Prior-itizing Privacy: A Bayesian Approach to Setting the Privacy Budget in Differential Privacy

Zeki Kazan and Jerome P. Reiter

April 23, 2023

Abstract

When releasing outputs from confidential data, agencies need to balance the analytical usefulness of the released data with the obligation to protect data subjects' confidentiality. For releases satisfying differential privacy, this balance is reflected by the parameter ε, known as the privacy budget. In practice, it can be difficult for agencies to select and interpret ε. We use Bayesian posterior probabilities of disclosure to provide a framework for setting ε. The agency decides how much posterior risk it is willing to accept in a data release at various levels of prior risk. Using a mathematical relationship among these probabilities and ε, the agency selects the maximum ε that ensures the posterior-to-prior ratios are acceptable for all values of prior disclosure risk. The framework applies to any differentially private mechanism.

1 Introduction

Differential privacy (DP) [8] is a gold standard definition of what it means to protect individuals' confidentiality when releasing sensitive data. DP is used by large tech companies, like Google [10] and Meta [25], both internally and for public data releases. A variant of DP is used by the U. S. Census Bureau for the release of 2020 redistricting data [1] and for the release of the 2020 demographic and housing characteristics file.

The confidentiality guarantee of DP is determined principally by an agency-selected parameter, typically referred to as the privacy budget ε. Smaller values of ε generally imply greater confidentiality protection. However, smaller values of ε also typically inject more noise into the released data, which can degrade the accuracy of analyses of the disclosure-protected data products. Thus, agencies must choose ε to balance confidentiality protection with analytical usefulness. This balancing act has resulted in a wide range of values of ε in practice. For example, early advice in the field recommends considering ε of "0.01, 0.1, or in some cases, log(2) or log(3)" [7], whereas recent large-scale implementations use values like $\varepsilon = 8.6$ in OnTheMap [23], $\varepsilon = 14$ in Apple's use of local DP for iOS 10.1.1 [29], and an equivalent of $\varepsilon = 17.14$ in the 2020 decennial census redistricting data release [1].

To navigate this trade off, decision makers inside agencies can benefit from familiar interpretations of the confidentiality protection guarantee afforded by ε. In particular, this can help agencies justify the choice of ε and lead to satisfactory trade offs of confidentiality protection and analytical usefulness. One familiar and interpretable quantity is the Bayesian posterior probability that an adversary learns sensitive information from the released data [6, 11, 16, 26]. Fortunately, posterior probabilities can be related to the randomness in algorithms that satisfy DP [2, 18, 19, 21, 24].

We propose that agencies utilize this relationship to select values of ε that accord with their desired confidentiality guarantees. The basic idea is as follows. First, the agency constructs a function that summarizes the maximum posterior probability of disclosure permitted for any prior probability of disclosure. For example, for a prior risk of 0.001, the agency may be comfortable with a ten-fold (or more) increase in the ratio of posterior risk to prior risk, whereas for a prior risk of 0.4, the agency may require the posterior-to-prior ratio not to exceed, say, 1.2. Second, for each prior risk value, the agency converts the posterior-to-prior ratio into the largest ε that still ensures the ratio is satisfied. Third, the agency selects the smallest ε among these values, using that value for the data release. Importantly, the agency does not use the confidential data in these computations—they are theoretical and data free—so that the selection of ε does not use privacy budget or require access to representative test data.

Our main contributions are as follows.

- We propose a framework for selecting ε that applies to any differentially private mechanism, does not use additional privacy budget, and can account for disclosure risk from both an individual's inclusion in the data and the sensitivity of the values in the data.
- We enable agencies to tune the choice of ε to achieve a posterior-to-prior risk profile. This can avoid setting ε unnecessarily small if, for example, the agency tolerates larger posterior-to-prior ratios for certain prior risks.
- We give complete theoretic justification for the framework and derive closed-form solutions for the ε implied by a range of risk profiles. For more complex risk profiles, we also provide a general form for ε as a minimization problem.

The remainder of this article is organized as follows. In Section 2, we describe notation and relevant definitions. In Section 3, we illustrate the approach and outline the choices a practitioner must make under the framework. In Section 4, we prove that our method for setting ε bounds the disclosure risk as desired. In Section 5, we compare to related methods from the literature on DP. Finally, in Section 6, we provide some concluding remarks. To streamline the discussion, throughout we focus on the release of discrete-valued statistics computed on discrete-valued data; extension to continuous-valued statistics and data is straightforward.

2 Background

We first describe differential privacy, followed by Bayesian probabilities of disclosure.

2.1 Differential Privacy

Let $\mathbf{P}$ represent a population of individuals. The agency has a subset of $\mathbf{P}$, which we call $\mathbf{Y}$, comprising n individuals measured on d variables. For any individual i, let Y_i be the length-d vector of values corresponding to individual i, and let $I_i = 1$ when individual i is in $\mathbf{Y}$ and $I_i = 0$ otherwise. For all i such that $I_i = 1$, let $\mathbf{Y}_{-i}$ be the $(n-1) \times d$ matrix of values for the $n-1$ individuals in $\mathbf{Y}$ excluding individual i.[1] The agency possessing the data—henceforth referred to as the data holder—wishes to release some function of the data, $T(\mathbf{Y})$. We assume $\mathbf{Y}$ and $T(\mathbf{Y})$ each have discrete support but may be many-dimensional. The data holder turns to DP and will release $T^*(\mathbf{Y})$, a noisy version of $T(\mathbf{Y})$ under ε-DP. We use the following definition of ε-DP.[2]

[1] For all i such that $I_i = 0$ we let $\mathbf{Y}_{-i}$ be an $(n-1) \times d$ matrix of individuals not including individual i.

[2] In Defintion 1, implicitly, for all individuals j such that $j \neq i$, the values of I_j in the numerator and denominator in (1) do not differ.

Definition 1 (ε-Differential Privacy)**.** *For a dataset* $\mathbf{Y} \subset \mathbf{P}$ *and* $\varepsilon \in (0, \infty)$*, a function* $T^*(\mathbf{Y})$ *satisfies* ε*-differential privacy if for all* i*, all* y *in the support of* Y_i*, all* $\mathbf{y}_{-i}$ *in the support of* $\mathbf{Y}_{-i}$*, and all* t^* *in the support of* $T^*(\mathbf{Y})$*,*

$$e^{-\varepsilon} \leq \frac{P[T^*(\mathbf{Y}) = t^* \mid Y_i = y, I_i = 1, \mathbf{Y}_{-i} = \mathbf{y}_{-i}]}{P[T^*(\mathbf{Y}_{-i}) = t^* \mid I_i = 0, \mathbf{Y}_{-i} = \mathbf{y}_{-i}]} \leq e^{\varepsilon}. \tag{1}$$

This DP definition, involving data with and without individual i, is referred to as unbounded DP. An alternative DP definition replaces the denominator in (1) by $P[T^*(\mathbf{Y}) = t^* \mid Y_i = y', I_i = 1, \mathbf{Y}_{-i} = \mathbf{y}_{-i}]$ for y' in the support of Y_i, that is, changing Y_i only; this is referred to as bounded DP. A mechanism satisfying unbounded DP also satisfies bounded DP, with the ε increased by a factor of 2. See [20] for more details on bounded and unbounded DP.

Commonly-used mechanisms satisfying DP involve releasing the sum of the statistic of interest and noise randomly sampled from an appropriate distribution. These noise distributions are centered at zero and have variance inversely proportional to ε. A commonly used mechanism in settings where the statistic of interest is a count is the geometric mechanism [14].

Definition 2 (Geometric Mechanism)**.** *Let* $T(\mathbf{Y}) \in \mathbb{Z}$ *be a count statistic and suppose we wish to release a noisy count* $T^*(\mathbf{Y}) \in \mathbb{Z}$ *satisfying* ε*-DP. The geometric mechanism produces a count centered at* T *with noise from a two-sided geometric distribution with parameter* $e^{-\varepsilon}$*. That is,*

$$P[T^*(\mathbf{Y}) = t^* \mid T(\mathbf{Y}) = t] = \frac{1 - e^{-\varepsilon}}{1 + e^{-\varepsilon}} e^{-\varepsilon |t^* - t|}, \qquad t^* \in \mathbb{Z}. \tag{2}$$

It is straightforward to show that under the Geometric Mechanism, the variance of $T^*(\mathbf{Y})$ is

$$\mathsf{V}[T^*(\mathbf{Y}) \mid T(\mathbf{Y}) = t] = \frac{2e^{-\varepsilon}}{(1 - e^{-\varepsilon})^2}. \tag{3}$$

The process for choosing ε has received scant attention in the literature [5, 27, 28]. Prior work focuses on either (i) scenarios where the data have yet to be collected, and the goal is to simultaneously select ε and determine how much to compensate individuals for their loss in privacy [4, 12, 17, 22], or (ii) settings where the population is already public information, and the goal is to protect which subset of individuals is included in a release [21]. We focus on the common setting where data have already been collected and the population they are drawn from is not public information.

2.2 Bayesian Measures of Disclosure Risk

Consider an adversary who desires to learn about some particular individual i in $\mathbf{Y}$ using the release of $T^*(\mathbf{Y})$. We assume the release mechanism for $T^*(\mathbf{Y})$ is known to the adversary. We suppose the adversary has a model, $\mathcal{M}$, for making predictions about Y_i based on auxiliary information about individual i, which does not directly use the confidential data. For example, the adversary could make $\mathcal{M}$ based on proprietary information or data from sources like administrative records. We require that the DP release mechanism does not depend on $\mathcal{M}$ and that, under $\mathcal{M}$, the observations are independent but not necessarily identically distributed. These conditions are formalized in Section 4. For our ultimate purpose, i.e., helping data holders set ε, the exact form of the adversary's $\mathcal{M}$ is immaterial. In fact, as we shall discuss, we are not concerned whether the adversary's predictions from $\mathcal{M}$ are highly accurate or completely awful.

On a technical note, we make the distinction that the data holder views $\mathbf{Y}$ and I_i as fixed quantities, since it knows which rows are in the collected data and what values are associated to each row. The adversary, however, views $\mathbf{Y}$ and I_i as random variables, and thus probabilistic statements about these quantities are well defined from the adversary's perspective. Notationally, we signify that a probabilistic statement is from the adversary's perspective by conditioning on the adversary's model, $\mathcal{M}$.

Let $\mathcal{S}$ be the subset of the support of Y_i that the data holder considers a privacy violation. For example, if $d = 1$ and $\mathbf{Y}$ is income data, then $\mathcal{S}$ may be the set of possible incomes within 5,000 or within 5% of the true income for individual i. If $d = 1$ and $\mathbf{Y}$ is binary, then $\mathcal{S}$ is a subset of $\{0, 1\}$. The selection of $\mathcal{S}$ must not depend on $\mathbf{P}$, as this might constitute a privacy violation.

The data holder may be concerned about the risk that the adversary determines individual i is in $\mathbf{Y}$ or the risk that the adversary makes a disclosure for individual i; that is, $I_i = 1$ and $Y_i \in \mathcal{S}$, respectively. Assuming that the adversary's model puts nonzero probability mass on these events, we can express their relevant prior probabilities as follows.

$$P[I_i = 1 \mid \mathcal{M}] = p_i, \qquad P[Y_i \in \mathcal{S} \mid I_i = 1, \mathcal{M}] = q_i. \tag{4}$$

For fixed p_i and q_i, we can measure the risk of disclosure for individual i in a number of ways. Drawing from [24], one measure of the risk to individual i is the relative disclosure risk, $r_i(p_i, q_i, t^*)$. Writing the noisy statistic as T^* and suppressing the dependence on $\mathbf{Y}$ or $\mathbf{Y}_{-i}$, this is defined as follows.

