Whose Data Is It Anyway? Towards a Formal Treatment of Differential Privacy for Surveys

James Bailie & Jörg Drechsler*

May 13, 2024

This is currently a working paper. Some parts of this paper remain unfinished. Comments and discussion are welcome; constructive criticism will be appreciated.

1 Introduction

The survey is the workhorse of statistical agencies. For example, the U.S. Census Bureau conducts more than 100 surveys annually [U.S. Census Bureau, 2023] including key data collections such as the American Community Survey (ACS), the Current Population Survey (CPS), the Survey of Income and Program Participation (SIPP) and the new Annual Business Survey (ABS). The data gathered through these surveys provide invaluable information on the US economy and on American society more generally. They are used by various stakeholders – for example, businesses, researchers, local and federal governments, media and not-for-profits – to make investment decisions, to inform policy and to allocate government funding, among many other uses.

On the other hand, national statistical organizations (NSOs) have acknowledged for decades their obligation to maintain the confidentiality of survey respondents due to legal and ethical considerations, but also to safeguard institutional trust and thus sustain the quality of their data products. To address these conflicting goals, various methods have been proposed over the years to protect the confidentiality of survey respondents while still maintaining the value of the data for the different stakeholders involved. In the last two decades, a new framework for assessing the privacy of statistical data products has emerged: *differential privacy* (DP) [Dwork et al., 2006b]. This framework is mathematically appealing as it offers a formal guarantee: any single unit's influence on the probabil-

*jamesbailie@g.harvard.edu, joerg.drechsler@iab.de

ity of observing a specific output is bounded. This guarantee translates into quantifiable measures of protection against an adversary seeking to learn the confidential responses, although not without some complications [Cuff and Yu, 2016, Tschantz et al., 2020, Kifer and Machanavajjhala, 2011, Bailie et al., 2024+b]. Beyond the formal guarantees, DP is attractive as it allows for full transparency of the methods that were used to protect the output, without further loss to respondents' privacy beyond that associated with publishing the outputted statistics and the DP specification. This is in contrast to many methods that are currently employed at statistical agencies, which rely on hiding some of the parameters of the privacy-protection mechanism (such as the variance of the noise term when noise addition is used to protect a continuous attribute) to ensure privacy. With DP, agencies typically release all details about the mechanism including the levels of the privacy parameters. This implies that – at least in principle – users of the protected data will be able to account for the additional uncertainty introduced through the protection step (although this turns out to be difficult for many of the algorithms used in practice so far). Finally, DP offers several additional properties such as immunity to postprocessing and composition of privacy budget (see for example Dwork and Roth [2014]). This second property makes the DP framework specifically interesting for statistical agencies as it allows for the quantification of the privacy loss over multiple data releases.

These attractive features have motivated the adoption of DP in the private sector [Erlingsson et al., 2014, Apple's Differential Privacy Team, 2017, Ding et al., 2017, Messing et al., 2020, Uber Security, 2017], as well as at some NSOs such as the Census Bureau [Machanavajjhala et al., 2008, Foote et al., 2019, Abowd and Hawes, 2023]. Still, all deployments of DP so far have focused on situations in which the data to be protected coincided with the population of interest. As pointed out above, this is rarely the case for data collected by NSOs. Except for censuses – which are typically only conducted every five to ten years – and some administrative databases, most data at statistical agencies are collected via probability surveys. In the survey context, information is only gathered from a small fraction of the population, but the careful design of the selection process and several adjustment steps after the survey has been conducted (such as weighting, editing and imputation) ensure that the resulting data can be used to obtain approximately unbiased estimates for the population of interest. (We will offer a more detailed review of the survey process in Section 2.3.) However, how to properly account for these particularities within the framework of DP is currently poorly understood (see Reiter [2019] and Drechsler [2023] for an in-depth discussion of the challenges that will arise in this context). Gaining a better understanding is especially critical as the Census Bureau has publicly committed to adopting DP for all its data products [US Census Bureau, 2018] – a resolution that has been recently reaffirmed in US Census Bureau [2022]. (In fact, the Census Bureau only recommitted to adopting "formal privacy"; however we are not aware of any other formal privacy frameworks for statistical data apart for DP.) In the same 2022 press release, the Census Bureau concluded that "the science does not yet exist" to implement DP for their flagship

survey – the ACS – highlighting the need for additional research in this area.

We are aware of only a few papers that address DP in the survey context and, moreover, all these papers only focus on specific aspects of this process. Lin et al. [2023] study how to estimate the mean of a binary variable under DP assuming stratified sampling using proportional allocation and simple random sampling within the strata. Bun et al. [2022] investigate if the complex sampling designs commonly used in the survey context can offer increased privacy protection building on previous results showing that simple sampling procedures such as simple random sampling or Poisson sampling will amplify the privacy protection [Balle et al., 2020]. We will summarize their findings in Section 4.1. Finally, in some preliminary work, Das et al. [2022] study the effects of imputation. They find that if DP is only considered when analyzing the imputed data, the required privacy loss budget can increase linearly with the number of missing cases. They also show that this problem can be avoided – at least for certain imputation schemes – if DP is already considered during imputation.

This paper aims to establish a framework for DP in the survey context by discussing the implications of (for example) whether the privacy guarantees should hold only for the sampled units or the entire population. We identify ten settings that vary in their assumptions about the data at different levels (the responding sample, the selected sample, the sampling frame, and the target population). Building on the framework introduced in Bailie et al. [2024+a], we formalize the DP flavors for these settings and discuss their implications on both data utility and privacy.

2 Background

2.1 Notation

We typically denote sets by upper-case calligraphic letters (for example, $\mathcal{S}, \mathcal{T}$ or $\mathcal{D}$) and sets of sets by upper-case script letters (for example, $\mathscr{D}$ or $\mathscr{F}$). Datasets are denoted by fraktur lower-case letters (for example, $\mathfrak{d}, \mathfrak{d}', \mathfrak{p}, \mathfrak{f}$ or $\mathfrak{s}$) when they are not stochastic, and by upper-case letters (for example, $\mathfrak{D}, \mathfrak{D}', \mathfrak{P}, \mathfrak{F}$ or $\mathfrak{S}$) when they are random variables. In general, we follow the convention that lower-case letters denote realizations of the corresponding upper-case random variable. However, we use the sans-serif superscript $^{\mathsf{R}}$ to denote a random set (for example, $\mathcal{S}^{\mathsf{R}}$); an upper-case calligraphic letter without this superscript often denotes a realization of the corresponding random set (for example, $\mathcal{S}$ denotes a realization of the random set $\mathcal{S}^{\mathsf{R}}$).

A record r is a set of attributes and a dataset $\mathfrak{d}$ is a set of records. Every record r is associated with a unit, which we denote by $u(r)$. The units of a dataset $\mathfrak{d}$ are given by the set $\mathcal{U}(\mathfrak{d}) = \{u(r) \mid r \in \mathfrak{d}\}$. We assume throughout that every unit is associated with at most one record in any given dataset, although a unit will often have multiple records spread across different datasets. The unique record

in the dataset $\mathfrak{d}$ associated with unit $i \in \mathcal{U}(\mathfrak{d})$ is denoted by $\mathfrak{d}_i$.

As an example, a unit could be a person, and the attributes of a record could describe some of the characteristics of that person, such as their age, income and occupation, as well as some identifiers, such as their name and address. Alternatively, a unit could be a company, and a record associated with a company could detail some business characteristics of that company. Less frequently, a unit may represent a group of people, or a population – in this way, we can encode population-level information in a dataset. Occasionally, it will be important to distinguish between the unit – which is an abstraction – and the real-world entity that is represented by the unit. Beyond their philosophical differences, discrepancies between a unit's data and the corresponding entity's characteristics can arise due to measurement error, non-response or imputation. Moreover, there can be multiple units which represent the same entity. Such over-counting can occur when, for example, units are constructed from a register of addresses (or phone numbers, identification numbers, etc.) because a single entity can have multiple addresses. Duplication is a common problem in surveying, particularly in the context of business statistics, and – as we will see – poses a complication for DP.

An attribute is a value of a variable. More exactly, an attribute of a unit i is the value of a variable that is taken by i. (For example, an attribute could be the value `40` and the associated variable could be `Age` (in years). This would signify that unit i has an age of 40 years.) Therefore, a record r is uniquely specified by its unit $u(r)$ and the variables associated to its attributes. Denote the set of the variables in a record r by $\mathcal{V}(r)$ and the variables in a dataset $\mathfrak{d}$ by $\mathcal{V}(\mathfrak{d}) = \bigcup_{i \in \mathcal{U}(\mathfrak{d})} \mathcal{V}(\mathfrak{d}_i)$. Although we do not require it, usually every record in a dataset has the same variables: $\mathcal{V}(\mathfrak{d}) = \mathcal{V}(\mathfrak{d}_i)$ for all $i \in \mathcal{U}(\mathfrak{d})$.

Given a set of units $\mathcal{U}$ and a set of variables $\mathcal{V}$, let $\mathfrak{d}(\mathcal{U}, \mathcal{V})$ denote the dataset $\{r \mid u(r) \in \mathcal{U}, \mathcal{V}(r) = \mathcal{V}\}$. This dataset $\mathfrak{d}(\mathcal{U}, \mathcal{V})$ is well-defined because every record is determined by its variables and its unit. Given a variable x and a unit i, let x_i denote i's value of the variable x. We can re-express $\mathfrak{d}(\mathcal{U}, \mathcal{V})$ as

$$\mathfrak{d}(\mathcal{U}, \mathcal{V}) = \Big\{ \{x_i \mid x \in \mathcal{V}\} \;\Big|\; i \in \mathcal{U} \Big\}.$$

2.2 Differential Privacy

DP studies data-release mechanisms – functions T which take as input a dataset $\mathfrak{d}$ and a random seed ω, and output a stochastic summary $T(\mathfrak{d}, \omega)$ of $\mathfrak{d}$.

Definition 2.1. A *data-release mechanism* is a function $T : \mathcal{D}_0 \times \Omega \to \mathcal{T}$ where

- $\mathcal{D}_0$ is the data space, the set of all theoretically-possible datasets $\mathfrak{d}$;
- Ω is the probability space of the seed ω with σ-algebra $\mathscr{F}_\Omega$ and probability P;
- $\mathcal{T}$ is equipped with a σ-algebra $\mathscr{F}_\mathcal{T}$; and

- $T(\mathfrak{d}, \cdot)$ is measurable for all $\mathfrak{d} \in \mathcal{D}_0$.

(See Bailie et al. [2024+a] for a slightly more general definition and for additional context.) ◇

Intuitively speaking, $\mathfrak{d}$ is the data that is considered confidential and hence must not be disclosed by the summary $T(\mathfrak{d}, \omega)$. DP measures how the probabilistic noise induced by the seed ω masks this input dataset $\mathfrak{d}$.

