Optimizing Discrete Noise Distributions for Rényi Differential Privacy

Atefeh Gilani*, Juan Felipe Gomez‡, Shahab Asoodeh§, Flavio P. Calmon†, Oliver Kosut*, Lalitha Sankar*

* School of Electrical, Computer and Energy Engineering, Arizona State University ({agilani2, okosut, lsankar}@asu.edu)

† School of Engineering and Applied Sciences, Harvard University (flavio@seas.harvard.edu)

‡ Department of Physics, Harvard University (juangomez@g.harvard.edu)

§ Department of Computing and Software, McMaster University (asoodehs@mcmaster.ca)

***Abstract*—An optimization framework is proposed to find discrete mechanisms supported on the integers that minimize the Rényi differential privacy subject to a cost constraint. The optimization problem is solved using gradient descent, and the resulting mechanisms are compared against the discrete Gaussian mechanism. Numerical results show that the optimized mechanisms outperform the discrete Gaussian with the same variance, in terms of both Rényi and (ϵ, δ) differential privacy.**

I. Introduction

Differential privacy (DP) [1] involves applying randomized mechanisms to queries based on sensitive data, so as to obscure the private information contained therein. The most popular mechanisms have been Laplacian noise [1], applied when pure-DP is desired, or Gaussian noise [2] for approximate DP (i.e., (ϵ, δ)-DP). However, as shown in [3], there is reason to believe that these mechanisms are **not** optimal for the *large composition setting* — i.e., when many iterations of the DP mechanisms are applied, such as when training a machine learning model over many iterations.

In addition, it has been observed that many implementations that generate Gaussian random variables have flaws that allow for exact reconstruction of the noise values, thereby challenging their *de facto* privacy guarantees [4]. For this reason, [5] proposed the discrete Gaussian mechanism, which is an integer-valued probability distribution with probability mass function (PMF) proportional to the Gaussian function. The restriction of noise distributions used in DP to an integer support allows for resilience against floating-point attacks while offering a better fit for applications where queries of interest are integer-valued (e.g., population counts in the US Census).

We explore other discrete mechanisms, also supported on the integers, that can achieve stronger privacy guarantees than the discrete Gaussian for the same variance. These mechanisms are derived by solving an optimization problem to find the PMF for the noise distribution. A similar approach was taken in [3], [6] to derive continuous distributions via an optimization problem based on the Kullback-Leibler divergence. Here, we directly optimize for the Rényi DP (RDP). RDP has the desirable properties that (i) it is a monotonic function of a convex function of the PMF of the noise distribution, (ii) it has good composition properties [7], and (iii) it can be transformed into approximate DP guarantees via the moments accountant [2]. Our main contributions are:

1) We formulate a finite-dimensional convex optimization problem to find noise PMFs that minimize the RDP subject to a cost constraint on the noise — typically a second moment constraint. The form of RDP requires that noise distribution have infinite support, so in order to make the optimization problem finite-dimensional, we only parameterize the PMF within a certain interval, and assume geometric tails beyond this interval.
2) We implement gradient descent to solve this convex optimization problem. Our implementation is available on GitHub [8].
3) We present numerical evidence that the optimized mechanisms can indeed outperform the discrete Gaussian. Specifically, we show that, for the same variance, the RDP at the target α (the Rényi parameter) for the optimized mechanism is strictly smaller than that of the discrete Gaussian. We also show that when RDP is converted into approximate-DP, the optimized mechanisms achieve better privacy parameters.

II. Preliminaries

We review some basic definitions and results from the DP literature. In this paper, all distributions will be discrete, so we can describe a probability measure via its probability mass function (PMF). Given a countable alphabet $\mathcal{X}$, let $\mathcal{P}(\mathcal{X})$ be the set of PMFs with support on $\mathcal{X}$. For $P, Q \in \mathcal{P}(\mathcal{X})$, the Rényi divergence of order α, for $\alpha \in (0,1) \cup (1, \infty)$, is

$$D_\alpha(P\|Q) = \frac{1}{\alpha - 1} \log \left(\sum_{x \in \mathcal{X}} P(x)^\alpha Q(x)^{1-\alpha} \right). \quad (1)$$

Let $\mathcal{D}$ be a set of possible datasets, and $\sim$ be a "neighboring" relation among elements of $\mathcal{D}$. That is, for $d, d' \in \mathcal{D}$ we write $d \sim d'$ to mean that d and d' are neighbors, which typically means that they differ in one entry. A *mechanism* is a function $\mathcal{M} : \mathcal{D} \to \mathcal{P}(\mathcal{X})$, which, for each $d \in \mathcal{M}$, selects a PMF $\mathcal{M}_d \in \mathcal{P}(\mathcal{X})$. This can be interpreted as a

conditional distribution for a random variable supported in $\mathcal{X}$ given a dataset from $\mathcal{D}$.

