algorithm, and let \mathcal{M}_2 be an ε_2 differentially private algorithm. Then the combined algorithm that outputs $(\mathcal{M}_1(x), \mathcal{M}_2(x))$ is $\varepsilon_1 + \varepsilon_2$ -differentially private.

Theorem 2 (Post-Processing Immunity). If a randomized algorithm \mathcal{M} is ε -differentially private, and f is any arbitrary function, then the composition $f \circ \mathcal{M}$ is also ε -differentially private.

2.3 Differentially Private Hierarchical Data Release Model

Motivated by Census applications, we consider the release of privatized hierarchical data by adding noise at each level of a hierarchy. Let $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_L$ denote the privacy budgets allocated from the highest level (ε_1) to the lowest (ε_L), with the total privacy budget given by the sequential composition:

$$\varepsilon_{\text{Total}} = \varepsilon_1 + \varepsilon_2 + \ldots + \varepsilon_L.$$

In this work, we add independent Laplace¹ noise (which is chosen for clarity in the analysis of bias and variance later in the paper) to each node in the hierarchy, and apply non-negativity post-processing to ensure that released counts remain valid for population or frequency data, as noise can render some counts negative:

$$\tilde{N} = \max\left(0, N + \operatorname{Lap}\left(\frac{1}{\varepsilon}\right)\right).$$
(1)

¹While we focus on the Laplace mechanism, similar analysis can be carried out under other mechanisms such as Gaussian, with qualitatively similar tradeoffs between bias, variance, and utility.

Independent noise addition can lead to inconsistencies across levels, such as lower-level counts not summing to their corresponding higher-level total. To address this, we apply a hierarchical consistency post-processing step, also described in Section 3.

3 The Optimal Allocation Problem

Given a fixed total privacy budget, determining how to allocate noise across hierarchical levels is non-trivial. On one hand, assigning more of the privacy budget to lower levels can improve the accuracy of fine-grained counts, which may be beneficial for downstream tasks. On the other hand, allocating more of the budget to higher levels may be preferable when hierarchical consistency constraints are imposed. Even if we understand whether a bottom-heavy or top-heavy allocation is preferable, it remains unclear how much budget should be assigned to each level.

3.1 Bias and Variance of \tilde{N}

A high-utility privatized release should minimize the mean squared error (MSE), defined as:

$$MSE = \mathbb{E}[(\tilde{N} - N)^2] = Bias(\tilde{N})^2 + Var(\tilde{N}), \qquad (2)$$

which decomposes into a bias and a variance term through the standard decomposition [15]. Thus, a high-utility hierarchical release should exhibit low bias and variance. We begin by deriving the bias introduced by the non-negativity post-processing applied after adding Laplace noise, as shown in prior work [16].

Proposition 1. For the population N, the bias of \tilde{N} is given in closed-form by:

$$\operatorname{Bias}(\tilde{N}) = \mathbb{E}\left[\tilde{N}\right] - N = \frac{1}{2\varepsilon} \exp\left(-N\varepsilon\right) > 0.$$

Proof. The proof is provided in Appendix A.

Note that the bias is strictly positive. As $\varepsilon \to \infty$ (no privacy) or $N \to \infty$, the $\text{Bias}(\tilde{N}) \to 0$. This implies that lower levels of the hierarchy—where each count is typically smaller—will suffer more from the positive bias.

We now derive the variance of the non-negativity post-processing applied after adding Laplace noise.

Proposition 2. For the population N, the variance of \tilde{N} is given in closed-form by:

$$\operatorname{Var}(\tilde{N}) = \frac{1}{\varepsilon^2} \left(2 - e^{-N\varepsilon} \right) - \frac{N}{\varepsilon} e^{-N\varepsilon} - \frac{1}{4\varepsilon^2} e^{-2N\varepsilon}.$$

Proof. The proof is provided in Appendix A.

As $\varepsilon \to \infty$ (no privacy), the variance $\operatorname{Var}(\tilde{N}) \to 0$. Conversely, as $N \to \infty$, the variance increases monotonically and approaches $\frac{2}{\varepsilon^2}$, which is the variance of the untruncated Laplace noise. This indicates that small population counts incur lower variance due to the truncation effect introduced by the non-negativity constraint, which reduces the overall variability. However, this reduction in variance comes at the cost of increased bias, as lower truncation shifts the mean upwards.

Using the standard decomposition in equation (2) and applying Proposition 1 and 2, the MSE admits the following closed-form expression:

$$MSE = Bias(\tilde{N})^2 + Var(\tilde{N}) = \frac{1}{\varepsilon^2} \left(2 - e^{-\varepsilon N}\right) - \frac{N}{\varepsilon} e^{-\varepsilon N}.$$
(3)

As $\varepsilon \to \infty$ (no privacy), the noise magnitude vanishes and MSE $\to 0$. Conversely, as $\varepsilon \to 0$, we have MSE $\to \infty$ due to increasing noise. When $N \to \infty$, the positive bias diminishes, and the MSE is dominated by the variance term. *This highlights that bias is most significant when true counts are small, whereas for large counts, the error is primarily driven by noise variance.* We now state two key propositions that play an important role in optimization and privacy budget allocation. We provide proofs in Appendix A.

Proposition 3. The MSE is strictly convex in ε over $\mathbb{R}_{>0}$ for any fixed N > 0.