We emphasize that, in order for T to be well-defined (as a function $\mathcal{D}_0 \times \Omega \to \mathcal{T}$), its input $\mathfrak{d}$ must contain all the data which has a non-zero probability (with respect to P) of being used by T. That is to say, the output $T(\mathfrak{d}, \omega)$ can only depend on data which is in $\mathfrak{d}$, or data that is generated from $\mathfrak{d}$ and ω, but not on other data. While it may seem we are belaboring an obvious point – of course, by definition $T(\mathfrak{d}, \omega)$ cannot be a function of anything but $\mathfrak{d}$ and ω – the input dataset $\mathfrak{d}$ is surprisingly slippery to specify in the context of surveying, as we illustrate with the following simplistic example.

Example 2.2. Suppose that a government agency is conducting a survey on the health of people in Massachusetts. The agency has a list of Massachusettsans (a *frame* $\mathfrak{f}$, see Subsection 2.3 below) from which they will randomly select a sample of individuals. They will then collect data $\mathfrak{S}$ on some of the health characteristics of the sampled individuals (e.g. blood pressure, heart rate, etc.) and publish some aggregate statistics based on these collected data.

As we will expand upon later in this article, the agency may decide to include the sampling procedure in their data-release mechanism T, since this can potentially increase the efficiency of the privacy-utility tradeoff (see Subsection 4.1). In this case, T takes as input the frame $\mathfrak{f}$; it "performs" the sampling and data collection steps outlined above; and then it calculates and outputs the aggregate statistics. There are two options for how T can "collect" the data $\mathfrak{S}$. The first option is that the data $\mathfrak{S}$ is generated (or modelled, depending on one's perspective) within the data-release mechanism T – i.e. $\mathfrak{S}$ is a function of T's input data and seed. The second option is that the data $\mathfrak{S}$ is itself included as part of T's input data.

We will see in Section 5 that the DP guarantee does not necessarily apply to data generated within a DP mechanism – it only applies to the mechanism's input data.[1] Hence, the first option is not appropriate if we want to guarantee the privacy protection of the sampled individuals' health characteristics. We must therefore resort to the second option and include the data $\mathfrak{S}$ as input to T. However, we do not know a-priori which individuals will be sampled. Since any individual in the frame $\mathfrak{f}$ has a non-zero probability of being sampled, any of the records in $\mathfrak{d}(\mathcal{U}(\mathfrak{f}), \mathcal{V}(\mathfrak{S}))$ may appear in the sample data $\mathfrak{S}$. As such, all of these records must be included as input – that is, T requires as input $\mathfrak{f}^* = \mathfrak{d}(\mathcal{U}(\mathfrak{f}), \mathcal{V}(\mathfrak{f}) \cup \mathcal{V}(\mathfrak{S}))$.

[1] This discussion is still missing at this point, but will be included in the final version of the paper.

We refer to $\mathfrak{f}^*$ as the *augmented frame*, since it includes all the variables that are collected in the survey as well as all the frame variables. In the context of survey sampling, $\mathfrak{f}^*$ is never observed. Yet, it must nevertheless serve as input to any data-release mechanism T, whenever T includes a sampling step and we wish to provide the sample data with a DP guarantee. The data in $\mathfrak{d}(\mathcal{U}(\mathfrak{f}), \mathcal{V}(\mathfrak{S}))$ are not available to the government agency at the time it starts its data collection. Rather, $\mathfrak{d}(\mathcal{U}(\mathfrak{f}), \mathcal{V}(\mathfrak{S}))$ is the 'theoretical' dataset from which the agency collects the survey data.

While the input $\mathfrak{f}^*$ described above can be observed if the agency surveyed all units in the frame, in some situations it is not even theoretically possible to observe the input to a DP data-release mechanism. It is not uncommon that a survey includes a minor intervention as part of its data collection. For example, the Massachusetts health survey could require administering an oral glucose load as part of a glucose tolerance test in the diagnosis of diabetes [Phillips, 2012], or it could direct the survey respondent to exercise on a stationary bike as part of a cardiac stress test [Bruce and McDonough, 1969]. Alternatively, in the context of a medical trial, the sampled individuals could be randomly assigned to receive a treatment or a placebo. In these cases, the data we wish to protect – the outcomes of these health tests – are only realized during the data collection process. When this data collection process is included within the data-release mechanism – as must necessarily be the case when the data-release mechanism T includes the sampling step of the survey – these data cannot possibly be included as input into T, because they do not even exist at the time the data-release mechanism begins! (One may argue that the process of any data collection or measurement – such as checking blood pressure – is itself an intervention and the collected data only come into existence at the point of collection. Under this perspective, the following remarks apply to all data.)

In such cases, the input data must necessarily include the potential outcome of each of the possible interventions (or treatments). To those familiar with causal inference, the dataset of these potential outcomes is known as the *science table* [Rubin, 2005]. The science table is never fully observable because the potential outcome under a counterfactual treatment is always unknown. Yet, if we want to protect the outcome under the non-counterfactual treatment – which is unknown at the start of T – we must include it as input to T, and we can only ensure it is included as input if we include all the potential outcomes as input.

We end this example by noting that, if T does not include a sampling step, then T need not include the data collection step either. As such, T's input data is simply the collected data, without concern to the counterfactual potential outcomes. ◇

It is convenient to think of a data-release mechanism as a function $\mathfrak{d} \mapsto \mathsf{P}_{\mathfrak{d}}(T \in \cdot)$. Here the probability distribution $\mathsf{P}_{\mathfrak{d}}(T \in \cdot)$ of the summary $T(\mathfrak{d}, \omega)$ is the push-forward measure induced by

the distribution P of the random seed $\omega \in \Omega$, taking $\mathfrak{d}$ as fixed:

$$\mathsf{P}_{\mathfrak{d}}(T \in E) := \mathsf{P}(\{\omega \in \Omega : T(\mathfrak{d}, \omega) \in E\}),$$

where $E \in \mathscr{F}_{\mathcal{T}}$ is any measurable subset of the output space $\mathcal{T}$. DP is the condition that the data-release mechanism is Lipschitz continuous – i.e. that the distance $d_{\mathrm{Pr}}(\mathsf{P}_{\mathfrak{d}}, \mathsf{P}_{\mathfrak{d}'})$ between outputs $\mathsf{P}_{\mathfrak{d}}$ and $\mathsf{P}_{\mathfrak{d}'}$ is at most a multiplicative factor of the distance $d_{\mathcal{D}_0}(\mathfrak{d}, \mathfrak{d}')$ between the corresponding inputs $\mathfrak{d}$ and $\mathfrak{d}'$.

Example 2.3. For pure ε-DP, as defined in Dwork et al. [2006b], the multiplicative factor is ε; the distance between inputs $\mathfrak{d}$ and $\mathfrak{d}'$ is the Hamming distance; and the distance between outputs $\mathsf{P}_{\mathfrak{d}}$ and $\mathsf{P}_{\mathfrak{d}'}$ is the multiplicative distance:

$$d_{\mathrm{MULT}}(\mathsf{P}_{\mathfrak{d}}, \mathsf{P}_{\mathfrak{d}'}) = \sup_{E \in \mathscr{F}_{\mathcal{T}}} \left| \ln \frac{\mathsf{P}_{\mathfrak{d}}(T \in E)}{\mathsf{P}_{\mathfrak{d}'}(T \in E)} \right|,$$

(For readers that are familiar with the definition of pure ε-DP in terms of neighboring datasets $\mathfrak{d}$ and $\mathfrak{d}'$, the Lipschitz condition for non-neighbors is implied by group privacy. Hence, the neighbor definition of pure ε-DP is the equivalent to the above definition.)

For approximate (ε, δ)-DP [Dwork et al., 2006a], the multiplicative factor is again ε; the distance between inputs is given by

$$d_{\mathcal{D}_0}^{\text{neighbors}}(\mathfrak{d}, \mathfrak{d}') = \begin{cases} 0 & \text{if } \mathfrak{d} = \mathfrak{d}', \\ 1 & \text{if } \mathfrak{d} \text{ and } \mathfrak{d}' \text{ are neighbors}, \\ \infty & \text{otherwise;} \end{cases}$$

and the distance between outputs is given by

$$d_{\mathrm{MULT}}^{\delta}(\mathsf{P}_{\mathfrak{d}}, \mathsf{P}_{\mathfrak{d}'}) = \sup_{E \in \mathscr{F}_{\mathcal{T}}} \left\{ \ln \frac{[\mathsf{P}_{\mathfrak{d}}(T \in E) - \delta]^+}{\mathsf{P}_{\mathfrak{d}'}(T \in E)}, \ln \frac{[\mathsf{P}_{\mathfrak{d}'}(T \in E) - \delta]^+}{\mathsf{P}_{\mathfrak{d}}(T \in E)}, 0 \right\},$$

(where $[x]^+ = \max\{x, 0\}$). Note that $d_{\mathcal{D}_0}^{\text{neighbors}}$ and $d_{\mathrm{MULT}}^{\delta}$ are not distances in the mathematical sense of a metric; we will instead refer to them as *premetrics* from herein. Since $d_{\mathrm{MULT}}^{\delta}$ does not satisfy the triangle inequality, approximate (ε, δ)-DP's group privacy budget does not increase linearly with the group size; hence we cannot replace $d_{\mathcal{D}_0}^{\text{neighbors}}$ with the Hamming distance, as we did for pure ε-DP. ◇

By definition, a data-release mechanism T satisfies DP if it is Lipschitz continuous. There are different *flavors* (i.e. types or versions) of DP; each of these flavors correspond to different ways

to specify continuity. For our purposes, there are four components to the specification of Lipschitz continuity. Most obviously, there are the premetrics $d_{\mathcal{D}_0}(\mathfrak{d}, \mathfrak{d}')$ and $d_{\mathrm{Pr}}(\mathsf{P}_{\mathfrak{d}}, \mathsf{P}_{\mathfrak{d}'})$. These premetrics measure the 'distance' between any two inputs $\mathfrak{d}$ and $\mathfrak{d}'$, or between any two output probabilities $\mathsf{P}_{\mathfrak{d}}$ and $\mathsf{P}_{\mathfrak{d}'}$. Secondly, there is the domain $\mathcal{D}_0$ of the data-release mechanism, which – as we shall see – serves as the parameter space of the attacker's inferential model.[2] Finally, there is the data multiverse $\mathscr{D}$, which allows the data custodian to restrict the Lipschitz continuity condition to certain pairs of inputs – as is often desirable in practice. For example, we may only want to compare samples drawn from the same population. This restriction is achieved by specifying the data multiverse $\mathscr{D}$.