Definition 1. *A mechanism $\mathcal{M} : \mathcal{D} \to \mathcal{P}(\mathcal{X})$ is said to be (ϵ, δ)-DP if for all $A \subset \mathcal{X}$,*

$$\mathcal{M}_d(A) \leq e^{\epsilon}\mathcal{M}_{d'}(A) + \delta \text{ for all } d, d' \in \mathcal{D},\ d \sim d'. \quad (2)$$

For a mechanism $\mathcal{M}$, we also define the best ϵ for a given δ as $\epsilon_{\mathcal{M}}(\delta) = \inf\{\epsilon : \mathcal{M} \text{ is } (\epsilon, \delta)\text{-DP}\}$.

Definition 2. *A mechanism $\mathcal{M} : \mathcal{D} \to \mathcal{P}(\mathcal{X})$ is said to be (α, γ)-RDP if*

$$D_\alpha(\mathcal{M}_d \| \mathcal{M}_{d'}) \leq \gamma \text{ for all } d, d' \in \mathcal{D},\ d \sim d'. \quad (3)$$

For a mechanism $\mathcal{M}$, also define the best γ for a given α as

$$\gamma_{\mathcal{M}}(\alpha) = \inf\{\gamma : \mathcal{M} \text{ is } (\alpha, \gamma)\text{-RDP}\}. \quad (4)$$

For two mechanisms $\mathcal{M}^{(1)}, \mathcal{M}^{(2)}$, each outputting a variable in $\mathcal{X}$, their *composition* $\mathcal{M} : \mathcal{D} \to \mathcal{P}(\mathcal{X} \times \mathcal{X})$ is

$$\mathcal{M}_d(x_1, x_2) = \mathcal{M}_d^{(1)}(x_1)\mathcal{M}_d^{(2)}(x_2). \quad (5)$$

Theorem 1 ([7])**.** *For any mechanisms $\mathcal{M}^{(1)}, \mathcal{M}^{(2)}$ and their composition $\mathcal{M}$,*

$$\gamma_{\mathcal{M}}(\alpha) \leq \gamma_{\mathcal{M}^{(1)}}(\alpha) + \gamma_{\mathcal{M}^{(2)}}(\alpha). \quad (6)$$

The above is non-adaptive composition, in that each mechanism works independently of the other's output. In contrast, in an adaptive composition, the second mechanism may depend on the output of the first. A similar composition result holds for the adaptive setting [7], but for brevity we omit the details. A particular consequence of Theorem 1 is that, if the same (or equivalent) mechanisms are composed N_c times, then the RDP is simply multiplied by N_c.

The moments accountant [2] provides a method to derive (ϵ, δ) guarantees from (α, γ) guarantees. The most basic form of the moments accountant is as follows.

Theorem 2 ([2])**.** *For any mechanism $\mathcal{M}$,*

$$\epsilon_{\mathcal{M}}(\delta) \leq \inf_{\alpha > 1} \gamma_{\mathcal{M}}(\alpha) + \frac{\log(1/\delta)}{\alpha - 1}. \quad (7)$$

While improvements to the moments accountant have been made in [9], [10], for simplicity we use only this version.

Let $\mathbb{Z}$ be the set of integers. The *discrete Gaussian distribution*, denoted $\mathcal{N}_{\mathbb{Z}}(\mu, \sigma^2)$, for $\mu \in \mathbb{Z}$ and $\sigma > 0$, is the PMF in $\mathcal{P}(\mathbb{Z})$ given by

$$P(x) = \frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sum_{y \in \mathbb{Z}} e^{-\frac{(y-\mu)^2}{2\sigma^2}}}, \quad x \in \mathbb{Z}. \quad (8)$$

The *discrete Gaussian mechanism*, for a query function $q : \mathcal{D} \to \mathbb{Z}$, is given by $\mathcal{M}_d = \mathcal{N}_{\mathbb{Z}}(q(d), \sigma^2)$. The following facts about the discrete Gaussian mechanism are shown in [5].