Definition 3 (Relative Disclosure Risk). *For fixed data $\mathbf{Y}$, individual i, prior model $\mathcal{M}$, and released $T^* = t^*$, the relative disclosure risk is the posterior-to-prior risk ratio,*

$$r_i(p_i, q_i, t^*) = \frac{P[Y_i \in \mathcal{S}, I_i = 1 \mid T^* = t^*, \mathcal{M}]}{P[Y_i \in \mathcal{S}, I_i = 1 \mid \mathcal{M}]}. \tag{5}$$

For interpretation, note that the relative risk can be decomposed into the posterior-to-prior ratio from inclusion (I_i) and the posterior-to-prior ratio from the values (Y_i). We have

$$r_i(p_i, q_i, t^*) = \frac{P[Y_i \in \mathcal{S} \mid I_i = 1, T^* = t^*, \mathcal{M}]}{P[Y_i \in \mathcal{S} \mid I_i = 1, \mathcal{M}]} \cdot \frac{P[I_i = 1 \mid T^* = t^*, \mathcal{M}]}{P[I_i = 1 \mid \mathcal{M}]}. \tag{6}$$

The relative risk, however, does not tell the full story. The data holder also may care about absolute disclosure risks, $a_i(p_i, q_i, t^*)$ [16].

Definition 4 (Absolute Disclosure Risk). *For fixed data $\mathbf{Y}$, individual i, prior model $\mathcal{M}$, and released $T^* = t^*$, the absolute disclosure risk is the posterior probability,*

$$a_i(p_i, q_i, t^*) = P[Y_i \in \mathcal{S}, I_i = 1 \mid T^* = t^*, \mathcal{M}]. \tag{7}$$

Since $r_i(p_i, q_i, t^*) = a_i(p_i, q_i, t^*)/(p_i q_i)$, it is straightforward to convert between these risk measures.

3 Using Posterior-to-prior Risks for Setting ε

The quantities from Section 2 can inform the choice of ε. For example, it has been shown that DP implies that for all p_i, q_i, t^*,

$$r_i(p_i, q_i, t^*) \leq e^{2\varepsilon}. \tag{8}$$

Symbol	Description
$\mathbf{P}$	Population of individuals the data is drawn from
$\mathbf{Y}$	$n \times d$ confidential data set
Y_i	Length-d vector of values for individual i
I_i	Indicator for whether individual i is included in $\mathbf{Y}$
$\mathbf{Y}_{-i}$	$(n-1) \times d$ matrix of the values in $\mathbf{Y}$ excluding those for individual i
$\mathcal{S}$	Subset of the support of Y_i constituting a privacy violation
$r^*(p_i, q_i)$	Function describing the data holder's desired relative risk bound
$\mathcal{M}$	Adversary's model for Y_i based on information other than $\mathbf{Y}$
T^*	Noisy estimate of $T(\mathbf{Y})$, the function being released

Table 1: Summary of notation.

Symbol	Definition	Description
p_i	$P[I_i = 1 \mid \mathcal{M}]$	Prior probability of inclusion
q_i	$P[Y_i \in \mathcal{S} \mid I_i = 1, \mathcal{M}]$	Prior probability values disclosed
$r_i(p_i, q_i, t^*)$	$\frac{P[Y_i \in \mathcal{S}, I_i = 1 \mid T^* = t^*, \mathcal{M}]}{P[Y_i \in \mathcal{S}, I_i = 1 \mid \mathcal{M}]}$	Relative disclosure risk
$a_i(p_i, q_i, t^*)$	$P[Y_i \in \mathcal{S}, I_i = 1 \mid T^* = t^*, \mathcal{M}]$	Absolute disclosure risk

Table 2: Summary of definitions.

See, for example, Theorem 1.3 in [15] for proof of this fact.[3] (8) implies a naive strategy for setting ε: select a desired bound on the relative risk, r^*, and set $\varepsilon = \log(r^*)/2$. Practically, however, this strategy suffers from two drawbacks, which cause the recommended ε to be smaller than necessary. First, for any particular prior probabilities p_i and q_i, the bound in (8) need not be tight. In fact, this bound is actually quite loose across a wide range of priors. Second, this strategy does not account for the fact that the data holder may be willing to tolerate different relative risks for different adversary priors. For example, if $p_i q_i = 0.25$, a data holder may wish to limit the adversary's posterior to $a_i(p_i, q_i, t^*) \leq 2 \times 0.25 = 0.5$, but for $p_i q_i = 10^{-6}$, the same data holder may find a limit of $a_i(p_i, q_i, t^*) \leq 2 \times 10^{-6}$ unnecessarily restrictive.

This suggests that, rather than restricting themselves to a constant relative risk bound, the data holder can consider tolerable relative risks as a function of a hypothetical adversary's priors. We refer to this function as the data holder's desired risk profile and denote it as $r^*(p_i, q_i)$. Thus, the data holder establishes a risk profile so that, for all p_i, q_i, and t^*,

$$r_i(p_i, q_i, t^*) \leq r^*(p_i, q_i). \tag{9}$$

As we show in Section 4, the requirement in (9) translates to a maximum value of ε.

3.1 Specifying the risk profile

Given $\mathcal{S}$, the data holder must select a form for $r^*(p_i, q_i)$. A default choice, equivalent to the naive strategy using (8) discussed above, is to set the bound to a constant $\tilde{r} > 1$, i.e.,

$$r^*(p_i, q_i) = \tilde{r}. \tag{10}$$

[3] [15] prove this under bounded DP in the case where $|\mathcal{S}| = 1$, but we show in Corollary 1 that, under our assumptions, it still holds for larger sets and for unbounded DP (with ε replaced by 2ε).

As we prove later in Theorem 3, the bound in (10) implies the data holder should set $\varepsilon = \log(\tilde{r})/2$. While a constant bound on relative risk is simple, data holders' that tolerate different risk profiles may be able to set ε to larger values, as we now illustrate.

As a first example, consider a data holder that requires the relative risk bound to hold on a subset of the (p_i, q_i) space and does not have any bounds outside that space. That is, rather than enforcing $r^*(p_i, q_i) = \tilde{r}$ for all $0 \leq p_i, q_i \leq 1$, the data holder only enforces this condition for $\tilde{p}_0 \leq p_i \leq \tilde{p}_1$ and $\tilde{q}_0 \leq q_i \leq \tilde{q}_1$, where $0 \leq \tilde{p}_0 \leq \tilde{p}_1 \leq 1$ and $0 \leq \tilde{q}_0 \leq \tilde{q}_1 \leq 1$ (for $\tilde{p}_1, \tilde{q}_1 > 0$). Formally,

$$r^*(p_i, q_i) = \begin{cases} \tilde{r}, & \text{if } \tilde{p}_0 \leq p_i \leq \tilde{p}_1 \text{ and } \tilde{q}_0 \leq q_i \leq \tilde{q}_1; \\ \infty, & \text{otherwise.} \end{cases} \tag{11}$$

This risk profile seems unlikely to map to realistic preferences, as it characterizes a data holder who does not care at all about the additional risks from releasing t^* for prior risks that are outside the defined range. However, it does serve to illustrate the potential of the framework to allow the data holder to increase ε based on their risk profile. In particular, as shown later in Theorem 4, the data holder with the risk profile in (11) can set

$$\varepsilon = \begin{cases} \log\left(\frac{2\tilde{p}_1(1-\tilde{q}_0)}{\sqrt{(1-\tilde{p}_1)^2+4\tilde{p}_1(1-\tilde{q}_0)\left(\frac{1}{\tilde{r}}-\tilde{p}_1\tilde{q}_0\right)}-(1-\tilde{p}_1)}\right), & \text{if } 0 \leq \tilde{q}_0 \leq \frac{1}{\tilde{r}+1}; \\ \log\left(\frac{2\tilde{p}_0(1-\tilde{q}_0)}{\sqrt{(1-\tilde{p}_0)^2+4\tilde{p}_0(1-\tilde{q}_0)\left(\frac{1}{\tilde{r}}-\tilde{p}_0\tilde{q}_0\right)}-(1-\tilde{p}_0)}\right), & \text{if } \frac{1}{\tilde{r}+1} < \tilde{q}_0 < 1 \text{ and } \tilde{p}_0 > 0; \\ \log\left(\tilde{r}\right), & \text{if } \frac{1}{\tilde{r}+1} < \tilde{q}_0 < 1 \text{ and } \tilde{p}_0 = 0; \\ \log\left(\frac{1-\tilde{p}_0}{\frac{1}{\tilde{r}}-\tilde{p}_0}\right), & \text{if } \tilde{q}_0 = 1. \end{cases} \tag{12}$$

Notably, this can produce a larger ε than $\log(\tilde{r})/2$.

As a second and more realistic example, consider a data holder that seeks to bound the relative risks for high prior probabilities and bound the absolute disclosure risk for low prior probabilities. For example, the data holder may not want adversaries whose prior probabilities are low to use t^* to increase those probabilities beyond 0.10. Simultaneously, the data holder may want to ensure adversaries with large prior probabilities cannot use t^* to triple their posterior probability. Such a data holder can specify a risk profile that requires either the relative risk be less than some $\tilde{r}$ or the absolute risk be less than some $\tilde{a} < 1$, as we now illustrate.

When the sensitivity of the values in the data is of primary concern (and the sensitivity of inclusion is secondary), the data holder can fix $\tilde{p}$ to some value $\tilde{p} \in (0, 1]$. For example, for a survey of size n_s of a population of size N, the data holder could set $\tilde{p} = n_s/N$, which effectively treats $\mathbf{Y}$ as a simple random sample from $\mathbf{P}$. The data holder could set $\tilde{p} = 1$ to imply an adversary that knows a priori that individual i is included in $\mathbf{Y}$. With a fixed $\tilde{p}$, the implied r^* is of the form

$$r^*(p_i, q_i) = \begin{cases} \max\left\{\frac{\tilde{a}}{\tilde{p}q_i}, \tilde{r}\right\}, & \text{if } p_i = \tilde{p}; \\ \infty, & \text{if } p_i \neq \tilde{p}. \end{cases} = \begin{cases} \frac{\tilde{a}}{\tilde{p}q_i}, & \text{if } p_i = \tilde{p} \text{ and } 0 < q_i < \frac{\tilde{a}}{\tilde{p}\tilde{r}}; \\ \tilde{r}, & \text{if } p_i = \tilde{p} \text{ and } \frac{\tilde{a}}{\tilde{p}\tilde{r}} \leq q_i < 1; \\ \infty, & \text{if } p_i \neq \tilde{p}. \end{cases} \tag{13}$$

An example data holder with this risk function is presented in the first column of Figure 1. The upper plot displays the risk profile as a function of q_i when $p_i = \tilde{p}$ and the lower plot displays the maximal ε for which the relative risk bound holds for each p_i. We recommend the data holder