Definition 2.4 (Bailie et al. [2024+a])**.** A *differential privacy flavor* is a quadruple $(\mathcal{D}_0, \mathscr{D}, d_{\mathcal{D}_0}, d_{\mathrm{Pr}})$ where:

1. The *domain* $\mathcal{D}_0$ is the *data space* – the set of all (theoretically-possible) input datasets.
2. The *multiverse* $\mathscr{D} \subset 2^{\mathcal{D}_0}$ is a set of *universes*, which are denoted by $\mathcal{D}$ or $\mathcal{D}'$.
3. The *input premetric* $d_{\mathcal{D}_0}$ is a premetric on $\mathcal{D}_0$ – i.e. a function $\mathcal{D}_0 \times \mathcal{D}_0 \to \mathbb{R}^{\geq 0}$ such that $d_{\mathcal{D}_0}(\mathfrak{d}, \mathfrak{d}) = 0$ for all $\mathfrak{d} \in \mathcal{D}_0$.
4. The *output premetric* d_{Pr} is a premetric on the space of all probability distributions $\mathcal{P}$ – i.e. a function $\mathcal{P} \times \mathcal{P} \to \mathbb{R}^{\geq 0}$ of probabilities $\mathsf{P}, \mathsf{Q} \in \mathcal{P}$ such that
   - $d_{\mathrm{Pr}}(\mathsf{P}, \mathsf{P}) = 0$ for all $\mathsf{P} \in \mathcal{P}$; and
   - $d_{\mathrm{Pr}}(\mathsf{P}, \mathsf{Q}) = \infty$ for probabilities P, Q which live on different measurable spaces. ◇

Once we have specified the four components for Lipschitz continuity via a DP flavor, we also need to specify the multiplicative constant (known as the Lipschitz constant) which controls the rate between input and output variations. Together, choices for these five components are called a DP specification:

Definition 2.5. A *differential privacy specification* is a quintuple $(\mathcal{D}_0, \mathscr{D}, d_{\mathcal{D}_0}, d_{\mathrm{Pr}}, \varepsilon_{\mathcal{D}})$ consisting of a DP flavor $(\mathcal{D}_0, \mathscr{D}, d_{\mathcal{D}_0}, d_{\mathrm{Pr}})$ and a privacy-loss budget $\varepsilon_{\mathcal{D}} : \mathscr{D} \to \mathbb{R}^{\geq 0}$. We denote a DP specification by $\varepsilon_{\mathcal{D}}$-DP$(\mathcal{D}_0, \mathscr{D}, d_{\mathcal{D}_0}, d_{\mathrm{Pr}})$.

A data-release mechanism $T : \mathcal{D}_0 \times \Omega \to \mathcal{T}$ *satisfies* the DP specification $\varepsilon_{\mathcal{D}}$-DP$(\mathcal{D}_0, \mathscr{D}, d_{\mathcal{D}_0}, d_{\mathrm{Pr}})$ if, for all data universes $\mathcal{D} \in \mathscr{D}$, and all $\mathfrak{d}, \mathfrak{d}' \in \mathcal{D}$,

$$d_{\mathrm{Pr}}\big[\mathsf{P}_{\mathfrak{d}}(T \in \cdot), \mathsf{P}_{\mathfrak{d}'}(T \in \cdot)\big] \leq \varepsilon_{\mathcal{D}} d_{\mathcal{D}_0}(\mathfrak{d}, \mathfrak{d}'). \tag{2.1}$$

[2] This discussion is still missing at this point, but will be included in the final version of the paper.

Let $\mathcal{M}(\mathcal{D}_0, \mathscr{D}, d_{\mathcal{D}_0}, d_{\mathrm{Pr}}, \varepsilon_{\mathcal{D}})$ denote the set of data-release mechanisms which satisfy the DP specification $\varepsilon_{\mathcal{D}}$-DP$(\mathcal{D}_0, \mathscr{D}, d_{\mathcal{D}_0}, d_{\mathrm{Pr}})$. ◇

For the purposes of understanding DP in the context of survey sampling, the relevant components of a DP flavor $(\mathcal{D}_0, \mathscr{D}, d_{\mathcal{D}_0}, d_{\mathrm{Pr}})$ are its domain $\mathcal{D}_0$ and its multiverse $\mathscr{D}$.

We need the following technical definition:

Definition 2.6. Let $\mathcal{D}_0$ be a domain and $\mathscr{D}, \mathscr{D}' \subset 2^{\mathcal{D}_0}$ be two multiverses of $\mathcal{D}_0$. We say $\mathscr{D}'$ is a *coarsening* of $\mathscr{D}$ if, for all $\mathcal{D} \in \mathscr{D}$, there exists $\mathcal{D}' \in \mathscr{D}'$ with $\mathcal{D} \subset \mathcal{D}'$. ◇

When $\mathscr{D}'$ is a coarsening of $\mathscr{D}$ we write $\mathscr{D} \leq \mathscr{D}'$. The following lemma justifies this notation by establishing that $\mathscr{D}$ is a weaker condition than $\mathscr{D}'$ if $\mathscr{D} \leq \mathscr{D}'$.

Lemma 2.7. *Let $\mathcal{D}_0$ be a domain and $\mathscr{D}, \mathscr{D}' \subset 2^{\mathcal{D}_0}$ be multiverses such that $\mathscr{D} \leq \mathscr{D}'$. Then, for all budgets $\varepsilon_{\mathcal{D}'} : \mathscr{D}' \to \mathbb{R}^{\geq 0}$,*

$$\mathcal{M}(\mathcal{D}_0, \mathscr{D}', d_{\mathcal{D}_0}, d_{\mathrm{Pr}}, \varepsilon_{\mathcal{D}'}) \subset \mathcal{M}(\mathcal{D}_0, \mathscr{D}, d_{\mathcal{D}_0}, d_{\mathrm{Pr}}, \varepsilon_{\mathcal{D}}),$$

where $\varepsilon_{\mathcal{D}} = \inf\{\varepsilon_{\mathcal{D}'} : \mathcal{D}' \in \mathscr{D}' \text{ s.t. } \mathcal{D} \subset \mathcal{D}'\}$.

Definition 2.8. Given a DP flavor $(\mathcal{D}_0, \mathscr{D}, d_{\mathcal{D}_0}, d_{\mathrm{Pr}})$, the multiverse $\mathscr{D}$ is *complete* if $d_{\mathcal{D}_0}(\mathfrak{d}, \mathfrak{d}') < \infty$ for all $\mathfrak{d}, \mathfrak{d}' \in \mathcal{D}$ and all $\mathcal{D} \in \mathscr{D}$. ◇

Definition 2.9. Given a DP flavor $(\mathcal{D}_0, \mathscr{D}, d_{\mathcal{D}_0}, d_{\mathrm{Pr}})$, two datasets $\mathfrak{d}, \mathfrak{d}' \in \mathcal{D}_0$ are *comparable* when 1) $\mathfrak{d} \neq \mathfrak{d}'$; 2) $d_{\mathcal{D}_0}(\mathfrak{d}, \mathfrak{d}') < \infty$ or $d_{\mathcal{D}_0}(\mathfrak{d}', \mathfrak{d}) < \infty$; and 3) there exists a data universe $\mathcal{D} \in \mathscr{D}$ such that $\mathfrak{d}, \mathfrak{d}' \in \mathcal{D}$. ◇

Definition 2.10. Given a DP flavor $(\mathcal{D}_0, \mathscr{D}, d_{\mathcal{D}_0}, d_{\mathrm{Pr}})$, denote the protection objects *connected* to $\mathfrak{d} \in \mathcal{D}_0$ by

$$[\mathfrak{d}] = \{\mathfrak{d}' \in \mathcal{D}_0 : d_{\mathcal{D}_0}(\mathfrak{d}, \mathfrak{d}') < \infty\}.$$

Then the *completion* $\overline{\mathscr{D}}$ of the data multiverse $\mathscr{D}$ is defined as

$$\overline{\mathscr{D}} = \{\mathcal{D} \cap [\mathfrak{d}] : \mathcal{D} \in \mathscr{D}, \mathfrak{d} \in \mathcal{D}\}.$$ ◇

Lemma 2.11. *Let $(\mathcal{D}_0, \mathscr{D}, d_{\mathcal{D}_0}, d_{\mathrm{Pr}})$ be a DP flavor where $d_{\mathcal{D}_0}$ is a metric. Then, the completion $\overline{\mathscr{D}}$ of $\mathscr{D}$ is complete and, for all budgets $\varepsilon_{\mathcal{D}} : \mathscr{D} \to \mathbb{R}^{\geq 0}$,*

$$\mathcal{M}(\mathcal{D}_0, \overline{\mathscr{D}}, d_{\mathcal{D}_0}, d_{\mathrm{Pr}}, \varepsilon_{\mathcal{D}'}) = \mathcal{M}(\mathcal{D}_0, \mathscr{D}, d_{\mathcal{D}_0}, d_{\mathrm{Pr}}, \varepsilon_{\mathcal{D}}),$$

where

$$\varepsilon_{\mathcal{D}'} = \inf\{\varepsilon_{\mathcal{D}} : \mathcal{D} \in \mathscr{D} \text{ s.t. } \mathcal{D}' \subset \mathcal{D}\}.$$

2.3 Survey Sampling

Surveys are conducted to learn some characteristics of a well-defined population by collecting information from a random subset of this population. Most survey sampling processes rely on three key ingredients: the *target population* of interest; the *sampling frame* from which the random sample to be surveyed is drawn; and the *sampling design* for drawing this sample.

The target population (also known as the universe in some survey sampling texts, although we will not use this term to avoid confusion with the notion of a universe $\mathcal{D}$ in a DP flavor) is the scope of the survey; it is the population the survey is aiming to learn about. It is typically defined conceptually, while the sampling frame $\mathfrak{f}$, on the other hand, is an existent register containing names (or other identifiers), contact information (postal or physical address, email, and/or telephone number) and possibly some basic demographic information of the survey units. The sampling frame $\mathfrak{f}$ serves as the source from which the sample is drawn. For the discussions in the remainder of the paper it is important to clearly distinguish between the target population and the sampling frame $\mathfrak{f}$. While the sampling frame aims to cover all the units from the target population, it might include units that are not part of the target population (*overcoverage*), and it might also miss units that should be included (*undercoverage*). To formalize the difference between the target population and the sampling frame, we suppose that the frame is not constructed from the target population data, but from a fixed dataset we term the *pseudo-population* dataset $\mathfrak{p}$. Typically, the frame is constructed from previous censuses' data, administrative records and canvassing. The pseudo-population dataset $\mathfrak{p}$ is the collection of all such data, so that $\mathcal{U}(\mathfrak{f}) \subset \mathcal{U}(\mathfrak{p})$. By introducing the concept of the pseudo-population, we allow for undercoverage and overcoverage, as well as duplications in the frame (where a single unit in the target population corresponds to multiple units in the frame).