Proposition 4. For any fixed $\varepsilon > 0$ and $N \ge 0$, the MSE is bounded as

$$\frac{1}{\varepsilon^2} \le \text{MSE} < \frac{2}{\varepsilon^2}.$$

In the context of hierarchical data release, the *total MSE* across a multi-level hierarchy with L levels (each allocated a privacy budget ε_{ℓ}) is given by:

$$\mathsf{MSE}_{\mathrm{Total}} = \sum_{\ell=1}^{L} \sum_{j \in R_{\ell}} \left[\frac{1}{\varepsilon_{\ell}^{2}} (2 - e^{-\varepsilon_{\ell} N_{j}}) - \frac{N_{j}}{\varepsilon_{\ell}} e^{-\varepsilon_{\ell} N_{j}} \right],\tag{4}$$

where R_{ℓ} denotes the set of nodes at level ℓ . This expression simply sums the per-node MSE across all levels of the hierarchy. Then, Proposition 4 suggests the following: under a uniform privacy budget (i.e., $\varepsilon_1 = \cdots = \varepsilon_L$), lower levels, containing as many or more nodes than higher levels, contribute more to MSE_{Total}, as the total MSE incurred at these levels is at least as large as that of the higher level. Furthermore, since a finite sum of strictly convex functions remains strictly convex, MSE_{Total} is strictly convex in ε by Proposition 3. This property is crucial, as it ensures that the privacy budget allocation problem can be formulated as a convex optimization problem.

3.2 Optimization Programs

To allow flexibility in prioritizing accuracy at different levels, we define a *weighted MSE objective* that incorporates level-wise preferences through weights:

$$\mathsf{MSE}_{\mathbf{w}}(\boldsymbol{\varepsilon}) = \sum_{\ell=1}^{L} w_{\ell} \cdot \underbrace{\sum_{j \in R_{\ell}} \left[\frac{1}{\varepsilon_{\ell}^{2}} (2 - e^{-\varepsilon_{\ell} N_{j}}) - \frac{N_{j}}{\varepsilon_{\ell}} e^{-\varepsilon_{\ell} N_{j}} \right]}_{\mathsf{MSE at level } \ell}, \tag{5}$$

where $\mathbf{w} = (w_1, \ldots, w_L) \in \mathbb{R}_{\geq 0}^L$ denotes the level-wise weights, and $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_L)$ is the vector of privacy budgets. These weights allow the data curator to emphasize accuracy at specific levels based on downstream utility needs. In Section 4.1, we focus on the case of equal weights $(w_1 = w_2 = \cdots = w_L)$, minimizing the total MSE across the hierarchy without prioritizing any particular level. An ablation study examining how different weight choices affect overall and per-level accuracy is then presented in Section 4.2. We now formalize two natural optimization problems.

(1) Minimize MSE subject to a total privacy constraint:

$$\underset{\boldsymbol{\varepsilon} \in \mathbb{R}^{L}_{\geq 0}}{\text{minimize}} \quad \mathsf{MSE}_{\mathbf{w}}(\boldsymbol{\varepsilon}) \tag{6a}$$

s.t.
$$\sum_{\ell=1}^{L} \varepsilon_{\ell} \leq \varepsilon_{\text{Total}}.$$
 (6b)

This program reflects settings where a fixed total privacy loss must be allocated across levels to minimize total error. It is important to note that to satisfy ε -DP, one must base the optimization on previously released statistics or a similar data structure, rather than directly computing the total MSE from the raw statistics to be privatized. This is because using raw statistics in the algorithm to determine the privacy budget would itself leak information. However, if only the total privacy loss is released (rather than individual allocations at each hierarchical level), then it may be permissible to use the ground statistics for budget allocation.

(2) Minimize total privacy loss subject to a utility target:

$$\min_{\boldsymbol{\varepsilon} \in \mathbb{R}_{\geq 0}^{L}} \sum_{\ell=1}^{L} \varepsilon_{\ell}$$
(7a)

s.t.
$$MSE_{w}(\varepsilon) \leq \tau$$
. (7b)

This formulation applies when the desired level of utility is fixed (e.g., when the data curator specifies a target error threshold τ) and seeks to minimize the total privacy cost. Such scenarios are common in practice when utility guarantees must be met under strict privacy budgets [17]. The optimization determines how to allocate privacy budgets across levels to meet the accuracy constraint with minimal cumulative privacy loss.

Convexity and Optimal Allocation. The objective $MSE_{\mathbf{w}}(\varepsilon)$ (5) is strictly convex in ε over $\mathbb{R}_{>0}^{L}$. This follows from Proposition 3, which establishes the strict convexity of each node-level MSE in ε_{ℓ} for fixed N_{j} , and the fact that convexity is preserved under nonnegative weighted sums and finite summations. As a result, the optimization problems (6) and (7) are convex and can be efficiently solved in polynomial time using interior-point (IP) methods.

Furthermore, we establish a key theoretical result showing that, under equal weighting, the optimal privacy budget allocation is monotonic across levels and allocates more budget to lower levels in the hierarchy.



Figure 2: Privacy budget allocation using Optimization Program (7) in Wyoming.

Theorem 3. Under equal weighting, the optimal allocation of the privacy budget satisfies

$$\varepsilon_1 \leq \varepsilon_2 \leq \cdots \leq \varepsilon_L.$$

Proof. The proof is provided in Appendix A.

3.3 Hierarchical Post-Processing

It is often desirable to post-process the privatized data to enforce hierarchical consistency (i.e., the sum of lower-level counts must match the total) since independent noise addition can violate this constraint. Due to post-processing immunity, this adjustment does not compromise privacy. The following program enforces consistency, as also described in [18, 5]:

$$\min_{v \in \mathbb{R}^{n}_{\geq 0}} \|v - \{\tilde{N}_{i}\}_{i=1}^{n}\|_{2}$$
(8a)

$$s.t.\sum_{i} v_i = \tilde{N}_{\text{total}}.$$
(8b)

In multi-level hierarchies, this procedure is applied recursively in a top-down manner: first between levels 1 and 2, then between each level-2 node and its children at level 3, and so on.