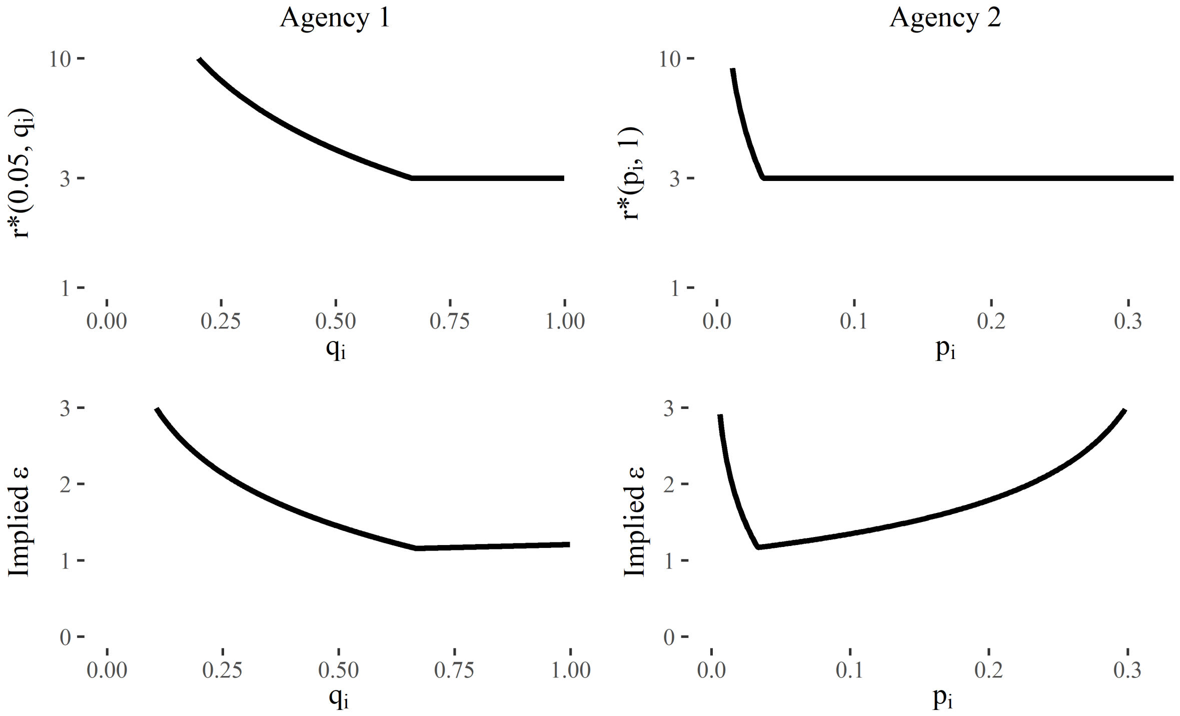

Figure 1: Each column corresponds to a particular hypothetical agency. The first row presents the agency's risk profile and the second row presents the profile's implied maximal allowable ε at each point on the curve. Agency 1's risk profile is given by (13) with $\tilde{a} = 0.1$, $\tilde{p} = 0.05$, and $\tilde{r} = 3$, while Agency 2's risk profile is given by (15) with $\tilde{a} = 0.1$, $\tilde{q} = 1$, and $\tilde{r} = 3$.

select the smallest ε on this curve for their release. As we show later in Theorem 5, the minimal point on this curve has the following form.

$$\varepsilon = \begin{cases} \log\left(\frac{2(\tilde{p}\tilde{r}-\tilde{a})}{\sqrt{\tilde{r}^2(1-\tilde{p})^2+4(\tilde{p}\tilde{r}-\tilde{a})(1-\tilde{a})}-\tilde{r}(1-\tilde{p})}\right), & \text{if } \frac{\tilde{a}}{\tilde{p}} < \tilde{r}; \\ \log\left(\frac{\tilde{a}(1-\tilde{p})}{\tilde{p}(1-\tilde{a})}\right), & \text{if } \frac{\tilde{a}}{\tilde{p}} \geq \tilde{r}. \end{cases} \tag{14}$$

It can be shown that the ε produced by (14) is bounded below by $\log(\tilde{r})/2$ and may be much larger.

Alternatively, when the sensitivity of inclusion in the data is of primary concern (and the sensitivity of the values in the data are secondary), the data holder can fix some $\tilde{q} \in (0, 1]$, giving an implied r^* of the form

$$r^*(p_i, q_i) = \begin{cases} \max\left\{\frac{\tilde{a}}{p_i\tilde{q}}, \tilde{r}\right\}, & \text{if } q_i = \tilde{q}; \\ \infty, & \text{if } q_i \neq \tilde{q}. \end{cases} = \begin{cases} \frac{\tilde{a}}{p_i\tilde{q}}, & \text{if } q_i = \tilde{q} \text{ and } 0 < p_i < \frac{\tilde{a}}{\tilde{q}\tilde{r}}; \\ \tilde{r}, & \text{if } q_i = \tilde{q} \text{ and } \frac{\tilde{a}}{\tilde{q}\tilde{r}} \leq p_i < 1; \\ \infty, & \text{if } q_i \neq \tilde{q}. \end{cases} \tag{15}$$

An example data holder with this risk function is presented in the second column of Figure 1. The upper plot displays the risk profile as a function of p_i when $q_i = \tilde{q}$ and the lower plot displays the maximal ε for which the relative risk bound holds for each p_i. We show later in Theorem 6 that the minimal point on this curve has the following form.

$$\varepsilon = \begin{cases} \frac{1}{2}\log\left(\frac{1-\tilde{q}}{\frac{1}{\tilde{r}}-\tilde{q}}\right), & \text{if } 0 < \tilde{q} \le \frac{1}{\tilde{r}+1}; \\ \log\left(\frac{2\tilde{a}(1-\tilde{q})}{\sqrt{(\tilde{r}\tilde{q}-\tilde{a})^2+4\tilde{q}(1-\tilde{q})(1-\tilde{a})}-(\tilde{r}\tilde{q}-\tilde{a})}\right), & \text{if } \frac{1}{\tilde{r}+1} < \tilde{q} < 1; \\ \log\left(\frac{\tilde{r}-\tilde{a}}{1-\tilde{a}}\right), & \text{if } \tilde{q} = 1. \end{cases} \tag{16}$$

The ε produced by (16) is bounded below by $\log(\tilde{r})/2$ and may be much larger.

For r^* of other forms, there may not be a closed form for the recommended ε. Instead, the optimal ε can be determined by numerically solving the following minimization problem.[4]

$$\varepsilon = \min_{p_i,q_i\in(0,1]} \begin{cases} \log\left(\frac{2p_i(1-q_i)}{\sqrt{(1-p_i)^2+4p_i(1-q_i)\left(\frac{1}{r^*(p_i,q_i)}-p_iq_i\right)}-(1-p_i)}\right), & \text{if } 0 < q_i < 1; \\ \log\left(\frac{1-p_i}{\frac{1}{r^*(p_i,1)}-p_i}\right), & \text{if } q_i = 1. \end{cases} \tag{17}$$

Regardless of the data holder's desiderata for a risk profile, we recommend that they keep the following in mind when setting its functional form. First, for any region where $r^*(p_i, q_i) > 1/(p_iq_i)$, the risk profile generates a bound on the posterior probability that exceeds 1. Of course, the posterior probabilities themselves cannot exceed 1; thus, in these regions, the risk profile effectively does not bound the posterior risk. For example, an agency that sets $r^*(p_i, 1) = 3$ in the region where $p_i \ge 1/3$ (as in the right column of Figure 1) implicitly is willing to accept an unbounded ε for prior probabilities $p_i \ge 1/3$. Second, when bounding the absolute disclosure risk below some $\tilde{a}$ in some region of (p_i, q_i), the data holder should require $p_iq_i < \tilde{a}$ in that region. When $p_iq_i = \tilde{a}$, the recommended $\varepsilon = 0$ since the data holder requires T^* to offer no information about Y_i. This also suggests that a data holder bounding absolute disclosure risk in a region of (p_i, q_i) that set $\tilde{a}$ close to some value of p_iq_i in the region is willing to accept only small ϵ values.

3.2 Examples of Risk Profiles

The use of non-constant posterior-to-prior ratios can lead to different, and potentially larger, ε than the data holder might select otherwise. To demonstrate, in this section we present two examples, beginning with a setting inspired by a case study in [9].

Example 1. *A healthcare provider possesses a data set comprising demographic information about individuals diagnosed with COVID-19 in a particular community. They plan to release the count of individuals diagnosed with COVID-19 in various demographic groups via the Geometric Mechanism. They are concerned this release, if insufficient noise is added, could reveal which individuals in the community had COVID-19 and wish to choose ε appropriately.*

In this example, the primary concern is with respect to inclusion in the data set. That is, for a given individual i, the adversary's prior probability $p_i = P[I_i = 1 \mid \mathcal{M}]$ is the key quantity, whereas the $q_i = P[Y_i \in \mathcal{S} \mid I_i = 1, \mathcal{M}]$ is not as important for any given $\mathcal{S}$. Suppose the data holder is most concerned about adversaries who already know individual i's demographic information, i.e., $q_i = 1$, and suppose the data holder is generally willing to accept a maximum absolute disclosure

[4] We provide `R` code at `https://github.com/zekicankazan/choosing_dp_epsilon` to determine the recommended ε for a provided risk profile.

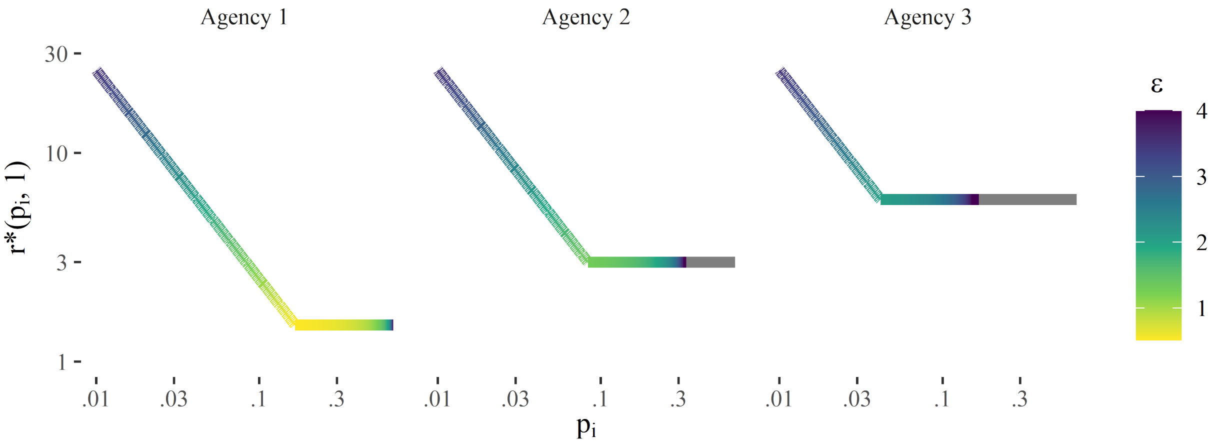

Figure 2: The risk profiles for three agencies with risk profile given by (18). The lines represent the risk profiles for $q_i = 1$ as a function of p_i, and the colors represent the implied ε at each point on the curve. The left plot sets $\tilde{r} = 1.5$, the center sets $\tilde{r} = 3$, and the right sets $\tilde{r} = 6$.