The sample is the set $\mathcal{U}^{\mathsf{R}}(\mathfrak{S})$ of units of the sample dataset $\mathfrak{S}$. The sample is a random set whose distribution is given by the sampling design. The sampling design is defined as a probability measure $\tau_{\mathfrak{f}}$ on $2^{\mathcal{U}(\mathfrak{f})}$. The units $\mathcal{U}^{\mathsf{R}}(\mathfrak{S})$ of the sample dataset $\mathfrak{S}$ are a draw from $\tau_{\mathfrak{f}}$. That is, $\mathcal{U}^{\mathsf{R}}(\mathfrak{S}) \sim \tau_{\mathfrak{f}}$. For each subset $\mathcal{S} \subset \mathcal{U}(\mathfrak{f})$, $\tau(\mathcal{S})$ is the probability that the realized sample $\mathcal{U}(\mathfrak{s})$ is $\mathcal{S}$. Sometimes the frame $\mathfrak{f}$ contains basic demographic information on the survey units, which can be used to construct the sample selection probabilities $\tau(\mathcal{S})$. The sampling designs used in practice are often complex multi-stage designs, with different sampling strategies (e.g. cluster sampling, stratified sampling, probability-proportional-to-size (PPS) sampling) for each of the different stages. In determining the sample design $\tau_{\mathfrak{f}}$, the frame $\mathfrak{f}$ is usually taken into consideration, which can complicate the deployment of DP.

To illustrate the relevance of this discussion, we look at the Current Population Survey (CPS) conducted by the Census Bureau for the Bureau of Labor Statistics (BLS). The *target population*

of the CPS is the civilian noninstitutionalized population in the US, or, more exactly,

> "all people residing in the 50 states [of the US] and the District of Columbia who are not confined to institutions such as nursing homes and prisons, and who are not on active duty in the US Armed Forces. Included are citizens of foreign countries who reside in the United States but do not live on the premises of an embassy. The civilian noninstitutional population ages 16 and older is the base population group used for CPS statistics" [U.S. Bureau of Labor Statistics, 2018].

The survey uses two different *sampling frames*: one for households and one for group quarters. Both are derived from the master address file (MAF) of the Census Bureau: "The MAF is a national inventory of addresses that is continually updated by the U.S. Census Bureau to support its decennial programs and demographic surveys" [U.S. Census Bureau, 2019]. The CPS uses a stratified two-stage *sampling design*. In the first stage, the population is divided into geographical clusters and one cluster is sampled within each stratum using PPS sampling. A small group of households is selected in the second stage using systematic sampling based on a list sorted by demographic composition and geographic proximity. (See Section 2.2 in U.S. Census Bureau [2019] for a full description of the sampling methodology.)

2.4 Survey Weights

A distinctive feature of survey data is that they typically contain survey weights. Survey weights are provided by statistical agencies as a convenient tool to account for the sampling design and additional data preparation steps such as nonresponse adjustments when analyzing the data. Because complex sampling designs are often used (as we described in the previous section) and because not all sampled units actually respond to the survey, the resulting dataset cannot be treated as a simple random sample from the target population. Most estimators need to be adjusted to take these complications into account. For example, the (unweighted) sample mean can no longer be treated as an unbiased estimator for the mean in the population if the probability of being included in the responding-sample varies between the units. Instead, it is typical to use weighted estimators, where individual data points are weighted according to their survey weights.

Survey weights are typically generated in three stages. In the first stage, *design weights* are generated that reflect the sampling design. In the second stage, *nonresponse adjustment weights* are used to account for different response propensities in different subgroups of the population. Finally, *calibration weights* try to correct for any deficiencies in the sampling frame and also help to reduce the variance of the final estimates.

The design weights w^D are defined as the inverse of the probability of selection: $w_i^D = 1/\pi_i$, where π_i is the probability that unit $i \in \mathcal{U}(\mathfrak{f})$ is selected into the sample $\mathcal{U}^{\mathsf{R}}(\mathfrak{S})$. Nonresponse adjustment

weights try to adjust for potential biases that might arise due to unequal response propensities. The idea is to estimate the probability to respond for each unit. The nonresponse adjustment weights w^{NR} are calculated as the inverse of the estimated response probabilities p_i^R; that is, $w_i^{NR} = 1/p_i^R$ for $i \in \mathcal{U}^{\mathsf{R}}(\mathfrak{S})$. Note that the response probabilities can be used to compute the final probability to be included in the sample: $p_i^{(inc)} = \pi_i p_i^R$. Hence, the inverse of $p_i^{(inc)}$ can be used as a weight that accounts for both the complex sampling design and the nonresponse.

The final weighting step is commonly to calibrate the survey data to information that is known about the population of interest from other sources. For example, the total number of people living in the U.S. by age and gender might be known from the previous Census. Common calibration techniques are post-stratification, raking or the GREG estimator. Describing the details of these adjustment methods is beyond the scope of this paper (see Valliant et al. [2018] for further details). It suffices to note that all these methods can be reflected by adjusting the survey weights obtained from the previous two steps.

3 DP Flavors for Survey Statistics

As we have seen in Subsection 2.3, there are multiple phases in the creation of survey statistics: defining the target population, compiling the sampling frame, selecting the sample according to the sampling design, and collecting data from the responding units. (From herein, we use the term 'target sample' to refer to the sample of units selected by the sampling design from the sampling frame, in order to differentiate this sample with the responding sample – the set of units which were selected and responded.) The data output by each phase of this pipeline is fed into the subsequent phase as input. For example, data about the target population is used to compile the frame and data on the frame is used to select the sample.

The data custodian (e.g. the NSO) could plausibly start the data-release mechanism T at any point along this data pipeline. That is, the data-release mechanism could take as input the dataset corresponding to any of the various phases. Moreover, the custodian could also plausibly condition on previous phases in the data pipeline (taking their data as invariant). Thus, the data custodian is faced with two decisions: what should the protection domain $\mathcal{D}_0$ be? And what should the data multiverse $\mathscr{D}$ be?

In this section, we formalize the various options for these two decisions in terms of their corresponding DP flavors. In Sections 4 and 5, we show why these two decisions are important by describing the consequences of each option on both data utility and privacy.

Definition 3.1. Let $\mathcal{D}_0^{\text{pp}}$ be the set of all possible pseudo-population datasets; $\mathcal{D}_0^{\text{fr}}$ the set of all possible frames (from all possible pseudo-populations); $\mathcal{D}_0^{\text{samp}}$ the set of all possible target sample

datasets (from all possible frames); and $\mathcal{D}_0^{\text{resp}}$ the set of all possible responding sample datasets (from all possible target samples). We say that a DP flavor $(\mathcal{D}_0, \mathscr{D}, d_{\mathcal{D}_0}, d_{\Pr})$ is *population-level* if $\mathcal{D}_0 = \mathcal{D}_0^{\text{pp}}$. The definitions of *frame-level*, *(target-)sample-level* and *responding-sample-level* DP flavors are analogous. $\diamond$

In the above definition, we have been deliberately vague in specifying $\mathcal{D}_0^{\text{pp}}$. The precise definition of the set $\mathcal{D}_0^{\text{pp}}$ depends on the data custodian's assessment of what pseudo-populations are considered 'possible'. In general, 'possible' should be interpreted liberally, so that this set $\mathcal{D}_0^{\text{pp}}$ is generously large. (See Section 5 for an explanation of why this matters and Bailie et al. [2024+a] for a more extensive discussion.)

We can be more specific in the definition of $\mathcal{D}_0^{\text{fr}}$, since the construction of a frame is a real-world process undertaken by an NSO (although in practice this process is often messy, complex and hard to precisely describe). This process takes as input a pseudo-population $\mathfrak{p} \in \mathcal{D}_0^{\text{pp}}$ and outputs a frame for that population. Then $\mathcal{D}_0^{\text{fr}}$ is the set of all outputs from this process, across all possible pseudo-populations $\mathfrak{p} \in \mathcal{D}_0^{\text{pp}}$.

When defining the set $\mathcal{D}_0^{\text{samp}}$ of all possible samples, we assume that there is a given sampling design $\tau_{\mathfrak{f}}$ and we only consider those sample datasets $\mathfrak{s}$ with non-zero probability $\tau_{\mathfrak{f}}(\mathcal{U}(\mathfrak{s})) > 0$. However, as is frequently the case, the sampling design $\tau_{\mathfrak{f}}$ can depend on the realized frame $\mathfrak{f}$. (For example, the stratum sample sizes are part of a stratified sampling design, and these sizes are partially based on the sizes of the strata in the frame $\mathfrak{f}$.) Thus,

$$\mathcal{D}_0^{\text{samp}} = \{\mathfrak{s} : \tau_{\mathfrak{f}}(\mathcal{U}(\mathfrak{s})) > 0, \mathfrak{f} \in \mathcal{D}_0^{\text{fr}}\}.$$

Definition 3.2 (Primitive data multiverses)**.** Define the primitive data multiverses:

1. $\mathscr{D}_{\text{fr}|\text{pp}} = \{\mathcal{D}_{\mathfrak{p}} : \mathfrak{p} \in \mathcal{D}_0^{\text{pp}}\}$, where $\mathcal{D}_{\mathfrak{p}}$ is the set of all possible frames constructed from the pseudo-population $\mathfrak{p} \in \mathcal{D}_0^{\text{pp}}$;

2. $\mathscr{D}_{\text{samp}|\text{pp}} = \{\mathcal{D}_{\mathfrak{p}} : \mathfrak{p} \in \mathcal{D}_0^{\text{pp}}\}$, where $\mathcal{D}_{\mathfrak{p}}$ is the set of all possible target sample datasets drawn from the pseudo-population $\mathfrak{p} \in \mathcal{D}_0^{\text{pp}}$:

$$\mathcal{D}_{\mathfrak{p}} = \{\mathfrak{s} : \tau_{\mathfrak{f}}(\mathcal{U}(\mathfrak{s})) > 0, \mathfrak{f} \text{ is a possible frame constructed from the pseudo-population } \mathfrak{p}\}.$$

3. $\mathscr{D}_{\text{samp}|\text{fr}} = \{\mathcal{D}_{\mathfrak{f}} : \mathfrak{f} \in \mathcal{D}_0^{\text{fr}}\}$, where $\mathcal{D}_{\mathfrak{f}}$ is the set of all possible target samples drawn from the frame $\mathfrak{f} \in \mathcal{D}_0^{\text{fr}}$:

$$\mathcal{D}_{\mathfrak{f}} = \{\mathfrak{s} : \tau_{\mathfrak{f}}(\mathcal{U}(\mathfrak{s})) > 0\}.$$

4. The data multiverses $\mathscr{D}_{\text{resp}|\text{pp}}, \mathscr{D}_{\text{resp}|\text{fr}}$ and $\mathscr{D}_{\text{resp}|\text{samp}}$ can be defined analogously, as the set of

$\mathcal{D}_0^{\text{pp}}$	population agnostic			
$\mathcal{D}_0^{\text{fr}}$	population agnostic	population invariant		
$\mathcal{D}_0^{\text{samp}}$	+frame agnostic	+frame agnostic	+frame invariant	
$\mathcal{D}_0^{\text{resp}}$	+sample agnostic	+sample agnostic	+sample agnostic	+sample invariant

Table 3.1: Overview of the possible settings for the different levels.

data universes $\mathcal{D}_{\mathfrak{d}}$, with $\mathcal{D}_{\mathfrak{d}}$ the set of all possible responding samples drawn from, respectively, the population, frame, or target sample $\mathfrak{d}$. ◇