Theorem 3. *If the query function q satisfies the sensitivity bound $|q(d) - q(d')| \leq s$ for all $d \sim d'$, then the discrete Gaussian mechanism with parameter σ satisfies*

$$\gamma_{\mathcal{M}}(\alpha) \leq \frac{s^2\alpha}{2\sigma^2} \quad (9)$$

with equality if αs is an integer. In addition, the discrete Gaussian mechanism satisfies the variance condition

$$\mathbb{E}_{X \sim \mathcal{M}_d}\left[(X - q(d))^2\right] \leq \sigma^2. \quad (10)$$

III. Optimized Discrete Rényi DP Mechanisms

In a discrete setting, our aim is to solve a problem of the following form:

$$\begin{aligned} &\underset{P_Z \in \mathcal{P}(\mathbb{Z})}{\text{minimize}} \ \max_{t \in \{-s, \ldots, s\}} D_\alpha\left(P_Z \| T_t P_Z\right), \\ &\text{subject to} \quad \mathbb{E}[c(Z)] \leq C. \end{aligned} \quad (11)$$

Here, we use T_t to denote the shift operator, i.e., for a real-valued function f the function $T_t f$ is defined as $(T_t f)(x) \coloneqq f(x - t)$. The parameter $s \in \mathbb{N}$, where $\mathbb{N}$ is the set of positive integers, is the sensitivity of a query, $c : \mathbb{Z} \to [0, \infty)$ is a symmetric cost function, and $C \in \mathbb{R}^+$ is an upper bound on the cost. To form a finite-dimensional optimization problem, we assume P_Z is a distribution over the integers in the following form:

$$P_Z(i) = \begin{cases} p_{|i|}, & \text{for } i \in \mathbb{Z}, \text{with } |i| \leq N \\ p_N r^{|i|-N}, & \text{for } i \in \mathbb{Z}, \text{with } |i| > N, \end{cases} \quad (12)$$

where $r \in (0, 1)$ and $N \in \mathbb{N}$. For fixed r, this distribution is parameterized by the finite-dimensional vector $p = (p_0, \ldots, p_N)$.

Theorem 4. *Fix $r \in (0, 1)$ and $N \in \mathbb{N}$, and let $c(z) = z^2$ be the quadratic cost function. Within the distribution family introduced in (12), the optimization described in (11) can be reformulated as the optimization problem in ($\star$).*

Note that the objective in ($\star$) is the quantity inside the log of the Rényi divergence — thus the RDP can be easily calculated from the optimal objective. A detailed proof is in the appendix.

IV. Numerical Results

In this section, we numerically compare the optimized discrete mechanism from ($\star$), which we henceforth refer to as an *discrete RDP mechanism*, against the discrete Gaussian mechanism. We employ (9) to obtain (α, γ)-RDP guarantees of the discrete Gaussian mechanism.

We begin by noting that (9) attains equality when αs is an integer. For the experiments presented here, we meticulously select our parameters to satisfy this constraint, i.e., αs is an integer. We use Theorem 2 to derive (ϵ, δ)-DP guarantees from (α, γ)-RDP guarantees. To quantify utility, we focus on the variance of distributions. Our results demonstrate that discrete RDP mechanisms can outperform discrete Gaussian mechanisms with the same variance[1] in terms of both (α, γ)-RDP and (ϵ, δ)-DP guarantees.

Figure 1: Figures 1a and 1b depict the optimal distributions obtained by solving ($\star$) for various parameter sets. These

[1]While in general, the parameter σ and the standard deviation of a discrete Gaussian need not be the same, for the range of σ values we consider, they are imperceptibly close (i.e., their difference is in the range of $\approx 10^{-15}$), and hence, assumed to be the same.