4 Experimental Results

Next, we present empirical evidence demonstrating the effectiveness of the proposed optimization method. We evaluate both bias and variance, as well as the impact of enforcing hierarchical constraints. Experiments are conducted on the release of three-level population statistics, where the hierarchy follows the geographic structure: State \rightarrow Census Tract \rightarrow Census Block.

Dataset and Setting. Our experiments use a subset of the U.S. Census data from the 2020 Census Privacy-Protected Microdata File (PPMF) [19], released under the disclosure avoidance methodology adopted for the 2020 Census. Specifically, we use the 2010 release version of the PPMF [20] to compute the budget allocation that satisfies exact ε -DP.

In the main experiment, we focus on the state of Wyoming (1 state, 128 Census Tracts, and 20,975 Blocks), where ε_1 , ε_2 , and ε_3 denote the privacy budgets for the State, Tract, and Block levels, respectively. Additional results for other states are provided in the Appendix D, showing the same clear trend and reinforcing that the analysis presented in this section applies broadly.

Evaluation Metrics and Baseline. We evaluate both Bias² and Variance, following the standard decomposition of mean squared error (MSE). Our baseline is uniform allocation of the total privacy budget, a commonly suggested heuristic for hierarchical data structures [6]. For example, in this setting, a total privacy budget of $\varepsilon_{\text{total}}$ is divided evenly: $\varepsilon_1 = \varepsilon_2 = \varepsilon_3 = \frac{\varepsilon_{\text{total}}}{3}$.



Figure 3: Hierarchical data release performance in Wyoming.

4.1 Optimized Allocation: Bias² and Variance

We begin by comparing the performance of the DP-post-processed estimates under a non-negativity constraint, without enforcing hierarchical consistency, as shown in Figure 3 under "No Hier." The optimization-based approach significantly outperforms uniform allocation in terms of Bias², with the uniform method yielding 10 times higher Bias². Variance is also approximately four times as high under the uniform approach. This shows that, for the same total privacy budget, both bias and variance can be substantially reduced through optimized budget allocation. Figure 4 illustrates the resulting privacy budget distribution from our proposed method (6), exhibiting the relationship $\varepsilon_1 \leq \varepsilon_2 \leq \varepsilon_3$ as shown by Theorem 3.

Next, we consider the case with hierarchical post-processing, as shown in Figure 3 under "w/ Hier." The overall trend remains similar: the optimized allocation continues to outperform uniform allocation at similar rate. Importantly, both bias and variance are slightly reduced after enforcing hierarchical consistency. By constraining the lower-level population counts to match the upper level, the post-processing step helps mitigate strong positive bias in small counts, which in turn slightly reduces the overall variance.

Finally, we consider the case of minimizing total privacy loss subject to a utility target using program (7), as shown in Figure 2. As the total error tolerance τ increases, the required total privacy budget ε_{total} decreases. For each value of τ , we observe that the majority of the privacy budget is allocated to the lower level, consistent with Theorem 3. Despite using 2010 PPMF data to predict the privacy budget required for the 2020 release,



Figure 4: Optimized privacy budget allocation for Wyoming using program (6).

our estimates closely align with observed utility. For example, at $\tau = 20,000$, the optimization allocates approximately $\varepsilon_{\text{total}} \approx 2$, and in Figure 3, we see that the total MSE (bias² + variance) at this budget is indeed around 20,000.

4.2 Ablation on Weights

To study the effect of weighting, we vary w_3 in equation (5), the weight assigned to level 3, and measure the resulting MSE at each level. The remaining weights are distributed equally between levels 1 and 2, i.e., $w_1 = w_2 = \frac{1-w_3}{2}$. The total privacy budget is fixed at $\varepsilon = 2.0$, and all other parameters are held constant.

Experimental Result. We begin by analyzing how total and level-wise MSE change as a function of the weight w_3 . As shown in Figure 5, increasing w_3 leads to a reduction in MSE at level 3 (MSE₃), as more weight on this level increases the allocated privacy budget ε_3 . Conversely, MSE at levels 1 and 2 increases with w_3 , since their corresponding weights, and thus ε_1 and ε_2 , decrease. When all three weights are equal (i.e., $w_1 = w_2 = w_3 = \frac{1}{3}$), the optimization corresponds to minimizing total MSE. This setting, indicated by the orange vertical line in Figure 5, yields the lowest total error. Deviating from this uniform weighting shifts accuracy toward the prioritized level but increases overall error.



Figure 5: Total and level-wise MSE as a function of varying w_3 (left), and privacy budget allocation across levels from program (6) under varying w_3 (right) in Wyoming.

5 Downstream Allocation Task

Motivated by Census applications where hierarchical data inform budget allocation, such as the distribution of Title I funds [21] or congressional apportionment [22], we analyze the impact of downstream decision-making tasks that rely on privatized hierarchical statistics. Many such tasks distribute a fixed budget *B* across subgroups according to their size or assessed need. These tasks often incorporate a *weight function* to encode policy preferences; for example, sublinear weighting may be used to prioritize minority populations. We formalize this setting using a generalized allocation framework that captures a broad class of weighted budget rules.