Definition 3.3 (Population-, frame- and sample-invariance)**.** A DP flavor $(\mathcal{D}_0, \mathscr{D}, d_{\mathcal{D}_0}, d_{\text{Pr}})$ is:

1. *population-invariant* if $\overline{\mathscr{D}} \leq \mathscr{D}_{\cdot|\text{pp}}$, where
   - for frame-level flavors: $\mathscr{D}_{\cdot|\text{pp}} = \mathscr{D}_{\text{fr}|\text{pp}}$,
   - for sample-level flavors: $\mathscr{D}_{\cdot|\text{pp}} = \mathscr{D}_{\text{samp}|\text{pp}}$, and
   - for responding-sample-level flavors: $\mathscr{D}_{\cdot|\text{pp}} = \mathscr{D}_{\text{resp}|\text{pp}}$;
2. *frame-invariant* if $\overline{\mathscr{D}} \leq \mathscr{D}_{\cdot|\text{fr}}$, where
   - for sample-level flavors: $\mathscr{D}_{\cdot|\text{fr}} = \mathscr{D}_{\text{samp}|\text{fr}}$, and
   - for responding-sample-level flavors: $\mathscr{D}_{\cdot|\text{fr}} = \mathscr{D}_{\text{resp}|\text{fr}}$;
3. *sample-invariant* if $\overline{\mathscr{D}} \leq \mathscr{D}_{\text{resp}|\text{samp}}$ (for responding-sample-level flavors). ◇

The intuition behind these definitions is very simple. The idea is to restrict the comparable datasets (Definition 2.9). Population-invariance means that comparable frames (or samples or responding samples) must be from the same pseudo-population. (That is, a pair of frames are comparable only if they are constructed from the same pseudo-population.) Analogously, frame-invariance means that comparable sample datasets must be drawn from the same frame.

If a DP flavor is not population-invariant (resp. frame-invariant or sample-invariant), then we say it is *population-agnostic* (resp. *frame-agnostic* or *sample-agnostic*). Frame-agnosticism implies that there are two comparable samples which are drawn from different frames.

Because invariance at one level implies invariance at previous data pipeline phases, we identify ten settings (which together exhaust the potential options for where the DP mechanism starts and which phases are taken as invariant): one setting for population-level flavors; two for frame-level flavors; three for sample-level flavors; and four for responding-sample-level flavors (see Table 3.1 for illustration.)

4 Utility Considerations

In this section we consider the possible implications of the different DP flavors on the achievable level of accuracy of the noisy outcome given a desired level of privacy (expressed by fixing the privacy parameters). Two components are relevant when evaluating the accuracy for DP estimates from survey data: the privacy amplification effects from sampling, which imply that less noise needs to be infused to achieve a given privacy level and the increased sensitivity of the weighted estimator (where weights are included to account for the sampling design, nonresponse, and potentially for other data deficiencies such as over- or undercoverage of the sampling frame), which typically implies that more noise is required. We discuss the effects of the different flavors on both components in the following chapters.

4.1 Privacy Amplification via Sampling

Previous research has shown that simple sampling designs offer privacy amplification, that is, the privacy offered when running a DP algorithm on a random subset of the population is higher than if the same algorithm with the same privacy parameters is run on the full population. Balle et al. [2018] proof the following theorem for simple random sampling with replacement (they also obtain similar results for Poisson sampling and simple random sampling without replacement):

Theorem 4.1 (Balle et al. [2018])**.** *Let $\mathcal{C}$ be a sampling scheme that uniformly randomly samples n values out of N possible values without replacement. Given an (ε, δ)-bounded differentially private mechanism $\mathcal{M}$, we have that $\mathcal{M} \circ \mathcal{C}$ is (ε', δ')-bounded differentially private for $\varepsilon' = \log(1 + \frac{n}{N}[e^{\varepsilon} - 1])$ and $\delta' = \frac{n}{N}\delta$.*

In this theorem bounded differential privacy refers to the scenario in which neighboring datasets are obtained by changing the values of one record in the data while keeping the size of the data fixed. Note that for small ε and small sampling rates this implies that $\varepsilon' \approx n/N\varepsilon$, i.e., the amplification is proportional to the sampling rate. Based on these results Bun et al. [2022] studied to what extent privacy amplification can also be achieved for the more complex sampling designs commonly used at statistical agencies. Their findings can be summarized as follows:

- Cluster sampling using simple random sampling without replacement to draw the clusters offers negligible amplification in practice except for small ε (less than 0.5) and very small cluster sizes (less than 15 units).
- With minor adjustments, stratified sampling using proportional allocation can provide privacy amplification.
- Data dependent allocation functions such as Neyman allocation for stratified sampling will

likely result in privacy degradation (the effects will depend on the sensitivity of the allocation function).

- With PPS sampling at the individual level, the privacy amplification will linearly depend on the maximum probability of inclusion (for small ε).
- Systematic sampling will only offer amplification if the ordering of the population is truly random. In all other cases, systematic sampling will suffer from the same effects as cluster sampling leading to no amplification (assuming the ordering is known to the attacker).

In practice this implies that for the multi-stage sampling designs that typically start with (multiple stages of) stratified cluster sampling, amplification effects can generally only be expected from those stages at which individual units are selected (typically the last stage of selection).

4.2 Privacy Amplification for Different DP Flavors

Before discussing the implications of the DP flavors introduced in Section 3, it is important to consider at which stages of the data production pipeline amplification effects could occur. Conceptually, three different sampling steps can be defined when moving from the population to the responding sample. The most obvious step (and the only one that is fully controlled) is the selection of the target sample from the sampling frame. However, if nonresponse is treated as a stochastic process (as is commonly done in the survey literature), moving from the target to the responding sample can be interpreted as another sampling step. The same is true when moving from the pseudo-population to the sampling frame if we assume that each unit in the pseudo-population has a certain probability to be included in the frame. Still, the amplification effects of these two steps are difficult to take into account in practice as the inclusion probabilities are unknown and would need to be estimated. Errors when modeling these probabilities would lead to invalid statements regarding the amplification effects. Besides, the amplification effects when moving from the pseudo-population to the sampling frame will typically be negligible given that the probability to be included in the frame should be well above 90% for high quality frames.

Considering the DP flavors, we can distinguish four scenarios: If the responding sample dataset is given as input, the DP mechanism can only be applied at the responding sample level. This scenario boils down to the standard setting considered in most DP papers. There is no (sub)sampling step within the data release mechanism T and thus there is no amplification effect. Interestingly, this scenario offers the same privacy guarantees as the more restrictive assumption that the attacker knows who participated in the survey. In all other scenarios, privacy is amplified through the (sub)sampling process.

In the second scenario, the DP flavor is at the target sample-level. In this scenario, amplification can

only arise from the subsampling step when moving from the target sample to the responding sample. As response rates are often less than 20% in practice, this subsampling might offer some privacy amplification. However, as mentioned earlier, quantifying this effect will be difficult in practice as response probabilities are unknown and will likely differ between the units. In the third scenario, the (augmented) frame $\mathfrak{f}^*$ is taken as input to the data-release mechanism. This scenario will offer privacy amplification as discussed in Bun et al. [2022] in addition to the theoretical amplification offered from nonresponse. Finally, if the DP flavor has domain $\mathcal{D}_0^{\text{PP}}$, a third layer of amplification is possible by moving from the pseudo-population to the sampling frame. As discussed above, this layer will typically be negligible for sampling frames commonly used in practice.

4.3 Weighting

Using weighted estimators generally increases the amount of noise that needs to be added to achieve a desired level of privacy protection. This is because the sensitivity of the result, i.e., the maximum possible change in the result when changing a single record, increases when incorporating the survey weights. To illustrate, we can consider the simple example of a counting query. A counting query simply counts the number of units in a database that satisfy a given set of conditions, for example, the total number of unemployed men between 30 and 40. Counting queries are attractive under DP as they have low sensitivity and thus require limited amounts of noise to achieve DP (as the noise scales with the sensitivity of the query). Under unbounded DP (i.e., defining neighboring datasets by adding or removing one record) the sensitivity of a counting query is 1.

In the survey literature a counting query is called a total and the most convenient way to estimate a total for complex sampling designs is to use the Horvitz-Thompson estimator [Horvitz and Thompson, 1952], which provides approximately unbiased estimates for most sampling designs. The Horvitz-Thompson estimator for a total is given as $\hat{t}_x = \sum_{i \in \mathcal{U}^{\mathsf{R}}(\mathfrak{S})} w_i x_i$, where $\hat{t}_x$ is the estimated total in the population for the target variable x and w_i is the survey weight for unit i. In our example, x_i is a binary indicator which equals 1 if unit i satisfies the conditions of interest (i.e. unit i is unemployed, male, and between 30 and 40 years old) and is zero otherwise. Using the Horvitz-Thompson estimator, the L_1-sensitivity increases to $\max(w_i)$ (where the maximum is taken either over the records in the sample (under target sample invariance), the frame (under frame invariance) or over the entire population (under population invariance)).

Since the amount of noise that is required typically scales with the sensitivity of the output, this implies that much more noise needs to be added when trying to protect a weighted survey estimate. However, the considerations so far assume that the weights can be considered fixed. This assumption is never justified for the final survey weights. This is because the nonresponse adjustments and calibration steps rely on models that are estimated from the data. Changing one record in the data will change these models and thus the weights. How to account for this variability in the final

DP Setting	Effects on Design Weights
Target sample invariance	Can be treated as fixed
Frame invariance	Can be treated as fixed
Population invariance	Sensitivity needs to be considered
Population agnostic	Sensitivity needs to be considered

Table 4.1: Overview of the implications on the design weights for different types of invariance. We note that the final weights that also account for nonresponse can never be treated as fixed.

DP Setting	Amplification from
Responding-sample level $\mathcal{D}_0^{\text{resp}}$	–
Target-sample level $\mathcal{D}_0^{\text{samp}}$	NR
Frame level $\mathcal{D}_0^{\text{fr}}$	NR&S
Population level $\mathcal{D}_0^{\text{pp}}$	FR&S&NR

Table 4.2: Overview of the implications on privacy amplification for different levels of DP. (The abbreviations are NR=nonresponse,S=sampling,FR=frame).

weights when computing the sensitivity of a survey weighted estimate has not been addressed in the DP literature so far.

But even if we only consider the design weights, the assumption of constant weights is only justified, if changing one record in the database does not change the probability of inclusion for any of the records in the pseudo-population. Whether this is a realistic assumption will depend on the DP flavor to be considered but also on the properties of the sampling design.

In general, the design weights can only be treated as fixed under the frame-invariant or target sample invariant scenario. In all other scenarios the weights will typically change. How much the weights will change will depend on the sensitivity of the sampling design, which in turn depends on how data dependent the sampling design is. To illustrate, data dependence will be small for single stage cluster sampling designs especially if the clusters are selected using simple random sampling (such a design is used for example for the German Microcensus). For such a design, the probability of selection does not change over neighboring frames (as long as the definition of the clusters does not change). On the other hand, PPS sampling will generally be highly data dependent as the probability of selection directly depends on some features of the data. This will be less problematic if PPS sampling is used to select the clusters as the probability of selection will only depend on the size of the clusters and these sizes will only change by one record over neighboring databases. However, if PPS sampling is used to select individual units, the probabilities of selection can change arbitrarily over neighboring datasets. Thus, for these designs the sensitivity of the final estimate might increase considerably and it seems difficult to correctly quantify this sensitivity in practice.