Agency	$\tilde{r}$	ε Recommendation	Noise Std. Dev.	Prob. Exact
1	1.5	0.51	2.74	25%
2	3	1.30	1.02	57%
3	6	2.04	0.59	77%

Table 3: For each of the three risk profiles in Figure 2, we present the ε recommended by our framework. For a release satisfying ε-DP using the Geometric Mechanism, we present the corresponding standard deviation of the noise distribution and the probability that the exact value is released, i.e., the noise distribution's probability mass at zero.

risk of 0.25 for adversaries with small prior probabilities. A reasonable risk profile for this data holder might be of the form, for some $\tilde{r} > 1$,

$$r^*(p_i, q_i) = \begin{cases} \max\left\{\frac{0.25}{p_i}, \tilde{r}\right\}, & \text{if } q_i = 1; \\ \infty, & \text{if } q_i \neq 1. \end{cases} \tag{18}$$

Three example risk functions of this form are presented in Figure 2. Agency 1 corresponds to a risk averse data holder, agency 3 corresponds to a utility seeking data holder, and agency 2 corresponds to a data holder that sits in between in terms of risk and utility. For adversaries with high prior probabilities, agencies 1, 2, and 3 bound the relative disclosure risk at $\tilde{r} = 1.5$, $\tilde{r} = 3$, and $\tilde{r} = 6$, respectively. Table 3 presents the maximal ε which satisfies the desired risk profile for each agency, computed via (16). To provide intuition on the amount of noise implied by these ε's, in Table 3 we display the standard deviation of the noise distribution for each statistic under the Geometric Mechanism and the probability that the Geometric Mechanism will output the exact value of each statistic. The risk averse data holder is recommended a ε that results in a release with a high standard deviation and fairly low probability of releasing the exact value of the statistic. The utility seeking data holder is recommended a ε that results in a release with a fairly low standard deviation and high probability of releasing the exact value of the statistic.

The ε recommendations from these risk profiles, which are tailored to the specific setting and data holder preferences, are much higher than the recommendations from a corresponding simple

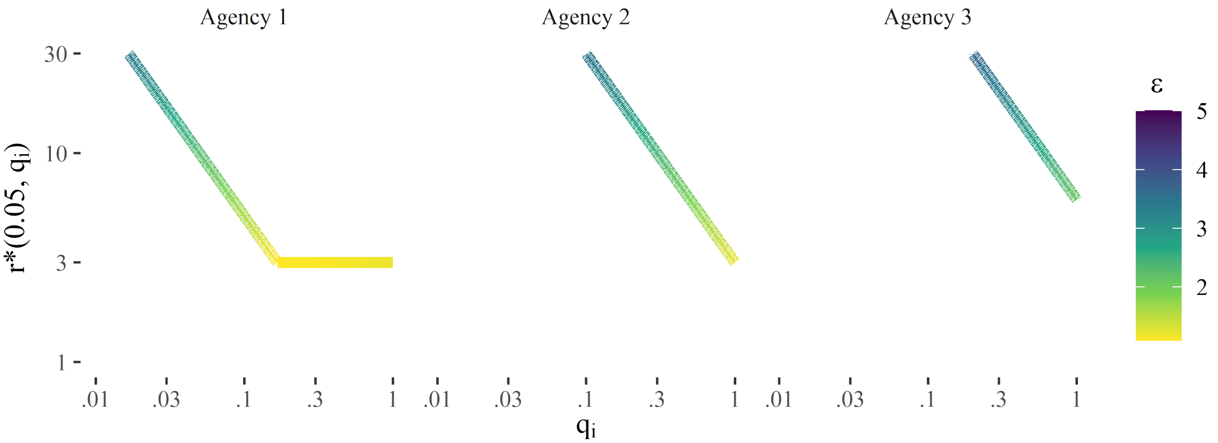

Figure 3: The risk profiles for three agencies with risk profile given by (19). The lines represent the risk profiles for $p_i = 0.05$ as a function of q_i, and the colors represent the implied ε at each point on the curve. The left plot sets $\tilde{a} = 0.025$, the center sets $\tilde{a} = 0.15$, and the right sets $\tilde{a} = 0.3$.

Agency	$\tilde{a}$	ε Recommendation	Noise Std. Dev.	Prob. Exact
1	0.025	1.09	1.24	50%
2	0.15	1.21	1.10	54%
3	0.3	2.10	0.56	78%

Table 4: For each of the three risk profiles in Figure 3, we present the ε recommended by our framework. For a release satisfying ε-DP using the Geometric Mechanism, we present the corresponding standard deviation of the noise distribution and the probability that the exact value is released (i.e., the noise distribution's probability mass at zero).

risk profile of $r^*(p_i, q_i) = \tilde{r}$ for all priors. For comparison, this simple profile yields $\varepsilon \approx 0.20$, $\varepsilon \approx 0.55$, and $\varepsilon \approx 0.90$ for $\tilde{r} = 1.5$, $\tilde{r} = 3$, and $\tilde{r} = 6$, respectively, which are less than half the recommended ε's above.

We now consider a second example that alters Example 1. This example is inspired by Example 16 in [30].

Example 2. *A survey is performed on a sample of individuals in the community of interest. 5% of the community is surveyed, and respondents are asked whether they have had COVID-19 along with a series of demographic questions. The data holder plans to release the counts of surveyed individuals who have and have not been diagnosed with COVID-19 in various demographic groups via the Geometric Mechanism. They are concerned this release, if insufficient noise is added, could reveal which individuals in the community had COVID-19 and wish to choose ε appropriately.*

In this example, the primary concern is with respect to the values in the data set. For ease of notation, let Y_i be a d-vector of binary values and let the first element, Y_{1i}, be an indicator for whether individual i has had COVID-19. Set $\mathcal{S}$ to be the subset of the support of Y_i for which individual i has had COVID-19, i.e., $\mathcal{S} = \{y \in \{0,1\}^d : y_1 = 1\}$. For individual i, the adversary's $q_i = P[Y_i \in \mathcal{S} \mid I_i = 1, \mathcal{M}]$ is the key quantity, and their $p_i = P[I_i = 1 \mid \mathcal{M}]$ is not of interest. Suppose the data holder is most concerned about adversaries whose only prior knowledge is that individual i is in the population, but not whether they were surveyed, i.e., $p_i = 0.05$, and suppose the data holder is generally willing to accept a maximum relative disclosure risk of 3 for adversaries with large prior probabilities. A reasonable risk profile for this data holder might be of the form, for some $\tilde{a} < 1$,

$$r^*(p_i, q_i) = \begin{cases} \max\left\{\frac{\tilde{a}}{0.05q_i}, 3\right\}, & \text{if } p_i = 0.05; \\ \infty, & \text{if } p_i \neq 0.05. \end{cases} \tag{19}$$

Three example risk functions of this form are presented in Figure 3. Once again, agency 1 corresponds to a risk averse data holder, agency 3 corresponds to a utility seeking data holder, and agency 2 corresponds to a data holder that sits in between on risk and utility. Agencies 1, 2, and 3 are willing to allow adversaries to achieve an absolute disclosure risk of $\tilde{a} = 0.025$, $\tilde{a} = 0.15$, and $\tilde{a} = 0.3$, respectively. Table 4 presents the ε recommendations for each agency along with the standard deviation of the noise and probability of releasing the exact value of each statistic under the Geometric Mechanism.

As in Table 3, the ε recommendations appear to match the data holder's desired balance between privacy and accuracy. They also are much higher than the recommendations from a corresponding simple risk profile of $r^*(p_i, q_i) = 3$ for all priors, which implies $\varepsilon \approx 0.55$. Even the most risk averse data holder is recommended an ε that is much larger than this baseline risk profile. This gain is primarily due to the assumption that the survey is a simple random sample from the population and the adversary has no prior knowledge about which individuals are surveyed. Essentially, the additional uncertainty from the sampling mechanism allows for an ε recommendation with less noise injected. This is consistent with prior work showing that DP mechanisms applied to random subsamples provide better privacy guarantees [3].

4 Theoretical Results

We now describe the main theoretical results used to develop the expressions in Section 3. Omitted proofs can be found in Appendix A.2. We begin by formalizing the assumptions on the release mechanism for T^* and the adversary's model, $\mathcal{M}$.

The first assumption implies the following three conditions: (i) the mechanism for releasing T^* given $\mathbf{Y}$ is known to the adversary, (ii) the adversary does not assume a release mechanism that is different than the actual mechanism used by the data holder, and (iii) the adversary does not possess any additional information about T^* beyond what is present in $\mathbf{Y}$.

Assumption 1. *The release mechanism under the adversary's model, $\mathcal{M}$, is the same as the true release mechanism used by the data holder. That is, for all y in the support of Y_i, all $\mathbf{y}_{-i}$ in the support of $\mathbf{Y}_{-i}$, and all t^* in the support of T^*,*

$$P[T^*(\mathbf{Y}) = t^* \mid Y_i = y, \mathbf{Y}_{-i} = \mathbf{y}_{-i}, I_i = 1, \mathcal{M}] = P[T^*(\mathbf{Y}) = t^* \mid Y_i = y, \mathbf{Y}_{-i} = \mathbf{y}_{-i}, I_i = 1]$$

$$P[T^*(\mathbf{Y}_{-i}) = t^* \mid \mathbf{Y}_{-i} = \mathbf{y}_{-i}, I_i = 0, \mathcal{M}] = P[T^*(\mathbf{Y}_{-i}) = t^* \mid \mathbf{Y}_{-i} = \mathbf{y}_{-i}, I_i = 0].$$

Assumption 1 implies that the data holder fully describes release mechanism to the public. It also implies adversaries who are rational and do not use a mechanism other than the one used by the data holder. Additionally, it implies that the adversary has no additional prior knowledge about T^* that does not come from their prior knowledge about $\mathbf{Y}$.

The second assumption involves the adversary's prior distribution for $\mathbf{Y}_{-i}$, the values in the data excluding individual i. We assume that this distribution does not change whether or not individual i is included in the data nor does it depend on individual i's confidential values. We formalize this as follows.

Assumption 2. *Under the adversary's model $\mathcal{M}$, $\mathbf{Y}_{-i}$ is independent of $\{Y_i, I_i\}$. In particular, for all y in the support of Y_i and all $\mathbf{y}_{-i}$ in the support of $\mathbf{Y}_{-i}$*

$$P[\mathbf{Y}_{-i} = \mathbf{y}_{-i} \mid I_i = 1, Y_i = y, \mathcal{M}] = P[\mathbf{Y}_{-i} = \mathbf{y}_{-i} \mid I_i = 0, \mathcal{M}].$$

Data with a network or hierarchical structure can violate this assumption.

We now show that given these assumptions and an adversary's model $\mathcal{M}$, we can relate ε to the distribution of T^* unconditional on $\mathbf{Y}_{-i}$ via the following lemma.

Lemma 1. *Under Assumption 1 and Assumption 2, and T^* adhering to DP, for all y in the support of Y_i and t^* in the support of T^*, we have*

$$e^{-\varepsilon} \leq \frac{P[T^* = t^* \mid Y_i = y, I_i = 1, \mathcal{M}]}{P[T^* = t^* \mid I_i = 0, \mathcal{M}]} \leq e^{\varepsilon}. \tag{20}$$

Proof. Let $\mathcal{Y}_{-i}$ be the support of $\mathbf{Y}_{-i}$ under $\mathcal{M}$. Then,

$$\begin{aligned}
&P[T^* = t^* \mid Y_i = y, I_i = 1, \mathcal{M}] \\
&\quad = \sum_{\mathbf{y}_{-i} \in \mathcal{Y}_{-i}} P[T^* = t^* \mid Y_i = y, \mathbf{Y}_{-i} = \mathbf{y}_{-i}, I_i = 1]\, P[\mathbf{Y}_{-i} = \mathbf{y}_{-i} \mid Y_i = y, I_i = 1, \mathcal{M}] && (21)\\
&\quad \leq \sum_{\mathbf{y}_{-i} \in \mathcal{Y}_{-i}} e^{\varepsilon} P[T^* = t^* \mid \mathbf{Y}_{-i} = \mathbf{y}_{-i}, I_i = 0]\, P[\mathbf{Y}_{-i} = \mathbf{y}_{-i} \mid Y_i = y, I_i = 1, \mathcal{M}] && (22)\\
&\quad = e^{\varepsilon} \sum_{\mathbf{y}_{-i} \in \mathcal{Y}_{-i}} P[T^* = t^* \mid \mathbf{Y}_{-i} = \mathbf{y}_{-i}, I_i = 0, \mathcal{M}]\, P[\mathbf{Y}_{-i} = \mathbf{y}_{-i} \mid I_i = 0] && (23)\\
&\quad = e^{\varepsilon} P[T^* = t^* \mid I_i = 0, \mathcal{M}]. && (24)
\end{aligned}$$

The equality in (21) follows from the law of total probability and Assumption 1. The inequality in (22) follows from the definition of DP in (1). The equality in (23) follows from Assumption 2. The equality in (24) follows from the law of total probability and Assumption 1. This completes the proof of the right inequality. The proof of the left inequality is identical with the other inequality in (1) applied in (22). □

We now generalize Lemma 1 from a single point $Y_i = y$ to a set $Y_i \in \mathcal{S}$. We include a proof in Appendix A.2.