Generalized budget allocation task. The process of budget allocation can be broadly described by the following steps:

1. Estimate the size of each group
$$i \in R_2$$
 relative to the total population: $P_i \triangleq \frac{N_i}{\sum_{j \in R_2} N_j} = \frac{N_i}{N_{\text{total}}}$

2. Apply a weight function $W : \mathbb{R} \to \mathbb{R}$ to each group's proportion to reflect policy preferences: $W_i \triangleq W(P_i)$.

3. *Distribute* the total budget *B* proportionally according to the computed weights.

If W is the identity function (i.e., W(p) = p), the allocation is directly proportional to group sizes.

5.1 Bias² and Variance under Different Weight Functions

We study the Bias² and Variance of the allocation derived from privatized weights \tilde{W}_i (i.e., using privatized statistics to compute proportions and apply the weight function) and assess how the choice of weight function W influences downstream performance. Specifically, we evaluate the bias and variance of the resulting *misallocation* as a percentage, rather than in terms of a fixed total budget B.

We evaluate three weight functions: *logarithmic* (which favors minority groups), *linear* (which is proportional), and *quadratic* (which favors majority groups), to examine how allocation preferences affect performance under different privacy mechanisms.

Dataset and Setting. We use the same dataset as in Section 4, where the 2010 PPMF is used as prior data and the 2020 PPMF serves as ground truth. All experiments are conducted on the state of Wyoming, with additional results for other states provided in Appendix D. In this experiment, we randomly select a Census Tract within Wyoming and allocate the privacy budget among the Census Blocks within that tract. The procedure involves: (1) allocating the privacy budget using program (6), (2) applying hierarchical post-processing for consistency using program (8), and (3) computing weighted group proportions based on the resulting privatized counts.

Experimental Results. We begin by evaluating performance under *linear weighting*, where the optimal allocation using Program 6 significantly reduces both bias and variance. This correction arises from better representation of group sizes in the allocation, reducing misallocation error.

For *nonlinear weight functions*, bias will shift by Jensen's inequality: since the weight function $W(\cdot)$ is applied to noisy proportions, we have

$$\mathbb{E}[W(\tilde{P}_i)] \neq W(\mathbb{E}[\tilde{P}_i]),$$



Figure 6: Bias² and Variance for three weight functions W(p): logarithmic $(\log(p+1))$, linear (p), and quadratic (p^2) , in a single Census Tract in Wyoming.

which introduces bias. For *convex* functions such as the quadratic, this leads to an positive bias in expected weights; for *concave* functions such as the logarithmic, it results in negative bias. As a result, Bias² is generally larger under nonlinear weighting than under linear.

Under *log weighting*, group size differences are compressed, promoting more equal allocation. Since uniform allocation ignores group sizes and tends to overestimate small populations due to higher noise, it performs more similarly to the optimal allocation under log weighting than under linear or quadratic, resulting in a smaller performance gap.

Regarding variance, log weighting tends to dampen the impact of noise, often resulting in lower variance. In contrast, quadratic weighting amplifies group size differences and can lead to increased variance. However, the behavior of variance is also influenced by the *skewness* of the population. In highly skewed populations, this contrast becomes less apparent. The log transformation flattens group differences, which makes small, noisy groups more influential after normalization, potentially increasing variance. Conversely, the quadratic transformation exaggerates group differences, allowing large, stable groups to dominate the normalization, which can reduce variance in the final allocation.

Overall, across all weight functions, *optimal allocation consistently outperforms uniform allocation*, although the degree of improvement varies with the weighting scheme and population structure.

These findings also highlight the importance of choosing a weighting scheme that aligns with the intended policy objective, as different schemes impact bias and variance in distinct ways. This suggests that practitioners should choose weight functions with careful attention to the fairness objectives or utility demands of their specific application.

6 Conclusion

This paper addresses the fundamental challenge of optimal privacy budget allocation for hierarchical data release under differential privacy. We propose a convex optimization framework that minimizes the total *mean squared error* (MSE), accounting for both bias and variance introduced by Laplace noise and non-negativity post-processing. Our theoretical analysis reveals that optimal allocations are bottom-heavy, assigning more of the privacy budget to lower levels of the hierarchy. This strategy is shown to be strictly superior to uniform allocation, both analytically and empirically.

Through extensive experiments on real U.S. Census data, the paper shows that the proposed allocation significantly reduces both bias and variance compared to a standard approach. Additionally, we extend our framework to evaluate downstream resource allocation tasks, showing that optimal allocation improves utility across a range of preference-weighted schemes, highlighting its practical and societal relevance in applications such as public funding and policy-driven resource distribution.

These findings emphasize the importance of structured privacy budget allocation in practical applications of differential privacy, especially in scenarios where hierarchical consistency and data utility are critical. Future directions include extending the approach to other privatization mechanisms, incorporating fairness constraints, and exploring its impact on broader policy-driven decisions.