Tables 4.1 and 4.2 summarize the implications of the different DP flavors considered in this paper. Together they highlight the inherent trade-off between the various flavors of DP for survey estimators. For example, considering the frame as invariant implies that the DP flavor is at the target- or responding-sample level and hence no utility improvements through amplification by sampling can be achieved. On the other hand, frame invariance allows treating the weights as fixed, which will generally reduce the sensitivity of the final estimates and thus the noise that needs to be added to ensure privacy. For the other flavors, utility improvements could be achieved through privacy amplification, but this benefit comes at the cost that the sensitivity of the weights needs to be considered. This increase might outweigh the benefits of amplification from sampling, especially since as Bun et al. [2022] have shown, the amplification effects tend to be small for sampling designs commonly used in practice. Which of the flavors will be most attractive from a utility perspective will crucially depend on the sampling design in practice as the design will have an effect on both the amplification and the sensitivity of the weights. It will also depend on the question whether response probabilities can be determined reliably.

4.4 Sensitivity Reduction from the Sampling Design

When the DP flavor is frame-invariant, the sampling design $\tau_{\mathfrak{f}}$ can reduce the sensitivity of a query such as the Horvitz-Thompson estimator. This is because only samples with non-zero probability are considered. Comparable sample datasets $\mathfrak{s}, \mathfrak{s}'$ must both have non-zero probability of being realized under the same sampling design $\tau_{\mathfrak{f}}$. This restricts the number of comparable sample datasets, and hence potentially reduces the sensitivity of a query.

For example, if the sampling design $\tau_{\mathfrak{f}}$ includes stratification, then the stratum sample sizes are constant between comparable sample datasets $\mathfrak{s}, \mathfrak{s}'$. Thus, if $\mathfrak{s}$ and $\mathfrak{s}'$ differ only on a single record, that record must belong to the same stratum in both $\mathfrak{s}$ and $\mathfrak{s}'$. When the difference between the possible values of x_i within strata is smaller than their difference across strata (which typically is the case whenever stratification is used to reduce the uncertainty in survey estimates), the sensitivity of the Horvitz-Thompson estimatoris reduced when the DP flavor is frame-invariant.

4.5 Utility Implications for the Horvitz-Thompson Estimator

In this section, we use the Horvitz-Thompson estimator $\hat{t}_x = \sum_{i \in \mathcal{U}^{\mathsf{R}}(\mathfrak{S})} w_i x_i$ discussed in Section 4.3 to illustrate the utility implications of the different settings. For simplicity, we assume the output of $\hat{t}_x$ is protected using the Laplace mechanism (we do not claim this mechanism is optimal for this estimator).

4.5.1 The Laplace Mechanism for the Horvitz-Thompson Estimator

If the Horvitz-Thompson estimator must be differentially private, the corresponding Laplace mechanism can be used in place of $\hat{t}_x$.

Definition 4.2 (Dwork et al. [2006b])**.** Let $\varepsilon_{\mathcal{D}}$-DP$(\mathcal{D}_0, \mathscr{D}, d_{\mathcal{D}_0}, d_{\Pr})$ be a DP specification with $d_{\Pr} = d_{\text{MULT}}$. Suppose $q : \mathcal{D}_0 \to \mathbb{R}^k$ is a non-stochastic function. The *Laplace mechanism corresponding to* q is the data-release mechanism

$$T_{q,\text{LAP}}(\mathfrak{d}, \omega) = q(\mathfrak{d}) + \Delta_q([\mathfrak{d}]_{\mathscr{D}})\omega,$$

where

- the seed $\omega \in \mathbb{R}^k$ is a vector of k iid Laplace random variables, each with PDF $f(\omega_i) = \frac{1}{2}\exp(-|\omega_i|)$,
- $[\mathfrak{d}]_{\mathscr{D}}$ is the connected component

$$[\mathfrak{d}]_{\mathscr{D}} = \{\mathfrak{d}' \in \mathcal{D}_0 \mid \text{there exists } \mathcal{D} \in \mathscr{D} \text{ s.t. } \mathfrak{d}, \mathfrak{d}' \in \mathcal{D} \text{ and } d_{\mathcal{D}_0}(\mathfrak{d}, \mathfrak{d}') < \infty\},$$

- for $\mathcal{D}^* \subset \mathcal{D}_0$, $\Delta_q(\mathcal{D}^*)$ is the ε-adjusted L_1-sensitivity

$$\Delta_q(\mathcal{D}^*) = \sup_{\substack{\mathcal{D} \in \mathscr{D} \\ \mathcal{D} \subset \mathcal{D}^*}} \sup_{\mathfrak{d}, \mathfrak{d}' \in \mathcal{D}} \frac{\|q(\mathfrak{d}) - q(\mathfrak{d}')\|_1}{\varepsilon_{\mathcal{D}} d_{\mathcal{D}_0}(\mathfrak{d}, \mathfrak{d}')},$$

(with $\|\cdot\|_1$ the L_1-norm, $0/0 := 0$ and $\sup \emptyset := 0$). ◇

Theorem 4.3. *The Laplace mechanism $T_{q,\text{LAP}}$ satisfies $\varepsilon_{\mathcal{D}}$-DP$(\mathcal{D}_0, \mathscr{D}, d_{\mathcal{D}_0}, d_{\Pr})$.*

4.5.2 Sensitivity of the Horvitz-Thompson Estimator

Suppose that $d_{\mathcal{D}_0}$ is the Hamming distance; the budget $\varepsilon_{\mathcal{D}} = \varepsilon$ is constant in $\mathcal{D}$; and q is the Horvitz-Thompson estimator $\hat{t}_x = \sum_{i \in \mathcal{U}^{\text{R}}(\mathfrak{S})} w_i x_i$, with w_i the design weights. Consider sample-level DP: the domain $\mathcal{D}_0$ is the set of all possible samples $\mathfrak{s}$. For

$$\mathcal{D}_{\mathfrak{f}} = \{\mathfrak{s} : \tau_{\mathfrak{f}}(\mathcal{U}(\mathfrak{s})) > 0\},$$

define the (unadjusted) L_1-sensitivity as

$$\Delta_q(\mathcal{D}_{\mathfrak{f}}) = \sup_{\mathfrak{s}, \mathfrak{s}' \in \mathcal{D}_{\mathfrak{f}}} |q(\mathfrak{s}) - q(\mathfrak{s}')|.$$

In this section, we will prove that the L_1-sensitivity $\Delta_q(\mathcal{D}_\mathfrak{f})$ is bounded by $|\max_{i\in\mathfrak{f}}(w_i x_i) - \min_{i\in\mathfrak{f}}(w_i x_i)|$. This is the relevant L_1-sensitivity for frame-invariant DP flavors. For frame-agnostic DP flavors, the relevant L_1-sensitivity is the global L_1-sensitivity $\Delta_q(\mathcal{D}_0)$, which can only be bounded by the worst-case

$$|\max w_i x_i - \min w_i x_i| + (n-1)(\max w_i - \min w_i)(|\max x_i| \vee |\min x_i|),$$

where n is the (fixed) size of the target sample and the maximums and minimums are all over $i \in \mathcal{U}(\mathfrak{p})$ and all possible $\mathfrak{p}$ because, in general, changing a single record may change the design weights of all other records.

5 Privacy Considerations

5.1 Privacy Semantics

5.1.1 Posterior-to-Posterior Comparisons

The aim of the posterior-to-posterior framework is to compare what an attacker would learn about a single unit, if this unit is included in the input dataset relative to a counterfactual world in which the unit is not included or his or her record is not used.

Adopting notation similar to that of Kifer et al. [2022], let P_A be the attacker's prior on the domain $\mathcal{D}_0$, i.e., the prior implies that the input dataset is treated as a random variable $\mathfrak{D}$ on the space $\mathcal{D}_0$. The goal of the attacker is to infer information about the record $\mathfrak{D}_i$ of a single unit $i \in \mathcal{U}^\mathsf{R}(\mathfrak{D})$ in the input dataset $\mathfrak{D}$. For this to be well-defined, we must assume that the units of $\mathfrak{D}$ are fixed (that is, $\mathcal{U}(\mathfrak{D})$ is a fixed set). A common practice in the literature is to assume that the units of $\mathfrak{D}$ are identified by the indices $1, \ldots, n$, where $n = |\mathcal{U}(\mathfrak{D})|$. Throughout this section we assume $d_{\mathcal{D}_0}$ is the Hamming distance. For simplicity, we also assume that $\mathcal{D}_0$ and $\mathcal{T}$ are countable spaces.

Let $t \in \mathcal{T}$ denote a realized output of the data release mechanism T. The posterior-to-posterior framework as adopted in Kifer et al. [2022] compares the posterior distribution $\mathsf{P}_\mathsf{A}(\mathfrak{D}_i \in \cdot \mid T(\mathfrak{D}) = t)$ with the counterfactual world in which the information of the selected unit is replaced by a random draw from the posterior distribution of the attacker assuming knowledge of everybody else. Let $psample[\mathfrak{d}] \sim \mathsf{P}_\mathsf{A}(\mathfrak{D} \mid \mathfrak{D}^- = \mathfrak{d}^-)$ denote this random draw, where $\mathfrak{D}^-$ and $\mathfrak{d}^-$ denote the random variable $\mathfrak{D}$ and dataset $\mathfrak{d}$ with the selected record ($\mathfrak{D}_i$ or $\mathfrak{d}_i$ respectively) being removed. As shown in Kifer et al. [2022] the ratio of these two posteriors is given by

$$\frac{\mathsf{P}_\mathsf{A}(\mathfrak{D}_i = r \mid T(\mathfrak{D}) = t)}{\mathsf{P}_\mathsf{A}(\mathfrak{D}_i = r \mid T(psample[\mathfrak{D}]) = t)} = \frac{\sum_{\mathfrak{d}^-} \mathsf{P}_\mathsf{A}(\mathfrak{d}^-)\mathsf{P}_\mathsf{A}(r \mid \mathfrak{d}^-)\mathsf{P}_{\mathfrak{d}^- \cup \{r\}}(T = t)}{\sum_{\mathfrak{d}^-} \mathsf{P}_\mathsf{A}(\mathfrak{d}^-)\mathsf{P}_\mathsf{A}(r \mid \mathfrak{d}^-) \sum_{r'} \mathsf{P}_\mathsf{A}(r' \mid \mathfrak{d}^-)\mathsf{P}_{\mathfrak{d}^- \cup \{r'\}}(T = t)}. \quad (5.1)$$

For ε-DP,

$$e^{-\varepsilon} \leq \frac{\mathsf{P}_{\mathfrak{d}^- \cup \{r\}}(T = t)}{\mathsf{P}_{\mathfrak{d}^- \cup \{r'\}}(T = t)} \leq e^{\varepsilon},$$

and hence the ratio of posteriors (5.1) is bounded between $e^{-\varepsilon}$ and e^{ε}, for all possible values r of $\mathfrak{D}_i$ (see Theorem 7.1 in Kifer et al. [2022]).