Lemma 2. *Under Assumption 1 and Assumption 2, if the release of $T^* = t^*$ satisfies DP, then for any subset $\mathcal{S}$ of the domain of Y_i, we have*

$$e^{-\varepsilon} \leq \frac{P[T^* = t^* \mid Y_i \in \mathcal{S}, I_i = 1, \mathcal{M}]}{P[T^* = t^* \mid I_i = 0, \mathcal{M}]} \leq e^{\varepsilon} \tag{25}$$

and

$$e^{-2\varepsilon} \leq \frac{P[T^* = t^* \mid Y_i \in \mathcal{S}, I_i = 1, \mathcal{M}]}{P[T^* = t^* \mid Y_i \notin \mathcal{S}, I_i = 1, \mathcal{M}]} \leq e^{2\varepsilon}. \tag{26}$$

For a given function r^* selected by the data holder, we can determine the ε that should be used for the release. This is due to the following result relating the relative risk to ε.

Theorem 1. *Under Assumption 1 and Assumption 2, if the release of $T^* = t^*$ satisfies DP, then*

$$r_i(p_i, q_i, t^*) \leq \frac{1}{q_i p_i + e^{-2\varepsilon}(1 - q_i)p_i + e^{-\varepsilon}(1 - p_i)}. \tag{27}$$

Proof. We begin by applying Bayes' Theorem to reverse the conditional in the relative risk.

$$
\begin{aligned}
r_i(p_i, q_i, t^*) &= \frac{P[Y_i \in \mathcal{S}, I_i = 1 \mid T^* = t^*, \mathcal{M}]}{P[Y_i \in \mathcal{S}, I_i = 1 \mid \mathcal{M}]} && (28)\\
&= \frac{\frac{P[T^*=t^* \mid Y_i \in \mathcal{S}, I_i=1, \mathcal{M}]\, P[Y_i \in \mathcal{S}, I_i=1 \mid \mathcal{M}]}{P[T^*=t^* \mid \mathcal{M}]}}{P[Y_i \in \mathcal{S}, I_i = 1 \mid \mathcal{M}]} && (29)\\
&= \frac{P[T^* = t^* \mid Y_i \in \mathcal{S}, I_i = 1, \mathcal{M}]}{P[T^* = t^* \mid \mathcal{M}]}. && (30)
\end{aligned}
$$

We may decompose the denominator via the law of total probability.

$$
\begin{aligned}
P[T^* = t^* \mid \mathcal{M}] = {} & P[T^* = t^* \mid Y_i \in \mathcal{S}, I_i = 1, \mathcal{M}]\, P[Y_i \in \mathcal{S} \mid I_i = 1, \mathcal{M}]\, P[I_i = 1 \mid \mathcal{M}]\\
& + P[T^* = t^* \mid Y_i \notin \mathcal{S}, I_i = 1, \mathcal{M}]\, P[Y_i \notin \mathcal{S} \mid I_i = 1, \mathcal{M}]\, P[I_i = 1 \mid \mathcal{M}]\\
& + P[T^* = t^* \mid I_i = 0, \mathcal{M}]\, P[I_i = 0 \mid \mathcal{M}] && (31)\\
= {} & P[T^* = t^* \mid Y_i \in \mathcal{S}, I_i = 1, \mathcal{M}]\, q_i p_i + P[T^* = t^* \mid Y_i \notin \mathcal{S}, I_i = 1, \mathcal{M}]\, (1 - q_i) p_i\\
& + P[T^* = t^* \mid I_i = 0, \mathcal{M}]\, (1 - p_i). && (32)
\end{aligned}
$$

Using this expansion in the expression for r_i and dividing through by the numerator yields

$$
r_i(p_i, q_i, t^*) = \frac{1}{q_i p_i + \frac{P[T^*=t^* \mid Y_i \notin \mathcal{S}, I_i=1, \mathcal{M}]}{P[T^*=t^* \mid Y_i \in \mathcal{S}, I_i=1, \mathcal{M}]}\, (1 - q_i) p_i + \frac{P[T^*=t^* \mid I_i=0, \mathcal{M}]}{P[T^*=t^* \mid Y_i \in \mathcal{S}, I_i=1, \mathcal{M}]}\, (1 - p_i)}. \quad (33)
$$

Using Lemma 2, we then have

$$
r_i(p_i, q_i, t^*) \leq \frac{1}{q_i p_i + e^{-2\varepsilon}(1 - q_i) p_i + e^{-\varepsilon}(1 - p_i)}. \quad (34)
$$

□

Using Theorem 1, one can solve for $e^{-\varepsilon}$ in (27) to determine the recommended ε, which is given by Theorem 2. A proof of this theorem is included in Appendix A.2.

Theorem 2. *For individual i, fix the adversary's prior probabilities, p_i and q_i, and a desired bound on the relative disclosure risk, $r^*(p_i, q_i)$. Under the conditions of Theorem 1, any statistic $T^* = t^*$ released under ε-DP with*

$$
\varepsilon \leq \begin{cases} \log\left(\frac{2p_i(1-q_i)}{\sqrt{(1-p_i)^2 + 4p_i(1-q_i)\left(\frac{1}{r^*(p_i,q_i)} - p_i q_i\right)} - (1-p_i)} \right), & \textit{if } 0 < q_i < 1;\\ \log\left(\frac{1-p_i}{\frac{1}{r^*(p_i,1)} - p_i} \right), & \textit{if } q_i = 1, \end{cases} \quad (35)
$$

will satisfy $r_i(p_i, q_i, t^*) \leq r^*(p_i, q_i)$.

By Theorem 2, to achieve $r_i(p_i, q_i, t^*) \leq r^*(p_i, q_i)$ for all (p_i, q_i), the data holder should set

$$
\varepsilon = \min_{p_i, q_i \in (0,1]} \begin{cases} \log\left(\frac{2p_i(1-q_i)}{\sqrt{(1-p_i)^2 + 4p_i(1-q_i)\left(\frac{1}{r^*(p_i,q_i)} - p_i q_i\right)} - (1-p_i)} \right), & \text{if } 0 < q_i < 1;\\ \log\left(\frac{1-p_i}{\frac{1}{r^*(p_i,1)} - p_i} \right), & \text{if } q_i = 1. \end{cases} \quad (36)
$$

Results regarding closed forms for the ε resulting from specific risk profiles are included in the appendix.

5 Relationship to Prior Work

In this section, we compare our framework to two previous works with similar aims. We begin with Lee and Clifton's "How Much is Enough? Choosing ε for Differential Privacy" [21], which also examines the problem of selecting ε from a Bayesian perspective. We then discuss Wood et al.'s "Differential Privacy: A Primer for a Non-Technical Audience" [30], which uses similar ideas in a different context.

5.1 "How Much is Enough? Choosing ε for Differential Privacy"

Lee and Clifton [21] focus on settings where the population, $\mathbf{P}$, of size n is public information and the adversary's goal is to determine which subset of individuals in $\mathbf{P}$ was used for a differentially private release of a statistic. We can characterize their setting with the notation of Section 2 as follows. We define $\mathbf{Y}$ to be the subset of individuals' values in $\mathbf{P}$ used to compute the statistic of interest, $T(\mathbf{Y})$, and its released DP counterpart, $T^*(\mathbf{Y})$. In their examples, the authors focus on the setting where only one individual is removed from $\mathbf{P}$ to create $\mathbf{Y}$, and the adversary's goal is to determine which i was removed.

We can apply our framework to this setting with a minor modification. For this comparison, we assume the adversary's $q_i = P[Y_i \in \mathcal{S} \mid I_i = 0, \mathcal{M}] = 1$ for any set $\mathcal{S}$ (although we note that this is a weaker assumption than that of Lee and Clifton, since we do not assume $\mathbf{P}$ is public). We redefine p_i and the risk measures to be in terms of $I_i = 0$, rather than $I_i = 1$.

$$p_i = P[I_i = 0 \mid \mathcal{M}] \tag{37}$$

$$r_i(p_i, 1, t^*) = \frac{P[I_i = 0 \mid T^* = t^*, \mathcal{M}]}{P[I_i = 0 \mid \mathcal{M}]} \tag{38}$$

$$a_i(p_i, 1, t^*) = P[I_i = 0 \mid T^* = t^*, \mathcal{M}]. \tag{39}$$

Lee and Clifton [21] focus on the case of $p_i = 1/n$, and seek to enforce the bound $a_i(1/n, 1, t^*) \leq \tilde{a}$ for some constant $\tilde{a}$ and all t^*, which implies the relative risk bound

$$r^*(p_i, q_i) = \begin{cases} n\tilde{a}, & \text{if } p_i = \frac{1}{n}, q_i = 1; \\ \infty, & \text{otherwise.} \end{cases} \tag{40}$$

From an analogy to Theorem 2 with the redefined p_i, it follows that under these conditions, our method sets

$$\varepsilon = \log\left(\frac{1 - \frac{1}{n}}{\frac{1}{n\tilde{a}} - \frac{1}{n}}\right) = \log\left(\frac{(n-1)\tilde{a}}{1 - \tilde{a}}\right). \tag{41}$$

In the motivating example from their paper, the authors set $n = 4$ and $\tilde{a} = 1/3$, giving $r^*(1/n, 1) = 4/3$. This results in $\varepsilon = \log(3/2) \approx 0.41$ from our method. When the release mechanism is the addition of Laplace noise, Lee and Clifton's method [21] arrives at a similar form, but with the recommendation scaled by a factor of $\Delta T/\Delta v$.

$$\varepsilon = \frac{\Delta T}{\Delta v} \log\left(\frac{(n-1)\tilde{a}}{1 - \tilde{a}}\right), \tag{42}$$

$$\Delta T = \max\left\{|T(\mathbf{Y}_1) - T(\mathbf{Y}_2)| : \mathbf{Y}_2 \subset \mathbf{Y}_1 \subset \mathbf{P}, |\mathbf{Y}_1| = n - 1, |\mathbf{Y}_2| = n - 2\right\} \tag{43}$$

$$\Delta v = \max\left\{|T(\mathbf{Y}_1) - T(\mathbf{Y}_2)| : \mathbf{Y}_1 \subset \mathbf{P}, \mathbf{Y}_2 \subset \mathbf{P}, |\mathbf{Y}_1| = |\mathbf{Y}_2| = n - 1\right\}. \tag{44}$$

The recommendation of Lee and Clifton's method [21] thus depends on both the population, $\mathbf{P}$, and the particular release function, T. This results in different ε's for the same n and $\tilde{a}$, as low as 0.34 and as high as 1.62 in the authors' examples, depending on the statistic of interest and the values in the data.