References

- Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In Shai Halevi and Tal Rabin, editors, *Theory of Cryptography*, pages 265–284, Berlin, Heidelberg, 2006. Springer Berlin Heidelberg.
- [2] Keyu Zhu, Pascal Van Hentenryck, and Ferdinando Fioretto. Bias and variance of post-processing in differential privacy. In *AAAI Conference on Artificial Intelligence*, pages 11177–11184. AAAI Press, 2021.
- [3] John Abowd. The u.s. census bureau adopts differential privacy. pages 2867–2867, 07 2018.
- [4] Fabrizio Boninsegna and Francesco Silvestri. Differentially private release of hierarchical origin/destination data with a topdown approach, 2025.
- [5] Ferdinando Fioretto, Pascal Van Hentenryck, and Keyu Zhu. Differential privacy of hierarchical census data: An optimization approach. *Artificial Intelligence Journal*, 296:103475, 2021.
- [6] Aloni Cohen, Moon Duchin, JN Matthews, and Bhushan Suwal. Census topdown: The impacts of differential privacy on redistricting. Schloss Dagstuhl Leibniz-Zentrum für Informatik, 2021.
- [7] Aloni Cohen, Moon Duchin, JN Matthews, and Bhushan Suwal. Private Numbers in Public Policy: Census, Differential Privacy, and Redistricting. *Harvard Data Science Review*, (Special Issue 2), jun 24 2022. https://hdsr.mitpress.mit.edu/pub/954ycugm.
- [8] 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.
- [9] Ryan Steed, Tailin Liu, Zhiwei Steven Wu, and Alessandro Acquisti. Policy impacts of statistical uncertainty and privacy. *Science*, 377(6609):928–931, 2022.
- [10] Cuong Tran, Ferdinando Fioretto, Pascal Van Hentenryck, and Zhiyan Yao. Decision making with differential privacy under a fairness lens. In *International Joint Conference on Artificial Intelligence*, pages 560–566. ijcai.org, 2021.
- [11] Terrance Liu and Zhiwei Steven Wu. Private synthetic data with hierarchical structure, 2022.
- [12] Varun Chandrasekaran, Suman Banerjee, Diego Perino, and Nicolas Kourtellis. Hierarchical federated learning with privacy, 2022.
- [13] Matthew Dawson, Badih Ghazi, Pritish Kamath, Kapil Kumar, Ravi Kumar, Bo Luan, Pasin Manurangsi, Nishanth Mundru, Harikesh Nair, Adam Sealfon, and Shengyu Zhu. Optimizing hierarchical queries for the attribution reporting api, 2023.
- [14] Cynthia Dwork and Aaron Roth. The Algorithmic Foundations of Differential Privacy. 2014.
- [15] Christopher M. Bishop. Pattern Recognition and Machine Learning. Springer, New York, 2006.
- [16] Joonhyuk Ko, Juba Ziani, Saswat Das, Matt Williams, and Ferdinando Fioretto. Fairness issues and mitigations in (differentially private) socio-demographic data processes. *Proceedings of the AAAI Conference on Artificial Intelligence*, 39(27):28160–28167, Apr. 2025.
- [17] Differential Privacy Team. Learning with privacy at scale. Apple Machine Learning Journal, December 2017.
- [18] Yu-Hsuan Kuo, Cho-Chun Chiu, Daniel Kifer, Michael Hay, and Ashwin Machanavajjhala. Differentially private hierarchical count-of-counts histograms. *Proc. VLDB Endow.*, 11(11):1509–1521, July 2018.
- [19] U.S. Census Bureau. 2020 census privacy-protected microdata file (ppmf). https://www2.census. gov/programs-surveys/decennial/2020/technical-documentation/complete-tech-docs/ privacy-protected-microdata-file/2020census-privacy-protected-microdata-file.pdf, 2024. Prepared by the U.S. Census Bureau, Washington, DC.
- [20] David Van Riper, Tracy Kugler, and Jonathan Schroeder. Ipums nhgis privacy-protected 2010 census demonstration data, version 20210428 [database]. Minneapolis, MN: IPUMS, 2021. Accessed from IPUMS NHGIS.
- [21] W. Sonnenberg. Allocating grants for title i. Technical report, U.S. Department of Education, Institute for Education Sciences, 2016. Accessed: 2025-05-05.
- [22] Michel L. Balinski and H. Peyton Young. *Fair Representation: Meeting the Ideal of One Person, One Vote.* Brookings Institution Press, Washington, D.C., 2001.

A Missing Proofs

Proof of Proposition 1. Let $f(z) = \frac{\varepsilon}{2} \exp(-\varepsilon |z - N|)$ be the PDF of the Laplace centered at N with scale parameter $1/\varepsilon$. The expected value of the post-processed count is given by:

$$\mathbb{E}[\tilde{N}] = \int_{-\infty}^{\infty} \max(0, z) f(z) dz$$
$$= \int_{-\infty}^{0} 0 \cdot f(z) dz + \int_{0}^{N} z f(z) dz + \int_{N}^{\infty} z f(z) dz.$$
(9)

We now compute each term in (9) separately:

$$\int_{-\infty}^{0} 0 \cdot f(z) \, dz = 0, \tag{10}$$

$$\int_0^N zf(z) \, dz = \frac{1}{2}(N - \frac{1}{\varepsilon}) + \frac{1}{2\varepsilon} \exp(-\varepsilon N),\tag{11}$$

$$\int_{N}^{\infty} zf(z) dz = \frac{1}{2} \left(N + \frac{1}{\varepsilon} \right). \tag{12}$$

Combining equations (10)–(12), we obtain:

$$\mathbb{E}[\tilde{N}] = N + \frac{1}{2\varepsilon} \exp(-\varepsilon N).$$

Thus, the bias of \tilde{N} is:

$$\operatorname{Bias}(\tilde{N}) = \mathbb{E}[\tilde{N}] - N = \frac{1}{2\varepsilon} \exp(-\varepsilon N) > 0.$$