5.1.2 Implications for the Different Settings

The posterior-to-posterior semantics apply to the possible values r of a record $\mathfrak{d}_i$ from the input dataset $\mathfrak{d} \in \mathcal{D}_0$, which varies depending on the DP setting. Of particular importance is the domain $\mathcal{D}_0$ of the DP flavor, since this determines what dataset – the (augmented) pseudo-population dataset $\mathfrak{p}^*$, the (augmented) frame $\mathfrak{f}^*$, the sample dataset $\mathfrak{s}$, or the responding-sample dataset $\mathfrak{r}$ – is protected. Although not explicitly stated, the classical framework considered in most of the DP literature assumes the responding-sample-level setting, in which the domain $\mathcal{D}_0$ is the set of possible responding-sample datasets, $\mathcal{D}_0^{\text{resp}}$. In this case, the data-release mechanism takes as input the fixed responding sample $\mathfrak{r}$. As such, the protections supplied by the data-release mechanism – as measured by the posterior-to-posterior framework – apply to a record $\mathfrak{r}_i$ from the responding sample. That is, an ε-DP mechanism with domain $\mathcal{D}_0^{\text{resp}}$ ensures that the posterior-to-posterior ratio for a responding sample record $\mathfrak{r}_i$ is bounded in the interval $[e^{-\varepsilon}, e^{\varepsilon}]$.

If we change the DP flavor to be at the frame-level – so that we may benefit from privacy amplification by sampling – then the input to the data-release mechanism is the augmented frame $\mathfrak{f}^*$. As such, an ε-DP mechanism under this setting will protect an augmented frame record $\mathfrak{f}_i^*$ – rather than a responding sample record $\mathfrak{r}_i$ – within the nominal interval $[e^{-\varepsilon}, e^{\varepsilon}]$. This distinction is important, because protection at one level does not imply protection at another level. In fact, we will see in Subsection 5.1.4 that whenever there is privacy amplification due to sampling, a sample record's posterior-to-posterior ratio is not bounded within $[e^{-\varepsilon}, e^{\varepsilon}]$ for an ε-DP mechanism at the frame-level.

Beyond looking at the different starting points of the data release mechanism, it is also important to consider the impacts of different types of invariances. For example, treating the frame as invariant implies that neighboring datasets must come from the same fixed frame. This enforces restrictions on the possible values r of $\mathfrak{D}_i$. As a consequence two data release mechanisms that start at the same level, for example, $\mathcal{D}_0^{\text{fr}}$ and use the same privacy loss budget ε, will offer different privacy guarantees, if one of them is frame-invariant while the other is frame-agnostic. This illustrates the ever existing trade-off between utility and privacy. From a utility perspective, it seems desirable to identify scenarios, in which enforcing invariance substantially restricts the possible values of $\mathfrak{D}_i$ as this might considerably reduce the sensitivity of the query of interest. On the other hand, shrinkingn the data univeres $\mathcal{D} \in \mathscr{D}$ will implicitly reduce the privacy guarantees even if the privacy

loss parameter is held constant.

5.1.3 No Privacy Amplification if the Attacker Knows that Unit i Is in the Sample

In this section, we show that we cannot hope for privacy amplification by sampling if we assume that the attacker knows that unit i is included in the sample $\mathcal{U}^{\mathsf{R}}(\mathfrak{S})$. This is a risk scenario that statistical agencies commonly need to consider in practice. In the statistical disclosure control literature, this is often referred to as the "nosy neighbor" scenario, since a possible scenario in which this kind of knowledge is realistic is the situation in which a neighbor witnesses an interviewer entering the house next door and then hopes to learn sensitive information about the neighbor by trying to reidentify him or her in the data.

To illustrate, we consider a data-release mechanism that starts at the frame level and thus should offer privacy amplification from sampling. Specifically, suppose that T is a data-release mechanism at the sample-level and let $\mathcal{S}(\cdot)$ be the sampling function, which takes an frame $\mathfrak{f}$ and outputs the sample according to the given sample design τ. (That is, $\mathsf{P}(\mathcal{S}(\mathfrak{f}) = \mathcal{S}) = \tau(\mathcal{S})$ for all $\mathcal{S} \subset \mathcal{U}(\mathfrak{f})$.) The data-release mechanism that starts at the frame level is therefore the composition $T' = T \circ \mathcal{S}$. Conditioning on the fact that unit i is included in the sample, the lower bound of the posterior-to-posterior ratio under the assumption that T is ε-DP is (below we write $\mathcal{S}^{\mathsf{R}}$ for the sample $\mathcal{U}^{\mathsf{R}}(\mathfrak{S})$):

$$
\begin{aligned}
&\frac{\mathsf{P}_{\mathsf{A}}(\mathfrak{F}_i^* = r \mid T'(\mathfrak{F}^*) = t, i \in \mathcal{S}^{\mathsf{R}})}{\mathsf{P}_{\mathsf{A}}(\mathfrak{F}_i^* = r \mid T'(psample[\mathfrak{F}^*]) = t, i \in \mathcal{S}^{\mathsf{R}})} \\
&= \frac{\sum_{\mathfrak{f}^{*-}} \mathsf{P}_{\mathsf{A}}(\mathfrak{f}^{*-} \mid i \in \mathcal{S}^{\mathsf{R}}) \mathsf{P}_{\mathsf{A}}(r \mid \mathfrak{f}^{*-}, i \in \mathcal{S}^{\mathsf{R}}) \mathsf{P}_{\mathfrak{f}^{*-} \cup \{r\}}(T' = t \mid i \in \mathcal{S}^{\mathsf{R}})}{\sum_{\mathfrak{f}^{*-}} \mathsf{P}_{\mathsf{A}}(\mathfrak{f}^{*-} \mid i \in \mathcal{S}^{\mathsf{R}}) \mathsf{P}_{\mathsf{A}}(r \mid \mathfrak{f}^{*-}, i \in \mathcal{S}^{\mathsf{R}}) \sum_{r'} \mathsf{P}_{\mathsf{A}}(r' \mid \mathfrak{f}^{*-}, i \in \mathcal{S}^{\mathsf{R}}) \mathsf{P}_{\mathfrak{f}^{*-} \cup \{r'\}}(T' = t \mid i \in \mathcal{S}^{\mathsf{R}})} \\
&\geq \frac{\sum_{\mathfrak{f}^{*-}} \mathsf{P}_{\mathsf{A}}(\mathfrak{f}^{*-} \mid i \in \mathcal{S}^{\mathsf{R}}) \mathsf{P}_{\mathsf{A}}(r \mid \mathfrak{f}^{*-}, i \in \mathcal{S}^{\mathsf{R}}) \mathsf{P}_{\mathfrak{f}^{*-} \cup \{r\}}(T' = t \mid i \in \mathcal{S}^{\mathsf{R}})}{\sum_{\mathfrak{f}^{*-}} \mathsf{P}_{\mathsf{A}}(\mathfrak{f}^{*-} \mid i \in \mathcal{S}^{\mathsf{R}}) \mathsf{P}_{\mathsf{A}}(r \mid \mathfrak{f}^{*-}, i \in \mathcal{S}^{\mathsf{R}}) e^{\varepsilon} \mathsf{P}_{\mathfrak{f}^{*-} \cup \{r\}}(T' = t \mid i \in \mathcal{S}^{\mathsf{R}})} \\
&= e^{-\varepsilon}.
\end{aligned}
$$

Whenever the mechanism T is optimal (i.e. it achieves the bound $\mathsf{P}_{\mathfrak{s}}(T = t)/\mathsf{P}_{\mathfrak{s}'}(T = t) = \epsilon$ for some $\mathfrak{s}, \mathfrak{s}'$ with $d_{\mathcal{D}_0}(\mathfrak{s}, \mathfrak{s}') = 1$), the above inequality is achieved for some choice r and i. Using a similar argument, the upper bound of the ratio is e^{ε}. Thus, while the data release mechanism T' satisfies ε'-DP for $\varepsilon' < \varepsilon$, the posterior-to-posterior protection provided by T' when the attacker knows $i \in \mathcal{U}^{\mathsf{R}}(\mathfrak{S})$ is not bounded within the interval $[e^{-\varepsilon'}, e^{\varepsilon'}]$ but only in the interval $[e^{-\varepsilon}, e^{\varepsilon}]$. That is, the protection due to privacy amplification from sampling is lost: T' provides the same level of protection as T when the attacker knows the unit i is in the sample.

5.1.4 The Journalist and Sampling Amplification

In this section, we show that privacy amplification by sampling is not possible when the attacker does not have a particular target unit in mind, but instead wishes to learn about an arbitrary record. In the statistical disclosure control literature, this is often referred to as the "journalist" scenario, since a journalist often wants to expose the vulnerability of a data-release mechanism by learning any record, rather than focusing on attacking a particular record (e.g. the record belonging to their neighbor). In this situation, it makes sense for the journalist to focus on a record that is in the sample, since these records have the most influence on the data-release mechanism's output. As in the previous subsection, let T be an ε-DP mechanism, $\mathcal{S}(\cdot)$ be the sampling function and $T' = T \circ \mathcal{S}$, so that T' is ε'-DP with $\varepsilon' < \varepsilon$. As is common convention, let us identify the units of $\mathfrak{S}$ as $i = 1, \ldots, n$, where $n = \left|\mathcal{U}^{\mathsf{R}}(\mathfrak{S})\right|$. Then