Lee and Clifton's method [21] for selecting ε is tailored to the setting where the Laplace mechanism is used for the release, only disclosure of an individual's inclusion in the data is of concern, and the values of the entire population can be used to inform the choice of ε. Setting ε in this manner allows for the choice to be tailored to the particular statistic and the variance of the population, providing an ε recommendation that may be larger than that of our framework. The population-dependent nature of the choice, however, limits the generalizability of the method to settings where the values of the entire population are not public and where the privacy of the values in the release is of primary concern.

5.2 "Differential Privacy: A Primer for a Non-Technical Audience"

Another related work involves an example in [30] (corrected in [31]). The example considers an individual deciding whether or not to participate in a survey for which results will be released via DP with a particular ε. Using our notation, let $Z_i = f(Y_i, I_i) \in \{0, 1\}$ be a quantity of interest to the adversary, who wishes to learn whether $Z_i = 1$. They have some prior $q_i = P[Z_i = 1 \mid \mathcal{M}]$. Rather than considering the relative or absolute disclosure risk, the individual is interested in comparing the adversary's posterior probability if they participate in the survey, $a_{1i}(q_i, t^*) = P[Z_i = 1 \mid I_i = 1, T^* = t^*, \mathcal{M}]$, to the adversary's posterior probability if they do not participate, $a_{0i}(q_i, t^*) = P[Z_i = 1 \mid I_i = 0, T^* = t^*, \mathcal{M}]$. The authors of [30] state that for all q_i and all t^*,

$$a_{1i}(q_i, t^*) \leq \frac{a_{0i}(q_i, t^*)}{a_{0i}(q_i, t^*) + e^{-2\varepsilon}(1 - a_{0i}(q_i, t^*))}. \tag{45}$$

This expression is in the same spirit as the results from our framework with $p_i = 1$. By Theorem 1, we have

$$r_i(1, q_i, t^*) \leq \frac{1}{q_i + e^{-2\varepsilon}(1 - q_i)} \quad \implies \quad a_i(1, q_i, t^*) \leq \frac{q_i}{q_i + e^{-2\varepsilon}(1 - q_i)}. \tag{46}$$

The authors of [30] suggest that the individual considering survey participation use (45) to bound a_{1i} for various values of a_{0i}. The individual can examine these bounds to make an informed decision about whether to participate in the survey.

While this result is similar to results from our framework, the goals of the frameworks differ. The authors of [30] use (45) to characterize the individual's disclosure risks for a fixed ε, whereas we fix the data holder's disclosure risk profile in order to set ε.

6 Commentary

In this article, we propose a framework for selecting ε for DP using a data holder's disclosure risk profile. Essentially, we provide a method for data holders to trade the problem of selecting ε for a release for the problem of specifying their desired disclosure risk profile. This process involves focusing on particular classes of adversaries the data holder is most concerned about and tuning ε to ensure the risk from these adversaries is sufficiently low. We emphasize that, once applied, DP will protect against all attacks with the guarantee of DP, not just the attack used to tune ε.

Recent work has expressed concerns about the suitability of posterior-to-prior comparisons in the context of DP. For example, one concern in [19] is that the relative risk can be arbitrarily large without making assumptions on the prior, and consensus about prior specification will never be a settled issue. In our framework, the data holder sets a maximum allowable risk for every value of the prior probabilities p_i and q_i in their analysis. No consensus about reasonable priors is required. A second concern in [19] is that posterior-to-prior comparisons consider any information gain, including generalizable scientific knowledge, to be a privacy violation. This issue, while still present, is mitigated by the use of a disclosure risk profile to select ε. For example, when the adversary's prior probability is small, the data holder can account for potential gain in disclosure risks due to generalizable knowledge by allowing for a larger relative risk for such priors. Importantly, the data holder uses relative risks only to set ε; the resulting release is differentially private.

One avenue for future work involves incorporating a version of this framework into differentially private data analysis tools, such as OpenDP's [13] DP Creator. In particular, [28] recently interviewed users of DP Creator and found that interviewees wished for more explanation about how to select privacy parameters and better understanding of the effects of this choice. Relating this decision to statistical disclosure risks as in our framework could aid decision making within such tools.

Additional future extensions could involve examining whether similar results follow under weaker assumptions, for example, not requiring the independence of Assumption 2. It may be possible to extend the framework to settings with multiple differentially private releases, exploiting results relating the relative risk to DP composition theorems (e.g., Section S5 of [18]). Additionally, these results could be extended from the posterior-to-prior risks we discuss in this article to the sorts of posterior-to-posterior risks discussed in Section 5.2.

Acknowledgements

This research was supported by NSF grant SES-2217456. We would like to thank Salil Vadhan and Jörg Drechsler for insightful discussion about a preliminary version of this work.

References

[1] John M Abowd, Robert Ashmead, Ryan Cumings-Menon, Simson Garfinkel, Micah Heineck, Christine Heiss, Robert Johns, Daniel Kifer, Philip Leclerc, Ashwin Machanavajjhala, et al. The 2020 census disclosure avoidance system topdown algorithm. *arXiv preprint arXiv:2204.08986*, 2022.

[2] John M Abowd and Lars Vilhuber. How protective are synthetic data? In *International Conference on Privacy in Statistical Databases*, pages 239–246. Springer, 2008.

[3] Borja Balle, Gilles Barthe, and Marco Gaboardi. Privacy amplification by subsampling: Tight analyses via couplings and divergences. *Advances in Neural Information Processing Systems*, 31, 2018.

[4] Pranav Dandekar, Nadia Fawaz, and Stratis Ioannidis. Privacy auctions for recommender systems. *ACM Transactions on Economics and Computation (TEAC)*, 2(3):1–22, 2014.

[5] Jörg Drechsler. Differential privacy for government agencies—are we there yet? *Journal of the American Statistical Association*, 118(541):761–773, 2023.

[6] George T Duncan and Diane Lambert. Disclosure-limited data dissemination. *Journal of the American statistical association*, 81(393):10–18, 1986.

[7] Cynthia Dwork. A firm foundation for private data analysis. *Communications of the ACM*, 54(1):86–95, 2011.

[8] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In *Theory of cryptography conference*, pages 265–284. Springer, 2006.

[9] Amalie Dyda, Michael Purcell, Stephanie Curtis, Emma Field, Priyanka Pillai, Kieran Ricardo, Haotian Weng, Jessica C Moore, Michael Hewett, Graham Williams, et al. Differential privacy for public health data: An innovative tool to optimize information sharing while protecting data confidentiality. *Patterns*, 2(12):100366, 2021.

[10] Úlfar Erlingsson, Vasyl Pihur, and Aleksandra Korolova. Rappor: Randomized aggregatable privacy-preserving ordinal response. In *Proceedings of the 2014 ACM SIGSAC conference on computer and communications security*, pages 1054–1067, 2014.

[11] Stephen E Fienberg and Ashish P Sanil. A bayesian approach to data disclosure: Optimal intruder behavior for continuous data. *Journal of Official Statistics*, 13(1):75, 1997.

[12] Lisa K Fleischer and Yu-Han Lyu. Approximately optimal auctions for selling privacy when costs are correlated with data. In *Proceedings of the 13th ACM conference on electronic commerce*, pages 568–585, 2012.

[13] Marco Gaboardi, Michael Hay, and Salil Vadhan. A programming framework for opendp. *Manuscript, May*, 2020.

[14] Arpita Ghosh, Tim Roughgarden, and Mukund Sundararajan. Universally utility-maximizing privacy mechanisms. *SIAM Journal on Computing*, 41(6):1673–1693, 2012.

[15] Ruobin Gong and Xiao-Li Meng. Congenial differential privacy under mandated disclosure. In *Proceedings of the 2020 ACM-IMS on foundations of data science conference*, pages 59–70, 2020.

[16] V Joseph Hotz, Christopher R Bollinger, Tatiana Komarova, Charles F Manski, Robert A Moffitt, Denis Nekipelov, Aaron Sojourner, and Bruce D Spencer. Balancing data privacy and usability in the federal statistical system. *Proceedings of the National Academy of Sciences*, 119(31), 2022.

[17] Justin Hsu, Marco Gaboardi, Andreas Haeberlen, Sanjeev Khanna, Arjun Narayan, Benjamin C Pierce, and Aaron Roth. Differential privacy: An economic method for choosing epsilon. In *2014 IEEE 27th Computer Security Foundations Symposium*, pages 398–410. IEEE, 2014.

[18] Zeki Kazan and Jerome Reiter. Assessing statistical disclosure risk for differentially private, hierarchical count data, with application to the 2020 us decennial census. *arXiv preprint arXiv:2204.04253*, 2022.

[19] Daniel Kifer, John M Abowd, Robert Ashmead, Ryan Cumings-Menon, Philip Leclerc, Ashwin Machanavajjhala, William Sexton, and Pavel Zhuravlev. Bayesian and frequentist semantics for common variations of differential privacy: Applications to the 2020 census. *arXiv preprint arXiv:2209.03310*, 2022.

[20] Daniel Kifer and Ashwin Machanavajjhala. No free lunch in data privacy. In *Proceedings of the 2011 ACM SIGMOD International Conference on Management of data*, pages 193–204, 2011.

[21] Jaewoo Lee and Chris Clifton. How much is enough? choosing ε for differential privacy. In *International Conference on Information Security*, pages 325–340. Springer, 2011.

[22] Chao Li, Daniel Yang Li, Gerome Miklau, and Dan Suciu. A theory of pricing private data. *ACM Transactions on Database Systems (TODS)*, 39(4):1–28, 2014.

[23] Ashwin Machanavajjhala, Daniel Kifer, John Abowd, Johannes Gehrke, and Lars Vilhuber. Privacy: Theory meets practice on the map. In *2008 IEEE 24th international conference on data engineering*, pages 277–286. IEEE, 2008.

[24] David McClure and Jerome P Reiter. Differential privacy and statistical disclosure risk measures: An investigation with binary synthetic data. *Trans. Data Priv.*, 5(3):535–552, 2012.

[25] Chaya Nayak. New privacy-protected facebook data for independent research on social media's impact on democracy. *Facebook Research*, 2020.

[26] Jerome P Reiter. Estimating risks of identification disclosure in microdata. *Journal of the American Statistical Association*, 100(472):1103–1112, 2005.

[27] Jerome P Reiter. Differential privacy and federal data releases. *Annual review of statistics and its application*, 6:85–101, 2019.

[28] Jayshree Sarathy, Sophia Song, Audrey Haque, Tania Schlatter, and Salil Vadhan. Don't look at the data! how differential privacy reconfigures the practices of data science. *arXiv preprint arXiv:2302.11775*, 2023.

[29] Jun Tang, Aleksandra Korolova, Xiaolong Bai, Xueqiang Wang, and Xiaofeng Wang. Privacy loss in apple's implementation of differential privacy on macos 10.12. *arXiv preprint arXiv:1709.02753*, 2017.

[30] Alexandra Wood, Micah Altman, Aaron Bembenek, Mark Bun, Marco Gaboardi, James Honaker, Kobbi Nissim, David R O'Brien, Thomas Steinke, and Salil Vadhan. Differential privacy: A primer for a non-technical audience. *Vand. J. Ent. & Tech. L.*, 21:209, 2018.

[31] Alexandra Wood, Micah Altman, Kobbi Nissim, and Salil Vadhan. Designing access with differential privacy. *Handbook on Using Administrative Data for Research and Evidence-based Policy, Shawn Cole, Iqbal Dhaliwal, Anja Sautmann, and Lars Vilhuber (Eds.). Abdul Latif Jameel Poverty Action Lab, Cambridge, MA*, 2020.