Proof of Proposition 2. From Proposition 1, the expected value is:

$$\mathbb{E}[\tilde{N}] = N + \frac{1}{2\varepsilon} e^{-\varepsilon N}.$$

Let the PDF of the Laplace centered at N with the scale of $1/\varepsilon$ be:

$$f(z) = \frac{\varepsilon}{2} \exp(-\varepsilon |z - N|).$$

We split the second moment:

$$\mathbb{E}[\tilde{N}^2] = \int_{-\infty}^0 0^2 f(z) \, dz + \int_0^N z^2 f(z) \, dz + \int_N^\infty z^2 f(z) \, dz.$$

Region 1: $z \in [0, N]$

$$\int_0^N z^2 f(z) \, dz = \frac{\varepsilon}{2} \int_0^N z^2 e^{-\varepsilon(N-z)} dz = \frac{\varepsilon}{2} e^{-\varepsilon N} \int_0^N z^2 e^{\varepsilon z} dz.$$

Let $u = \varepsilon z$, so $dz = \frac{du}{\varepsilon}$, the limits become 0 to εN :

$$= \frac{1}{2\varepsilon^2} e^{-\varepsilon N} \int_0^{\varepsilon N} u^2 e^u du.$$

Using the identity:

$$\int_0^x u^2 e^u du = (x^2 - 2x + 2)e^x - 2,$$

we get:

$$\int_0^N z^2 f(z) \, dz = \frac{1}{2\varepsilon^2} \left(\varepsilon^2 N^2 - 2\varepsilon N + 2 - 2e^{-\varepsilon N} \right).$$

Region 2: $z \in [N, \infty)$

Let u = z - N, so:

$$\int_{N}^{\infty} z^2 f(z) dz = \frac{\varepsilon}{2} \int_{0}^{\infty} (u+N)^2 e^{-\varepsilon u} du = \frac{\varepsilon}{2} \int_{0}^{\infty} (u^2 + 2Nu + N^2) e^{-\varepsilon u} du$$

Expanding:

$$=\frac{\varepsilon}{2}\left(\int_0^\infty u^2 e^{-\varepsilon u} du + 2N\int_0^\infty u e^{-\varepsilon u} du + N^2\int_0^\infty e^{-\varepsilon u} du\right).$$

Using standard integrals:

$$\int_0^\infty e^{-\varepsilon u} du = \frac{1}{\varepsilon}, \quad \int_0^\infty u e^{-\varepsilon u} du = \frac{1}{\varepsilon^2}, \quad \int_0^\infty u^2 e^{-\varepsilon u} du = \frac{2}{\varepsilon^3},$$

$$= \frac{1}{2\varepsilon^2} \left(2 + 2\varepsilon N + \varepsilon^2 N^2 \right).$$

Combine both parts:

$$\mathbb{E}[\tilde{N}^2] = \frac{1}{\varepsilon^2} \left(\varepsilon^2 N^2 + 2 - e^{-\varepsilon N} \right).$$

Now compute the variance:

$$\operatorname{Var}(\tilde{N}) = \mathbb{E}[\tilde{N}^2] - \left(\mathbb{E}[\tilde{N}]\right)^2.$$

Recall:

$$\mathbb{E}[\tilde{N}] = N + \frac{1}{2\varepsilon} e^{-\varepsilon N} \Rightarrow \left(\mathbb{E}[\tilde{N}]\right)^2 = N^2 + \frac{1}{\varepsilon} N e^{-\varepsilon N} + \frac{1}{4\varepsilon^2} e^{-2\varepsilon N}.$$

Therefore:

$$\begin{split} \mathrm{Var}(\tilde{N}) &= \frac{1}{\varepsilon^2} \left(\varepsilon^2 N^2 + 2 - e^{-\varepsilon N} \right) - \left(N^2 + \frac{1}{\varepsilon} N e^{-\varepsilon N} + \frac{1}{4\varepsilon^2} e^{-2\varepsilon N} \right) \\ &= \frac{1}{\varepsilon^2} \left(2 - e^{-\varepsilon N} \right) - \frac{N}{\varepsilon} e^{-\varepsilon N} - \frac{1}{4\varepsilon^2} e^{-2\varepsilon N}. \end{split}$$

Proof of Proposition 3. Let

$$f(\varepsilon) = \frac{1}{\varepsilon^2} \left(2 - e^{-\varepsilon N}\right) - \frac{N}{\varepsilon} e^{-\varepsilon N}.$$

We show that $f''(\varepsilon) > 0$ for all $\varepsilon > 0$. First, compute the first derivative:

$$f'(\varepsilon) = \frac{N^2 e^{-\varepsilon N} \varepsilon^2 + 2N e^{-\varepsilon N} \varepsilon + 2e^{-\varepsilon N} - 4}{\varepsilon^3}.$$

Then, compute the second derivative:

$$f''(\varepsilon) = -\frac{N^3 e^{-\varepsilon N} \varepsilon^3 + 3 \left(N^2 e^{-\varepsilon N} \varepsilon^2 + 2N e^{-\varepsilon N} \varepsilon + 2e^{-\varepsilon N} - 4\right)}{\varepsilon^4}.$$

Let

$$g(\varepsilon) = -N^3 e^{-\varepsilon N} \varepsilon^3 - 3 \left(N^2 e^{-\varepsilon N} \varepsilon^2 + 2N e^{-\varepsilon N} \varepsilon + 2e^{-\varepsilon N} - 4 \right),$$

so that $f''(\varepsilon) = \frac{g(\varepsilon)}{\varepsilon^4}$. It suffices to show $g(\varepsilon) > 0$ for all $\varepsilon > 0$.