$$\begin{aligned}
\underset{p}{\text{minimize}} \max_{t\in\{1,\ldots,s\}} \; & \frac{p_N r}{1-r}\left(r^{t(1-\alpha)} + r^{t\alpha}\right) + \sum_{j=-N}^{N-t} p_{|t+j|}^{\alpha}\, p_{|j|}^{1-\alpha} + p_N^{\alpha} r^{\alpha(t-N)} \sum_{j=N-t+1}^{N} r^{\alpha j}\, p_{|j|}^{1-\alpha} \\
& + p_N^{1-\alpha} r^{-N(1-\alpha)} \sum_{j=-t-N}^{-N-1} p_{|t+j|}^{\alpha}\, r^{-j(1-\alpha)} \\
\text{subject to} \quad & 2\sum_{i=1}^{N-1} p_i\, i^2 + \frac{2p_N\left(-r^2(N-1)^2 + N^2(2r-1) - r\right)}{(r-1)^3}\, i^2 \le C, \\
& p_0 + 2\sum_{j=1}^{N-1} p_j + \frac{2\,p_N}{1-r} = 1, \\
& p_i \in (0,1) \quad \text{for all} \quad i \in \{0,\ldots,N\}.
\end{aligned} \tag{$\star$}$$

plots illustrate that the optimal distributions may exhibit either monotonic or non-monotonic behavior.

Figure 2a: This figure displays the RDP guarantees of the discrete RDP mechanism for $\alpha = 2$ optimized for different values of sensitivity s. The ratio σ/s is kept fixed for each curve in this plot. Similar to the (α,γ)-RDP guarantees of discrete Gaussian mechanisms, the (α,γ)-RDP guarantees of our mechanisms also depend almost entirely on the σ/s ratio rather than on σ and s individually.

Figure 2b: This figure compares the (α,γ)-RDP guarantees of a discrete Gaussian mechanism with a variance of 16 (parameter $\sigma = 4$ to very high precision) to a discrete RDP mechanism optimized for the same variance and a range of α values. Observe that the two achieve roughly the same performance for α close to 1, but the RDP mechanism significantly outperforms the discrete Gaussian for larger α.

Figure 3: Figure 3a compares $(\epsilon, \delta = 10^{-8})$-DP guarantees of discrete Gaussian and RDP mechanisms across different σ/s values for 10 compositions. To accomplish this, for a fixed σ/s, we first determine the optimal α^* and ϵ for a discrete Gaussian mechanism according to Theorem 2. Note that if we substitute the upper bound (which is tight for integer αs) provided in Theorem 3 into Theorem 2, the resulting expression becomes convex in α. Consequently, the optimal α can be easily determined by setting the derivative to zero, resulting in

$$\alpha^* = \sqrt{\frac{2\log 1/\delta}{N_c}}\frac{\sigma}{s} + 1, \tag{13}$$

where N_c is the number of compositions. Then, we optimize the RDP noise for that α^* and σ/s to obtain (α^*,γ)-guarantees and simply utilize Theorem 2 (without performing the minimization) to translate these guarantees into (ϵ,δ)-DP guarantees. We then use the same settings to create the relative change plot, as shown in Figure 3b. For a fixed σ/s ratio, the percent relative change is computed as $100 \times \left(1 - \frac{\epsilon \text{ of discrete RDP mechanism}}{\epsilon \text{ of discrete Gaussian mechanism}}\right)$.

V. Concluding Remarks

We have presented a framework for determining the optimal additive-noise discrete mechanism minimizing Rényi DP subject to a cost constraint. The resulting discrete RDP mechanism is in general non-monotonic, and is akin to the discrete Gaussian mechanism in that its privacy performance primarily depends on the sensitivity and variance through the ratio of the two quantities.

Another key take-away of our work is that the optimized discrete RDP mechanism assures better privacy than the discrete Gaussian, quantified via (i) RDP itself, and (ii) the resulting ϵ for a chosen δ.

The discrete Gaussian is presently the mechanism of choice in the top-down private Census summary statistics release algorithm [11], [12]. Our results presented here make a compelling case for the discrete RDP as an alternative and we hope to explore this further.

References

[1] C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," in *Proc. Theory of Cryptography (TCC)*, Berlin, Heidelberg, 2006, pp. 265–284.

[2] M. Abadi, A. Chu, I. Goodfellow, H. B. McMahan, I. Mironov, K. Talwar, and L. Zhang, "Deep learning with differential privacy," in *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security*, 2016, pp. 308–318.

[3] W. Alghamdi, S. Asoodeh, F. P. Calmon, O. Kosut, L. Sankar, and F. Wei, "Cactus mechanisms: Optimal differential privacy mechanisms in the large-composition regime," in *2022 IEEE International Symposium on Information Theory (ISIT)*, 2022, pp. 1838–1843.