A Omitted Results and Proofs

A.1 Omitted Results

In this section, we present all results omitted from the paper. Proofs of these results are included in Appendix A.2.

First, we state a corollary to Theorem 1, which generalizes Theorem 1.3 in [15] to sets where $|\mathcal{S}| > 1$ and to unbounded DP.

Corollary 1. *Under the conditions of Theorem 1, for all $p_i, q_i \in (0, 1]$ and all t^*,*

$$r_i(p_i, q_i, t^*) \leq e^{2\varepsilon}. \tag{47}$$

Next, we state a corollary to Theorem 2 which considers the special case where $p_i = 1$. This corresponds to a setting where individual i's inclusion in the data is known a priori by the adversary, for example data from a census or public social media platform.

Corollary 2. *Under the conditions of Theorem 2, if $p_i = 1$ and $0 < q_i < 1$, any statistic $T^* = t^*$ released under ε-DP with*

$$\varepsilon \leq \frac{1}{2} \log\left(\frac{1 - q_i}{\frac{1}{r^*(1,q_i)} - q_i}\right), \tag{48}$$

will satisfy $r_i(1, q_i, t^*) \leq r^*(1, q_i)$.

For particular forms of r^*, the optimization in (36) has a closed form solution. This is detailed by the following theorems, which are preceded by two lemmas used in the proofs.

Lemma 3. *Fix $p_i, q_i \in (0, 1)$ and let $r^*(p_i, q_i) = \frac{\tilde{a}}{p_i q_i}$. Then the function*

$$\varepsilon(p_i, q_i) = \log\left(\frac{2p_i(1 - q_i)}{\sqrt{(1 - p_i)^2 + 4p_i(1 - q_i)\left(\frac{1}{r^*(p_i,q_i)} - p_i q_i\right)} - (1 - p_i)}\right) \tag{49}$$

has partial derivatives such that

1. $\frac{\partial \varepsilon(p_i,q_i)}{\partial p_i} < 0$ *for all* $0 < p_i < 1$ *and* $0 < q_i < 1$
2. $\frac{\partial \varepsilon(p_i,q_i)}{\partial q_i} < 0$ *for all* $0 < p_i < 1$ *and* $0 < q_i < 1$.

Lemma 4. *Fix $p_i, q_i \in (0, 1)$ and let $r^*(p_i, q_i) = \tilde{r}$. Then the function*

$$\varepsilon(p_i, q_i) = \log\left(\frac{2p_i(1 - q_i)}{\sqrt{(1 - p_i)^2 + 4p_i(1 - q_i)\left(\frac{1}{r^*(p_i,q_i)} - p_i q_i\right)} - (1 - p_i)}\right) \tag{50}$$

has partial derivatives such that

1. $\frac{\partial \varepsilon(p_i,q_i)}{\partial p_i} < 0$ *if* $0 < q_i < \frac{1}{\tilde{r}+1}$ *and* $0 < p_i < 1$
2. $\frac{\partial \varepsilon(p_i,q_i)}{\partial p_i} = 0$ *if* $q_i = \frac{1}{\tilde{r}+1}$ *and* $0 < p_i < 1$

3. $\frac{\partial \varepsilon(p_i,q_i)}{\partial p_i} > 0$ *if* $\frac{1}{\tilde{r}+1} < q_i < 1$ *and* $0 < p_i < 1$

4. $\frac{\partial \varepsilon(p_i,q_i)}{\partial q_i} > 0$ *for all* $0 < p_i < 1$ *and* $0 < q_i < 1$.

Theorem 3. *Under the conditions of Theorem 2, if* $r^*(p_i, q_i) = \tilde{r} > 1$*, the solution to the minimization problem in (36) is*

$$\varepsilon = \frac{1}{2}\log\left(\tilde{r}\right). \tag{51}$$

Theorem 4. *Under the conditions of Theorem 2, let* $0 \le \tilde{p}_0 \le \tilde{p}_1 \le 1$, $0 \le \tilde{q}_0 \le \tilde{q}_1 \le 1$, $\tilde{p}_1, \tilde{q}_1 > 0$, *and* $\tilde{r} > 1$*. If the function* r^* *is such that* $r^*(p_i, q_i) = \tilde{r}$ *if* $p_i \in [\tilde{p}_0, \tilde{p}_1], q_i \in [\tilde{q}_0, \tilde{q}_1]$ *and* $r^*(p_i, q_i) = \infty$ *otherwise, then the solution to the minimization problem in (36) is*

$$\varepsilon = \begin{cases} \log\left(\frac{2\tilde{p}_1(1-\tilde{q}_0)}{\sqrt{(1-\tilde{p}_1)^2+4\tilde{p}_1(1-\tilde{q}_0)\left(\frac{1}{\tilde{r}}-\tilde{p}_1\tilde{q}_0\right)}-(1-\tilde{p}_1)}\right), & \textit{if } 0 \le \tilde{q}_0 \le \frac{1}{\tilde{r}+1}; \\ \log\left(\frac{2\tilde{p}_0(1-\tilde{q}_0)}{\sqrt{(1-\tilde{p}_0)^2+4\tilde{p}_0(1-\tilde{q}_0)\left(\frac{1}{\tilde{r}}-\tilde{p}_0\tilde{q}_0\right)}-(1-\tilde{p}_0)}\right), & \textit{if } \frac{1}{\tilde{r}+1} < \tilde{q}_0 < 1 \textit{ and } \tilde{p}_0 > 0; \\ \log\left(\tilde{r}\right), & \textit{if } \frac{1}{\tilde{r}+1} < \tilde{q}_0 < 1 \textit{ and } \tilde{p}_0 = 0; \\ \log\left(\frac{1-\tilde{p}_0}{\frac{1}{\tilde{r}}-\tilde{p}_0}\right), & \textit{if } \tilde{q}_0 = 1. \end{cases} \tag{52}$$

Theorem 5. *Under the conditions of Theorem 2, let* $\tilde{a} < 1$, $\tilde{p} \le 1$, *and* $\tilde{r} > 1$, *and* $0 < q_i < 1$*. If the function* r^* *is such that* $r^*(\tilde{p}, q_i) = \max\{\tilde{a}/(\tilde{p}q_i), \tilde{r}\}$ *and* $r^*(p_i, q_i) = \infty$ *if* $p_i \neq \tilde{p}$*, then the solution to the minimization problem in (36) is*

$$\varepsilon = \begin{cases} \log\left(\frac{2(\tilde{p}\tilde{r}-\tilde{a})}{\sqrt{\tilde{r}^2(1-\tilde{p})^2+4(\tilde{p}\tilde{r}-\tilde{a})(1-\tilde{a})}-\tilde{r}(1-\tilde{p})}\right), & \textit{if } \frac{\tilde{a}}{\tilde{p}} < \tilde{r}; \\ \log\left(\frac{\tilde{a}(1-\tilde{p})}{\tilde{p}(1-\tilde{a})}\right), & \textit{if } \frac{\tilde{a}}{\tilde{p}} \ge \tilde{r}. \end{cases} \tag{53}$$

Theorem 6. *Under the conditions of Theorem 2, let* $\tilde{a} < 1$, $\tilde{q} \le 1$, *and* $\tilde{r} > 1$, *and* $0 < p_i < 1$*. If the function* r^* *is such that* $r^*(p_i, \tilde{q}) = \max\{\tilde{a}/(p_i\tilde{q}), \tilde{r}\}$ *and* $r^*(p_i, q_i) = \infty$ *for* $q_i \neq \tilde{q}$*, then the solution to the minimization problem in (36) is*

$$\varepsilon = \begin{cases} \frac{1}{2}\log\left(\frac{1-\tilde{q}}{\frac{1}{\tilde{r}}-\tilde{q}}\right), & \textit{if } 0 < \tilde{q} \le \frac{1}{\tilde{r}+1}; \\ \log\left(\frac{2\tilde{a}(1-\tilde{q})}{\sqrt{(\tilde{r}\tilde{q}-\tilde{a})^2+4\tilde{q}(1-\tilde{q})(1-\tilde{a})}-(\tilde{r}\tilde{q}-\tilde{a})}\right), & \textit{if } \frac{1}{\tilde{r}+1} < \tilde{q} < 1; \\ \log\left(\frac{\tilde{r}-\tilde{a}}{1-\tilde{a}}\right), & \textit{if } \tilde{q} = 1. \end{cases} \tag{54}$$

A.2 Omitted Proofs

This section provides proofs of results from Section 4 and Appendix A.1, including Lemmas 2–4, Theorems 2–6, and Corollaries 1–2.

Lemma 2. *Under Assumptions 1-2, if the release of* $T^* = t^*$ *satisfies differential privacy, then for any subset* $\mathcal{S}$ *of the domain of* Y_i,

$$e^{-\varepsilon} \le \frac{P[T^* = t^* \mid Y_i \in \mathcal{S}, I_i = 1, \mathcal{M}]}{P[T^* = t^* \mid I_i = 0, \mathcal{M}]} \le e^{\varepsilon} \tag{55}$$

and

$$e^{-2\varepsilon} \le \frac{P[T^* = t^* \mid Y_i \in \mathcal{S}, I_i = 1, \mathcal{M}]}{P[T^* = t^* \mid Y_i \notin \mathcal{S}, I_i = 1, \mathcal{M}]} \le e^{2\varepsilon}. \tag{56}$$

Proof. We begin by applying Bayes' Theorem to the numerator in (25).

$$\frac{P[T^* = t^* \mid Y_i \in \mathcal{S}, I_i = 1, \mathcal{M}]}{P[T^* = t^* \mid I_i = 0, \mathcal{M}]} = \frac{\frac{P[Y_i \in \mathcal{S} | T^* = t^*, I_i = 1, \mathcal{M}]\, P[T^* = t^* | I_i = 1, \mathcal{M}]}{P[Y_i \in \mathcal{S} | I_i = 1, \mathcal{M}]}}{P[T^* = t^* \mid I_i = 0, \mathcal{M}]} \tag{57}$$

$$= \frac{P[T^* = t^* \mid I_i = 1, \mathcal{M}]\, P[Y_i \in \mathcal{S} \mid T^* = t^*, I_i = 1, \mathcal{M}]}{P[Y_i \in \mathcal{S} \mid I_i = 1, \mathcal{M}]\, P[T^* = t^* \mid I_i = 0, \mathcal{M}]} \tag{58}$$

We can then break the second term in the numerator into a summation and apply Bayes' Theorem to each term in the sum.

$$\frac{P[T^* = t^* \mid Y_i \in \mathcal{S}, I_i = 1, \mathcal{M}]}{P[T^* = t^* \mid I_i = 0, \mathcal{M}]} = \frac{P[T^* = t^* \mid I_i = 1, \mathcal{M}] \sum_{y \in \mathcal{S}} P[Y_i = y \mid T^* = t^*, I_i = 1, \mathcal{M}]}{P[Y_i \in \mathcal{S} \mid I_i = 1, \mathcal{M}]\, P[T^* = t^* \mid I_i = 0, \mathcal{M}]} \tag{59}$$

$$= \frac{P[T^* = t^* \mid I_i = 1, \mathcal{M}] \sum_{y \in \mathcal{S}} \frac{P[T^* = t^* | Y_i = y, I_i = 1, \mathcal{M}]\, P[Y_i = y | I_i = 1, \mathcal{M}]}{P[T^* = t^* | I_i = 1, \mathcal{M}]}}{P[Y_i \in \mathcal{S} \mid I_i = 1, \mathcal{M}]\, P[T^* = t^* \mid I_i = 0, \mathcal{M}]} \tag{60}$$

$$= \frac{\sum_{y \in \mathcal{S}} \frac{P[T^* = t^* | Y_i = y, I_i = 1, \mathcal{M}]}{P[T^* = t^* | I_i = 0, \mathcal{M}]}\, P[Y_i = y \mid I_i = 1, \mathcal{M}]}{P[Y_i \in \mathcal{S} \mid I_i = 1, \mathcal{M}]} \tag{61}$$