First, evaluate the function at $\varepsilon = 0$. Taking the limit,

$$\lim_{\varepsilon \to 0} g(\varepsilon) = -0 - 3(0 + 0 + 2 - 4) = 6.$$

Next, compute the derivative:

 $g'(\varepsilon) = N^4 e^{-\varepsilon N} \varepsilon^3 > 0 \quad \text{for all } \varepsilon > 0.$ Since g(0) = 6 > 0 and $g'(\varepsilon) > 0$, it follows that $g(\varepsilon) > 0$ for all $\varepsilon > 0$.

Therefore, $f''(\varepsilon) = \frac{g(\varepsilon)}{\varepsilon^4} > 0$ for all $\varepsilon > 0$, proving that f is strictly convex on $\mathbb{R}_{>0}$.

Proof of Proposition 4. Let

$$f(N) = \frac{1}{\varepsilon^2} (2 - e^{-\varepsilon N}) - \frac{N}{\varepsilon} e^{-\varepsilon N}.$$

First, evaluate the lower bound at N = 0:

$$f(0) = \frac{1}{\varepsilon^2}(2-1) - \frac{0}{\varepsilon} \cdot 1 = \frac{1}{\varepsilon^2}.$$

Now consider the limit as $N \to \infty$:

$$\lim_{N \to \infty} f(N) = \frac{1}{\varepsilon^2} \cdot 2 = \frac{2}{\varepsilon^2}.$$

To show monotonicity, compute the derivative:

$$\frac{d}{dN}f(N) = Ne^{-\varepsilon N}.$$

This is non-negative for all $N \ge 0$, and strictly positive for N > 0, which implies that f(N) is strictly increasing in N. Hence, the function is bounded below by its value at N = 0 and bounded above by its limit as $N \to \infty$. Therefore,

$$\frac{1}{\varepsilon^2} \le f(N) < \frac{2}{\varepsilon^2}.$$

Proof of Theorem 3. To show that $\varepsilon_1 \leq \varepsilon_2 \leq \cdots \leq \varepsilon_L$, it suffices to show that $\varepsilon_n \leq \varepsilon_{n+1}$ for any two consecutive levels in the hierarchy. Consider an arbitrary two-level hierarchy where the upper-level node has count N and the lower-level consists of subcounts N_i satisfying $\sum_i N_i = N$.

Define the per-node MSE as a function of $\varepsilon > 0$ and count $N \ge 0$:

$$f(\varepsilon, N) = \frac{1}{\varepsilon^2} (2 - e^{-\varepsilon N}) - \frac{N}{\varepsilon} e^{-\varepsilon N}$$

Suppose, for the sake of contradiction, that $\varepsilon_n > \varepsilon_{n+1}$ is optimal. Then, the marginal utility of privacy budget must be equal across levels:

$$\frac{\partial f(\varepsilon_n, N)}{\partial \varepsilon} = \sum_i \frac{\partial f(\varepsilon_{n+1}, N_i)}{\partial \varepsilon}.$$

Let us define:

$$g(\varepsilon,N) \triangleq \frac{\partial f(\varepsilon,N)}{\partial \varepsilon} = \frac{N^2 e^{-\varepsilon N} \varepsilon^2 + 2N e^{-\varepsilon N} \varepsilon + 2 e^{-\varepsilon N} - 4}{\varepsilon^3}.$$

First, observe that $g(\varepsilon, N)$ is strictly increasing in ε since $f(\varepsilon, N)$ is strictly convex in ε by Proposition 3. Therefore, for fixed N, we have:

$$g(\varepsilon_n, N) > g(\varepsilon_{n+1}, N).$$

Next, notice that $\frac{\partial g(\varepsilon,N)}{\partial N} = -N^2 e^{-\varepsilon N} \le 0$, so $g(\varepsilon,N)$ is monotonically decreasing in N. Furthermore, observe:

$$g(\varepsilon, 0) = -\frac{2}{\varepsilon^3}, \quad \lim_{N \to \infty} g(\varepsilon, N) = -\frac{4}{\varepsilon^3}$$

Hence, for all $N \ge 0$, we have the bound:

$$-\frac{4}{\varepsilon^3} \le g(\varepsilon, N) \le -\frac{2}{\varepsilon^3}.$$

Let k be the number of nodes at the lower level. Now, consider two cases: (1) If k = 1, then trivially $g(\varepsilon, N) = \sum_{i} g(\varepsilon, N_i)$, since there is only a single lower-level node and it is equal to N. (2) If $k \ge 2$, then using the bounds on $g(\varepsilon, N)$, we have:

$$g(\varepsilon, N) \ge -\frac{4}{\varepsilon^3} \ge k \cdot \left(-\frac{2}{\varepsilon^3}\right) \ge \sum_i g(\varepsilon, N_i),$$

since each $g(\varepsilon, N_i) \leq -\frac{2}{\varepsilon^3}$ and there are k such terms in the sum. Combining the two observations, we obtain

$$g(\varepsilon_n, N) > g(\varepsilon_{n+1}, N) \ge \sum_i g(\varepsilon_{n+1}, N_i),$$

which is a contradiction.

B Skewness Analysis



Figure 7: Bias incurred as a function of ε for different distributions of N_i , where the total sum is fixed at 100. The left plot represents the 2-region case $(N_1+N_2=100)$, and the right plot represents the 3-region case $(N_1+N_2+N_3=100)$.

In this section, we investigate how the skewness of a population distribution affects bias. Even when the total population and the number of regions are fixed, the bias introduced by non-negativity post-processing (1) can vary depending on how the population is distributed. As illustrated in Figure 7, bias decreases as the distribution becomes more uniform. In fact, the bias is minimized when the population is evenly distributed across all regions.