[4] I. Mironov, "On significance of the least significant bits for differential privacy," in *Proceedings of the 2012 ACM Conference on Computer and Communications Security*, ser. CCS '12. New York, NY, USA: Association for Computing Machinery, 2012, p. 650–661. [Online]. Available: https://doi.org/10.1145/2382196.2382264

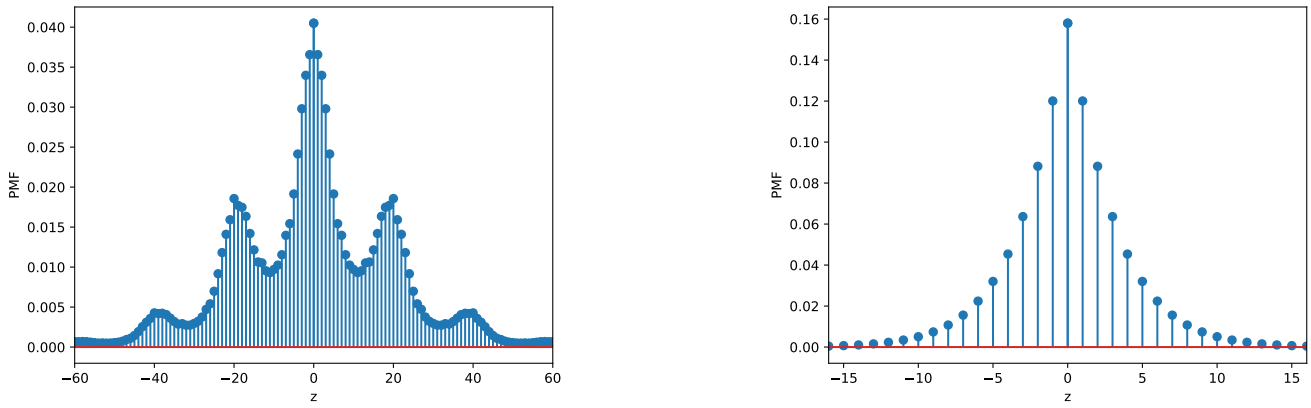
[5] C. L. Canonne, G. Kamath, and T. Steinke, "The discrete Gaussian for differential privacy," in *Proceedings of the 34th International Conference on Neural Information Processing Systems*, ser. NIPS '20. Red Hook, NY, USA: Curran Associates Inc., 2020.

[6] W. Alghamdi, S. Asoodeh, F. P. Calmon, J. Felipe Gomez, O. Kosut, and L. Sankar, "Optimal multidimensional differentially private mechanisms in the large-composition regime," in *2023 IEEE International Symposium on Information Theory (ISIT)*, 2023, pp. 2195–2200.

[7] I. Mironov, "Rényi differential privacy," in *Proc. IEEE Comp. Security Foundations Symp. (CSF)*, 2017, pp. 263–275.

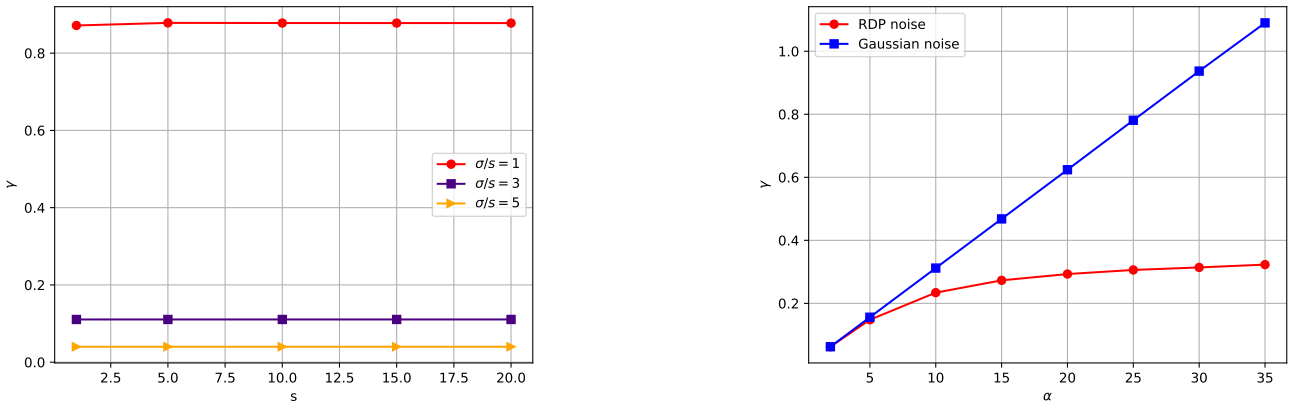
[8] "Renyi DP mechanism design," May 2024, https://github.com/SankarLab/Renyi-DP-Mechanism-Design.