By Lemma 1, to achieve the left bound in (25),

$$\frac{P[T^* = t^* \mid Y_i \in \mathcal{S}, I_i = 1, \mathcal{M}]}{P[T^* = t^* \mid I_i = 0, \mathcal{M}]} \geq \frac{\sum_{y \in \mathcal{S}} e^{-\varepsilon} P[Y_i = y \mid I_i = 1, \mathcal{M}]}{P[Y_i \in \mathcal{S} \mid I_i = 1, \mathcal{M}]} = e^{-\varepsilon}. \tag{62}$$

To achieve the right bound in (25),

$$\frac{P[T^* = t^* \mid Y_i \in \mathcal{S}, I_i = 1, \mathcal{M}]}{P[T^* = t^* \mid I_i = 0, \mathcal{M}]} \leq \frac{\sum_{y \in \mathcal{S}} e^{\varepsilon} P[Y_i = y \mid I_i = 1, \mathcal{M}]}{P[Y_i \in \mathcal{S} \mid I_i = 1, \mathcal{M}]} = e^{\varepsilon}. \tag{63}$$

We now turn to (26). First note that

$$\frac{P[T^* = t^* \mid Y_i \in \mathcal{S}, I_i = 1, \mathcal{M}]}{P[T^* = t^* \mid Y_i \notin \mathcal{S}, I_i = 1, \mathcal{M}]} = \frac{P[T^* = t^* \mid Y_i \in \mathcal{S}, I_i = 1, \mathcal{M}]}{P[T^* = t^* \mid I_i = 0, \mathcal{M}]} \cdot \frac{P[T^* = t^* \mid I_i = 0, \mathcal{M}]}{P[T^* = t^* \mid Y_i \notin \mathcal{S}, I_i = 1, \mathcal{M}]}. \tag{64}$$

By applying the bounds in (25) to $\mathcal{S}$ and $\mathcal{S}^C$,

$$e^{-\varepsilon} \leq \frac{P[T^* = t^* \mid Y_i \in \mathcal{S}, I_i = 1, \mathcal{M}]}{P[T^* = t^* \mid I_i = 0, \mathcal{M}]} \leq e^{\varepsilon} \quad \text{and} \quad e^{-\varepsilon} \leq \frac{P[T^* = t^* \mid Y_i \notin \mathcal{S}, I_i = 1, \mathcal{M}]}{P[T^* = t^* \mid I_i = 0, \mathcal{M}]} \leq e^{\varepsilon}. \tag{65}$$

Thus, to achieve the left inequality in (26),

$$\frac{P[T^* = t^* \mid Y_i \in \mathcal{S}, I_i = 1, \mathcal{M}]}{P[T^* = t^* \mid Y_i \notin \mathcal{S}, I_i = 1, \mathcal{M}]} \geq e^{-\varepsilon} \cdot (e^{\varepsilon})^{-1} = e^{-2\varepsilon}. \tag{66}$$

To achieve the right inequality in (26),

$$\frac{P[T^* = t^* \mid Y_i \in \mathcal{S}, I_i = 1, \mathcal{M}]}{P[T^* = t^* \mid Y_i \notin \mathcal{S}, I_i = 1, \mathcal{M}]} \leq e^{\varepsilon} \cdot \left(e^{-\varepsilon}\right)^{-1} = e^{2\varepsilon}. \tag{67}$$

□

The proof of the main theorem is provided below.

Theorem 2. *For individual i, fix the adversary's prior probabilities, p_i, q_i, and a desired bound on the relative disclosure risk, $r^*(p_i, q_i)$. Under the conditions of Theorem 1, any statistic $T^* = t^*$ released under ε-DP with*

$$\varepsilon \leq \begin{cases} \log\left(\frac{2p_i(1-q_i)}{\sqrt{(1-p_i)^2+4p_i(1-q_i)\left(\frac{1}{r^*(p_i,q_i)}-p_iq_i\right)}-(1-p_i)}\right), & \text{if } 0 < q_i < 1; \\ \log\left(\frac{1-p_i}{\frac{1}{r^*(p_i,1)}-p_i}\right), & \text{if } q_i = 1, \end{cases} \tag{68}$$

will satisfy $r_i(p_i, q_i, t^*) \leq r^*(p_i, q_i)$.

Proof. We begin with the simpler case where $q_i = 1$. By Theorem 1,

$$r_i(p_i, 1, t^*) \leq \frac{1}{p_i + e^{-\varepsilon}(1-p_i)}. \tag{69}$$

Since $\varepsilon \leq \log\left((1-p_i)/(1/r^*(p_i, 1) - p_i)\right)$, it follows that $e^{-\varepsilon} \geq (1/r^*(p_i, 1) - p_i)/(1-p_i)$. Thus, as desired,

$$r_i(p_i, 1, t^*) \leq \frac{1}{p_i + \frac{\frac{1}{r^*(p_i,1)}-p_i}{1-p_i}(1-p_i)} = \frac{1}{p_i + \left(\frac{1}{r^*(p_i,1)} - p_i\right)} = r^*(p_i, 1). \tag{70}$$

Now consider the case where $0 < q_i < 1$. By Theorem 1,

$$r_i(p_i, q_i, t^*) \leq \frac{1}{q_ip_i + e^{-2\varepsilon}(1-q_i)p_i + e^{-\varepsilon}(1-p_i)}. \tag{71}$$

Since

$$\varepsilon \leq \log\left(\frac{2p_i(1-q_i)}{\sqrt{(1-p_i)^2 + 4p_i(1-q_i)\left(\frac{1}{r^*(p_i,q_i)} - p_iq_i\right)} - (1-p_i)}\right), \tag{72}$$

it follows that

$$e^{-\varepsilon} \geq \frac{\sqrt{(1-p_i)^2 + 4p_i(1-q_i)\left(\frac{1}{r^*(p_i,q_i)} - p_iq_i\right)} - (1-p_i)}{2p_i(1-q_i)}. \tag{73}$$

Taking the square gives

$$e^{-2\varepsilon} \geq \frac{4p_i(1-q_i)\left(\frac{1}{r^*(p_i,q_i)} - p_iq_i\right) + 2(1-p_i)^2 - 2(1-p_i)\sqrt{(1-p_i)^2 + 4p_i(1-q_i)\left(\frac{1}{r^*(p_i,q_i)} - p_iq_i\right)}}{4p_i^2(1-q_i)^2} \tag{74}$$

$$= \frac{\frac{1}{r^*(p_i,q_i)} - p_iq_i}{p_i(1-q_i)} + \frac{(1-p_i)^2 - (1-p_i)\sqrt{(1-p_i)^2 + 4p_i(1-q_i)\left(\frac{1}{r^*(p_i,q_i)} - p_iq_i\right)}}{2p_i^2(1-q_i)^2}. \tag{75}$$

Thus,

$$
\begin{aligned}
&q_ip_i + e^{-2\varepsilon}(1-q_i)p_i + e^{-\varepsilon}(1-p_i) && (76)\\
&\geq q_ip_i + \frac{\frac{1}{r^*(p_i,q_i)} - p_iq_i}{p_i(1-q_i)} \cdot (1-q_i)p_i \\
&\quad + \frac{(1-p_i)^2 - (1-p_i)\sqrt{(1-p_i)^2 + 4p_i(1-q_i)\left(\frac{1}{r^*(p_i,q_i)} - p_iq_i\right)}}{2p_i^2(1-q_i)^2} \cdot (1-q_i)p_i \\
&\quad + \frac{\sqrt{(1-p_i)^2 + 4p_i(1-q_i)\left(\frac{1}{r^*(p_i,q_i)} - p_iq_i\right)} - (1-p_i)}{2p_i(1-q_i)} \cdot (1-p_i) && (77)\\
&= q_ip_i + \left(\frac{1}{r^*(p_i,q_i)} - p_iq_i\right) \\
&\quad + \left((1-p_i) - \sqrt{(1-p_i)^2 + 4p_i(1-q_i)\left(\frac{1}{r^*(p_i,q_i)} - p_iq_i\right)}\right)\frac{1-p_i}{2p_i(1-q_i)} \\
&\quad - \left((1-p_i) - \sqrt{(1-p_i)^2 + 4p_i(1-q_i)\left(\frac{1}{r^*(p_i,q_i)} - p_iq_i\right)}\right)\frac{1-p_i}{2p_i(1-q_i)} \\
& && (78)\\
&= \frac{1}{r^*(p_i,q_i)}. && (79)
\end{aligned}
$$

It is then immediate that

$$
r_i(p_i,q_i,t^*) \leq \frac{1}{q_ip_i + e^{-2\varepsilon}(1-q_i)p_i + e^{-\varepsilon}(1-p_i)} \leq \frac{1}{\frac{1}{r^*(p_i,q_i)}} = r^*(p_i,q_i). \tag{80}
$$

□

We now prove the generalization of Theorem 1.3 in [15].

Corollary 1. *Under the conditions of Theorem 1, for all $p_i, q_i \in (0,1]$ and all t^*,*

$$
r_i(p_i,q_i,t^*) \leq e^{2\varepsilon}. \tag{81}
$$

Proof. First note that since, $0 \leq e^{-\varepsilon} \leq 1$, it follows that $e^{-2\varepsilon} \leq e^{-\varepsilon}$. Then, applying the result of Theorem 1,

$$
r_i(p_i,q_i,t^*) \leq \frac{1}{q_ip_i + e^{-2\varepsilon}(1-q_i)p_i + e^{-\varepsilon}(1-p_i)} \leq \frac{1}{q_ip_i + e^{-2\varepsilon}(1-q_i)p_i + e^{-2\varepsilon}(1-p_i)} \tag{82}
$$

Combining terms and using the fact that $q_ip_i \geq 0$ gives

$$
r_i(p_i,q_i,t^*) \leq \frac{1}{q_ip_i + e^{-2\varepsilon}(1-q_ip_i)} \leq \frac{1}{e^{-2\varepsilon} + q_ip_i(1-e^{-2\varepsilon})} \leq \frac{1}{e^{-2\varepsilon}+0} = e^{2\varepsilon}. \tag{83}
$$

□

We now prove the corollary of Theorem 2 in the case of $p_i = 1$.

Corollary 2. *Under the conditions of Theorem 2, if $p_i = 1$ and $0 < q_i < 1$, then any statistic $T^* = t^*$ released under ε-DP with*

$$\varepsilon \leq \frac{1}{2} \log \left(\frac{1 - q_i}{\frac{1}{r^*(1,q_i)} - q_i} \right), \tag{84}$$

will satisfy $r_i(1, q_i, t^*) \leq r^*(1, q_i)$.

Proof. Plugging $p_i = 1$ into the expression from Theorem 2 yields

$$\varepsilon \leq \log \left(\frac{2(1 - q_i)}{\sqrt{0 + 4(1 - q_i)\left(\frac{1}{r^*(1,q_i)} - q_i\right)} - 0} \right) = \log \left(\sqrt{\frac{1 - q_i}{\frac{1}{r^*(1,q_i)} - q_i}} \right) = \frac{1}{2} \log \left(\frac{1 - q_i}{\frac{1}{r^*(1,q_i)} - q_i} \right). \tag{85}$$

□

Proofs for Lemmas 3–4 and Theorems 3–6 will be included here.