Theorem 4. For a fixed privacy budget ε , the bias is minimized when N_i is uniform for all $i \in \mathbb{R}$.

Proof. For a fixed ε , total population N, and number of regions |R|, minimizing the bias is equivalent to solving the following optimization problem:

$$\underset{N_i}{\text{minimize}} \sum_{i \in B} e^{-N_i} \tag{13a}$$

s.t.
$$\sum_{i \in R} N_i = N,$$
(13b)

$$N_i \ge 0, \quad \forall i \in R.$$
 (13c)

The Lagrangian for this problem is given by:

$$\mathcal{L}(N_i, \lambda, \mu_i) = \sum_{i \in R} e^{-N_i} + \lambda \left(\sum_{i \in R} N_i - N \right) - \sum_{i \in R} \mu_i N_i.$$

The necessary KKT conditions are:

A PREPRINT

1. Stationarity Condition

Rearranging, we obtain:

 $\frac{\partial \mathcal{L}}{\partial N_i} = -e^{-N_i} + \lambda - \mu_i = 0.$ $\lambda - \mu_i = e^{-N_i}.$

$$\sum_{i \in R} N_i = N, \quad N_i \ge 0, \quad \forall i \in R.$$

 $\mu_i > 0, \quad \forall i \in R.$

 $\mu_i N_i = 0, \quad \forall i \in R.$

 $\lambda = \frac{1}{e^{N_i}}.$

 $N_i = -\ln \lambda.$

3. Dual Feasibility

2. Primal Feasibility

- 4. Complementary Slackness
- If $N_i > 0$, then $\mu_i = 0$, so we obtain:

Taking the natural logarithm on both sides:

Using the primal feasibility, we have:
$$\sum_{i\in R}N_i=\sum_{i\in R}(-\ln\lambda)=|R|(-\ln\lambda)=N.$$

Solving for λ :

Substituting back:

$$N_i^* = \frac{N}{|R|}, \quad \forall i \in R.$$

 $\lambda = e^{-N/|R|}.$

Since the objective function is strictly convex over the feasible set, the optimal solution must be unique. This implies that the only possible minimizer is:

$$N_i^* = \frac{N}{|R|}, \quad \forall i \in R.$$

Thus, the optimal allocation distributes N evenly among all states $i \in R$.

This also implies that if two distributions of N_i , written in vector form as $\{N_i^B\}_{i \in R}$ and $\{N_i^A\}_{i \in R}$, satisfy $\{N_i^B\}$ majorizes $\{N_i^A\}$, then the total bias is higher under $\{N_i^B\}$. Since the function $f(x) = e^{-x}$ is convex, Karamata's inequality gives

$$\sum_{i \in R} f(N_i^A) \le \sum_{i \in R} f(N_i^B),$$

which implies greater cumulative bias in more skewed distributions.

C Limitations

We discuss two primary limitations of the paper.

First, the optimization program relies on previously released data to determine the allocation of privacy budget. If level-wise privacy budgets are also made public, then the program (6) is not differentially private unless the input counts used for computing the allocation are themselves privatized. However, as discussed in Section 4, if only the total privacy budget is released—without revealing the per-level allocation—then it is acceptable to use ground-truth statistics for budget allocation.

Second, the hierarchical post-processing method introduced in Section 3.3 assumes that lower levels typically contribute more to the total MSE under optimal allocation. While this is generally true, the contribution can vary depending on the level-wise weights. In particular, under extreme weight settings (e.g., $w_3 \approx 0.9$ in Figure 5), the upper level may contribute more to the total MSE. In such cases, a bottom-up projection approach—where higher-level counts are adjusted to match lower-level estimates—may be more appropriate than the top-down method used here.

 \square

D Additional Experiments on Other States

500 Optimal (No Hier.) Optimal (No Hier.) 200000 Uniform (No Hier.) Uniform (No Hier.) 400 Optimal (w/ Hier.) Optimal (w/ Hier.) Variance 150000 Uniform (w/ Hier.) Uniform (w/ Hier.) ²⁰⁰ Bias 100000 50000 100 0 2.0 2.5 3.0 3.5 4.0 2.0 2.5 3.0 3.5 4.0 $\epsilon_{\rm total}$ $\varepsilon_{\rm total}$ (a) New Mexico Optimal (No Hier.) Optimal (No Hier.) 80 60000 Uniform (No Hier.) Uniform (No Hier.) Optimal (w/ Hier.) Optimal (w/ Hier.) Variance Uniform (w/ Hier.) 60 Uniform (w/ Hier.) Bias² 40000 40 20000 20 0 2.5 2.5 2.0 3.0 3.5 4.0 2.0 3.0 3.5 4.0 $\epsilon_{\rm total}$ $\varepsilon_{\rm total}$ (b) Delaware

D.1 Additional Result on Section 4.1

Figure 8: Hierarchical data release performance in **New Mexico** (top) and **Delaware** (bottom). Each row shows bias (left) and variance (right) for three-level hierarchical releases.



Figure 9: Bias² and variance for three weight functions W(p)—logarithmic $(\log(p + 1))$, linear (p), and quadratic (p^2) —in a single Census Tract in **New Mexico** (left) and **Delaware** (right).



(b) Delaware

Figure 10: Privacy budget allocation using Optimization Program (7) for New Mexico (top) and Delaware (bottom).

D.2 Additional Result on Section 5



Figure 11: Bias² and variance for three weight functions W(p)—logarithmic (log(p + 1)), linear (p), and quadratic (p^2)—in a single Census Tract in **New Mexico** (top) and **Delaware** (bottom).