[9] B. Balle, G. Barthe, M. Gaboardi, J. Hsu, and T. Sato, "Hypothesis testing interpretations and Rényi differential privacy," in *Int. Conf. Art. Intelligence and Stat. (AISTAT)*, 2020, pp. 2496–2506.



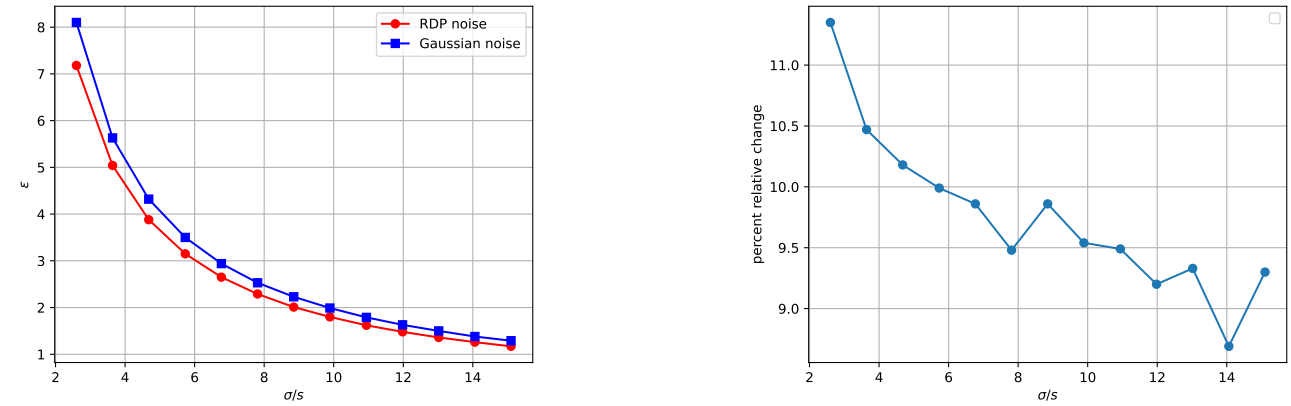
(a) (b)

Fig. 1: Figures 1a and 1b display the optimal distributions obtained by solving ($\star$) for the parameter sets $(s = 20, C = 400, N = 120, \alpha = 2, r = 0.9)$ and $(s = 1, C = 16, N = 22, \alpha = 35, r = 0.9)$, respectively.



(a) (b)

Fig. 2: Figure 2a illustrates the variation in $(\alpha = 2, \gamma)$-RDP guarantees across discrete RDP mechanisms optimized for different s values, each with a fixed σ/s ratio. Figure 2b illustrates the comparison of (α, γ)-RDP guarantees between discrete Gaussian and optimized RDP mechanisms, all with the same variance of 16, across various α values.



(a) (b)

Fig. 3: Figure 3a compares $(\epsilon, \delta = 10^{-8})$-DP guarantees of discrete Gaussian and RDP mechanisms for different σ/s values with $s = 1$, for 10 compositions. Figure 3b illustrates the percent relative change for the same setting as Fig. 3a.

[10] S. Asoodeh, J. Liao, F. P. Calmon, O. Kosut, and L. Sankar, "Three variants of differential privacy: Lossless conversion and applications," *IEEE Journal on Selected Areas in Information Theory*, vol. 2, no. 1, pp. 208–222, 2021.
[11] C. Willner, "An overview of differential privacy in the 2020 decennial census," May 2024, https://www.ctdata.org/blog/an-overview-of-differential-privacy-in-the-2020-decennial-census.
[12] S. Haney, W. Sexton, A. Machanavajjhala, M. Hay, and G. Miklau, "Differentially private algorithms for 2020 census detailed DHC race & ethnicity," *arXiv preprint arXiv:2107.10659*, 2021.

APPENDIX

PROOF OF THEOREM 4

We have

$$D_\alpha\left(P_Z \| T_a P_Z\right) = \frac{1}{\alpha-1}\log\sum_{i\in\mathbb{Z}} P_Z(i)^\alpha P_Z(i-a)^{1-\alpha} = \frac{1}{\alpha-1}\log\sum_{i\in\mathbb{Z}} p_i^\alpha\ p_{i-a}^{1-\alpha}, \tag{14}$$

where $p_{i-a} = p_N r^{|i-a|-N}$ for $i \geq a+N$ or $i \leq -N+a$. Let focus on the quantity inside the logarithm:

$$\sum_{i\in\mathbb{Z}} p_i^\alpha\ p_{i-a}^{1-\alpha} \tag{15}$$

$$= \sum_{i=-\infty}^{\min\{-N,-N+a\}-1} p_i^\alpha\ p_{i-a}^{1-\alpha} + \sum_{i=\min\{-N,-N+a\}}^{\max\{N,N+a\}} p_i^\alpha\ p_{i-a}^{1-\alpha} + \sum_{i=\max\{N,N+a\}+1}^{\infty} p_i^\alpha\ p_{i-a}^{1-\alpha}. \tag{16}$$

We have

$$\sum_{i=-\infty}^{\min\{-N,-N+a\}-1} p_i^\alpha\ p_{i-a}^{1-\alpha} = \sum_{i=-\infty}^{\min\{-N,-N+a\}-1} \left(p_N r^{-i-N}\right)^\alpha \left(p_N r^{-i+a-N}\right)^{1-\alpha} \tag{17}$$

$$= \sum_{i=-\infty}^{\min\{-N,-N+a\}-1} p_N r^{-\alpha a - i + a - N} \tag{18}$$

$$= p_N\ r^{a(1-\alpha)}\frac{r^{\max\{0,-a\}+1}}{1-r} \tag{19}$$

Similarly, we have

$$\sum_{i=\max\{N,N+a\}+1}^{\infty} p_i^\alpha\ p_{i-a}^{1-\alpha} = \sum_{i=\max\{N,N+a\}+1}^{\infty} \left(p_N r^{i-N}\right)^\alpha \left(p_N r^{i-a-N}\right)^{1-\alpha} \tag{20}$$

$$= \sum_{i=\max\{N,N+a\}+1}^{\infty} p_N r^{i-a-N+\alpha a} \tag{21}$$

$$= p_N\ r^{-a(1-\alpha)}\frac{r^{\max\{0,a\}+1}}{1-r} \tag{22}$$

Therefore,

$$\sum_{i=-\infty}^{\min\{-N,-N+a\}-1} p_i^\alpha\ p_{i-a}^{1-\alpha} + \sum_{i=\max\{N,N+a\}+1}^{\infty} p_i^\alpha\ p_{i-a}^{1-\alpha} \tag{23}$$

$$= p_N\ r^{a(1-\alpha)}\frac{r^{\max\{0,-a\}+1}}{1-r} + p_N\ r^{-a(1-\alpha)}\frac{r^{\max\{0,a\}+1}}{1-r} \tag{24}$$

$$= \frac{p_N r}{1-r}\left(r^{a(1-\alpha)+\max\{0,-a\}} + r^{-a(1-\alpha)+\max\{0,a\}}\right) \tag{25}$$

$$= \begin{cases} \frac{p_N r}{1-r}\left(r^{a(1-\alpha)} + r^{-a(1-\alpha)+a}\right) & \text{if } a \geq 0 \\ \frac{p_N r}{1-r}\left(r^{a(1-\alpha)-a} + r^{-a(1-\alpha)}\right) & \text{if } a < 0 \end{cases} \tag{26}$$

$$= \begin{cases} \frac{p_N r}{1-r}\left(r^{a(1-\alpha)} + r^{a\alpha}\right) & \text{if } a \geq 0 \\ \frac{p_N r}{1-r}\left(r^{-a\alpha} + r^{-a(1-\alpha)}\right) & \text{if } a < 0 \end{cases} \tag{27}$$

$$= \frac{p_N r}{1-r}\left(r^{|a|(1-\alpha)} + r^{|a|\alpha}\right). \tag{28}$$

Now we simplify the middle term in (15) as follows:

$$\sum_{i=\min\{-N,-N+a\}}^{\max\{N,N+a\}} p_i^\alpha\ p_{i-a}^{1-\alpha} \tag{29}$$

$$= \sum_{j=\min\{-a,0\}-N}^{\max\{-a,0\}+N} p_{j+a}^\alpha\ p_j^{1-\alpha} \tag{30}$$

$$= \sum_{j=-N}^{N} p_{|a|+j}^\alpha\ p_j^{1-\alpha} + \left[p_N^{1-\alpha} r^{-N(1-\alpha)} \sum_{j=-|a|-N}^{-N-1} p_{|a|+j}^\alpha\ r^{-j(1-\alpha)}\right] \mathbb{1}\{a \neq 0\} \tag{31}$$

$$= \sum_{j=-N}^{N-|a|} p_{|a|+j}^\alpha\ p_j^{1-\alpha} + p_N^\alpha r^{\alpha(|a|-N)} \sum_{j=N-|a|+1}^{N} r^{\alpha j}\ p_j^{1-\alpha} + \left[p_N^{1-\alpha} r^{-N(1-\alpha)} \sum_{j=-|a|-N}^{-N-1} p_{|a|+j}^\alpha\ r^{-j(1-\alpha)}\right] \mathbb{1}\{a \neq 0\} \tag{32}$$

The first equality follows from a change of variable where $i = j + a$.

So, overall, the quantity inside the logarithm of Rényi divergence simplifies to

$$\frac{p_N r}{1-r}\left(r^{|a|(1-\alpha)} + r^{|a|\alpha}\right) + \sum_{j=-N}^{N-|a|} p_{|a|+j}^\alpha\ p_j^{1-\alpha} + p_N^\alpha r^{\alpha(|a|-N)} \sum_{j=N-|a|+1}^{N} r^{\alpha j}\ p_j^{1-\alpha} + \left[p_N^{1-\alpha} r^{-N(1-\alpha)} \sum_{j=-|a|-N}^{-N-1} p_{|a|+j}^\alpha\ r^{-j(1-\alpha)}\right] \mathbb{1}\{a \neq 0\} \tag{33}$$

This quantity is symmetric in a; so, $\max_{a\in\{-s,\cdots,s\}}$ collapses to $\max_{a\in\{0,\cdots,s\}}$ and $|a|$ simplifies to a. Moreover, Rényi divergence attains its minimum, which is zero, when $a = 0$, so let us exclude $a = 0$ from the set $\{0, \cdots, s\}$. So, the task

of finding the worst-case shift of Rényi divergence collapses to the following optimization:

$$\max_{a\in\{1,\cdots,s\}} \frac{p_N r}{1-r}\left(r^{a(1-\alpha)}+r^{a\alpha}\right) + \sum_{j=-N}^{N-a} p_{a+j}^{\alpha}\, p_j^{1-\alpha} + p_N^{\alpha} r^{\alpha(a-N)} \sum_{j=N-a+1}^{N} r^{\alpha j}\, p_j^{1-\alpha} + p_N^{1-\alpha} r^{-N(1-\alpha)} \sum_{j=-a-N}^{-N-1} p_{a+j}^{\alpha}\, r^{-j(1-\alpha)} \tag{34}$$

Now we simplify the normalization constraint as follows:

$$1 = \sum_{i\in\mathbb{Z}} p_i \tag{35}$$

$$= p_0 + 2\sum_{i=1}^{N-1} p_i + 2\sum_{i=N}^{\infty} p_N\ r^{i-N} \tag{36}$$

$$= p_0 + 2\sum_{i=1}^{N-1} p_i + \frac{2p_N}{1-r}. \tag{37}$$

Let $c(i) = i^2$, the cost constraint simplifies to

$$C \geq \mathbb{E}[c(i)] = \sum_{i\in\mathbb{Z}} p_i\ i^2 = 2\sum_{i=1}^{N-1} p_i\ i^2 + 2\sum_{i=N}^{\infty} p_N\ r^{i-N}\ i^2, \tag{38}$$

where

$$\sum_{i=N}^{\infty} r^{i-N}\ i^2 = \frac{-r^2(N-1)^2 + N^2(2r-1) - r}{(r-1)^3}. \tag{39}$$