$$
\begin{aligned}
&\frac{\mathsf{P}_{\mathsf{A}}(\mathfrak{S}_i = r \mid T'(\mathfrak{F}^*) = t)}{\mathsf{P}_{\mathsf{A}}(\mathfrak{S}_i = r \mid T'(psample[\mathfrak{F}^*]) = t)} \\
&= \frac{\sum_{\mathfrak{s}^-} \mathsf{P}_{\mathsf{A}}(\mathfrak{s}^-)\mathsf{P}_{\mathsf{A}}(\mathfrak{S}_i = r \mid \mathfrak{s}^-)\mathsf{P}_{\mathsf{A}}(T'(\mathfrak{F}^*) = t \mid \mathfrak{S} = \mathfrak{s}^- \cup \{r\})}{\sum_{\mathfrak{s}^-} \mathsf{P}_{\mathsf{A}}(\mathfrak{s}^-)\mathsf{P}_{\mathsf{A}}(\mathfrak{S}_i = r \mid \mathfrak{s}^-) \sum_{r'} \mathsf{P}_{\mathsf{A}}(\mathfrak{S}_i = r' \mid \mathfrak{s}^-)\mathsf{P}_{\mathsf{A}}(T'(\mathfrak{F}^*) = t \mid \mathfrak{S} = \mathfrak{s}^- \cup \{r'\})} \\
&= \frac{\sum_{\mathfrak{s}^-} \mathsf{P}_{\mathsf{A}}(\mathfrak{s}^-)\mathsf{P}_{\mathsf{A}}(\mathfrak{S}_i = r \mid \mathfrak{s}^-)\mathsf{P}_{\mathfrak{s}^- \cup \{r\}}(T = t)}{\sum_{\mathfrak{s}^-} \mathsf{P}_{\mathsf{A}}(\mathfrak{s}^-)\mathsf{P}_{\mathsf{A}}(\mathfrak{S}_i = r \mid \mathfrak{s}^-) \sum_{r'} \mathsf{P}_{\mathsf{A}}(\mathfrak{S}_i = r' \mid \mathfrak{s}^-)\mathsf{P}_{\mathfrak{s}^- \cup \{r'\}}(T = t)} \\
&\geq \frac{\sum_{\mathfrak{s}^-} \mathsf{P}_{\mathsf{A}}(\mathfrak{s}^-)\mathsf{P}_{\mathsf{A}}(\mathfrak{S}_i = r \mid \mathfrak{s}^-)\mathsf{P}_{\mathfrak{s}^- \cup \{r\}}(T = t)}{\sum_{\mathfrak{s}^-} \mathsf{P}_{\mathsf{A}}(\mathfrak{s}^-)\mathsf{P}_{\mathsf{A}}(\mathfrak{S}_i = r \mid \mathfrak{s}^-) \sum_{r'} \mathsf{P}_{\mathsf{A}}(\mathfrak{S}_i = r' \mid \mathfrak{s}^-)e^{\varepsilon}\mathsf{P}_{\mathfrak{s}^- \cup \{r\}}(T = t)} \\
&= e^{-\varepsilon}.
\end{aligned}
$$

As in the previous subsection, if T is optimal then the above inequality is achieved for some choice of t, r and i. Analogous working shows that this posterior-to-posterior ratio is bounded above by e^{ε}, and moreover, this bound is achieved when T is optimal. Hence, as in the previous subsection, the additional privacy protection due to amplification from sampling is lost when the attacker targets an arbitrary record in the sample. That is, a sample record is not protected by the mechanism T' at the nominal privacy level ϵ' of T', but only at the privacy level ϵ.

Note that this result and the accompanying discussion applies more generally beyond the context of survey sampling. They holds for any DP mechanism T' which employs amplification by sampling. Such mechanisms are frequently used as modules in sanitized (i.e. privacy-protected) machine learning and neural networks as amplification by sampling is key to sanitized stochastic gradient descent algorithms [Abadi et al., 2016, Bu et al., 2020].

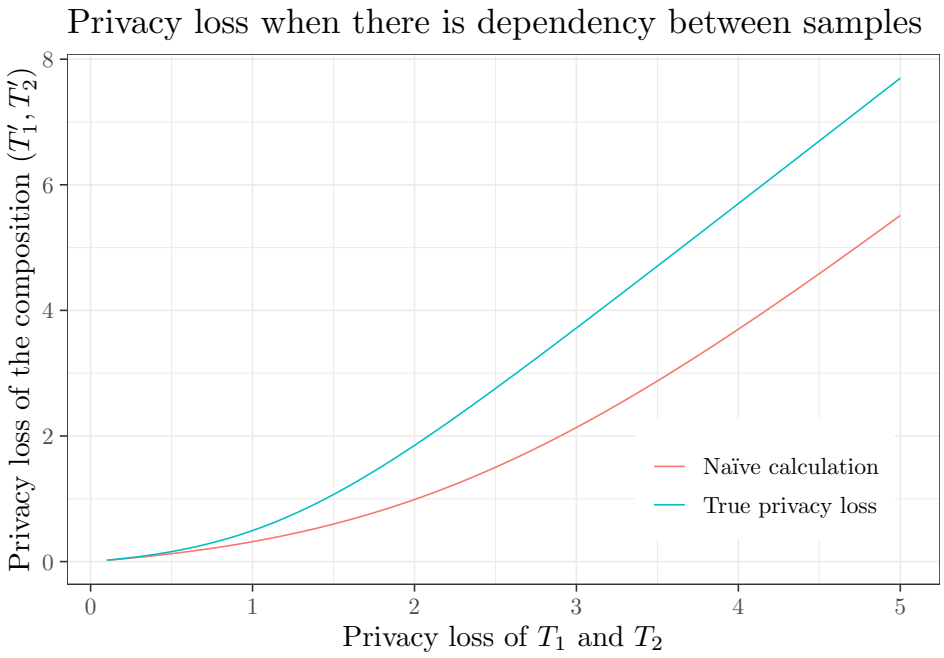

Figure 5.1: The total privacy loss over two mechanisms T_1' and T_2' which share the same sampling step. Here, $T_i' = T_i \circ S$ where S is simple random sampling without replacement (with sampling fraction $f = n/N = 0.1$). Both T_1 and T_2 satisfy ϵ-DP with privacy loss ϵ given on the x-axis. The total privacy loss of the composition of the two mechanisms T_1' and T_2' is given on the y-axis. The naïve calculation (in red) is given by the standard composition result of ϵ-DP which states that the privacy loss of (T_1', T_2') is the sum of privacy losses of T_1' and T_2'. That is, the red line is $2\log(1 + f[\exp(\mathrm{pl}(T_1)) - 1])$, where $\mathrm{pl}(T_1)$ is the privacy loss of T_1. (We assume $\mathrm{pl}(T_1) = \mathrm{pl}(T_2)$.) The true total privacy loss (in blue) is given by first composing T_1 and T_2 and then applying privacy amplification (Theorem 4.1): $\mathrm{pl}(T_1', T_2') = \log(1 + f[\exp(2\mathrm{pl}(T_1)) - 1])$.

5.2 Amplification by Sampling and Composition

An important consideration when discussing the benefits of privacy amplification from sampling is whether the composition property of DP still hold. Composition refers to the fact that the total privacy loss of two DP mechanisms with privacy loss ε_1 and ε_2, respectively is upper bounded by the sum $\varepsilon_1 + \varepsilon_2$ of the two losses. This is an important property as it helps to track the privacy loss over multiple data releases. This property is lost, however, in the context of privacy amplification through sampling as the following example illustrates: Consider two pure ε-DP mechanisms T_1 and T_2 with privacy loss $\varepsilon = 1$ and $\varepsilon = 2$ respectively. Suppose that they are two outputs from the same sample survey (i.e. they always use the same sample). For example, T_1 is the Laplace mechanism for querying the number of males in the sample and T_2 is the Laplace mechanism for querying the number of people in the sample with incomes over \$100,000. Suppose for simplicity that the sampling mechanism for the survey was simple random sampling without replacement (SRSWOR) with sampling fraction $f = n/N = 0.1$. Let T_1' and T_2' be the mechanisms which apply the sampling step and then run T_1 or T_2 respectively. These mechanisms have privacy loss 0.16 and 0.49 respectively (by amplification by sampling results given in Theorem 4.1). A naïve interpretation of the composition theorem implies that their composition (T_1', T_2') has privacy loss 0.65. However, the correct calculus would consider the composition (T_1, T_2) – which has privacy loss $\varepsilon = 3$ – and then apply the amplification by sampling result to get a privacy loss for the composition (T_1', T_2') of 1.07. We note that for small sampling rates f and small values ($\ll 1$) for both ε_1 and ε_2, the composition properties based on the amplified privacy guarantees would still hold approximately since these conditions would imply that the privacy loss of T_i' is approximately $\varepsilon_i' \approx n/N\varepsilon_i$, and thus $\varepsilon_1' + \varepsilon_2' \approx n/N\varepsilon_1 + n/N\varepsilon_2 = n/N(\varepsilon_1 + \varepsilon_2)$. However, for larger f or ε_i, the gap between the true privacy loss and the naïve calculation can be substantial, as illustrated by Figure 5.1.

The source of this apparent contradiction is the composition theorem's implicit assumption that the seeds ω_1' and ω_2' of T_1' and T_2' are independent. This assumption does not hold when T_1' and T_2' always select the same sample. More generally, suppose that T_1' and T_2' are mechanisms which include sample procedures with designs τ_1 and τ_2 respectively. Then the composition theorem's assumption is violated whenever the sample designs τ_1 and τ_2 are dependent. In such cases, the calculation of the total privacy loss across T_1' and T_2' cannot rely on applying the composition theorem to T_1' and T_2'. Instead, this calculation requires analyzing the privacy amplification of the sample designs τ_1 and τ_2 *jointly*, which will be difficult in general.

Dependency between sample designs is unfortunately a common occurrence at many NSOs. Beyond the above example where T_1' and T_2' use the same sample, there are (at least) two other common scenarios which lead to violations of the composition theorem's independence assumption. Firstly, because NSOs run many different survey collections concurrently, modern sample designs aim to

reduce respondent burden by controlling the overlap between the samples of different surveys. (For example, if a unit was selected for one survey, they will have a lower (or zero) probability of being selected in the near future for a different survey.) This introduces dependence between the sample designs of the NSO's different surveys. Secondly, sample rotation – which is a common feature in the collection of time series data, such as labor force statistics – introduces dependency between the sample designs across time for the same survey.

In all three of these scenarios, frame- (or population-)level DP mechanisms will not have independent seeds and hence the standard composition theorem does not apply to these mechanisms. This is an important consideration in determining the total privacy loss of an NSO across their multiple surveys. In situations traditionally encountered in the DP literature, the composition theorem allows for modular privacy analyses, but – without a generalized composition theorem which can account for dependency between seeds – an NSO will be forced to resort to a joint privacy analysis which must simultaneously analyze all the NSO's surveys. Therefore, an important (and novel, as far as we are aware) future research is to understand the composition property of DP under varying levels of seed dependency. Such an understanding will enable modular privacy analyses of dependent DP mechanisms to be combined into an overall privacy loss – as the standard composition theorem currently enables for independent DP mechanisms.

We conclude this subsection with the general comment that the composition of multiple mechanisms becomes more complex when these mechanisms share data-processing steps in common. Sampling is an example of one such data-processing step, but it is by no means the only example. Population-level DP mechanisms will also share the same process of frame construction (even if they use different frames, it is likely that there are dependencies between the construction of the two frames), which must be accounted for when determining the overall privacy loss.

6 Discussion

This paper develops theory for understanding and implementing differential privacy in the context of survey statistics. By recognizing the major phases in the survey-data pipeline, we identified ten different settings of DP. These settings correspond to different choices for 1) where the DP data-release mechanism starts in this pipeline; and for 2) which of the previous phases are taken as invariant. Section 3 formalized these ten settings into ten different conditions on the DP flavor.

Sections 4 and 5 show that the choice of the setting has significant impacts in terms of both privacy and utility. Therefore, while DP is invariant to post-processing, pre-processing steps matter. Moreover, the data custodian must necessarily choose a setting – they cannot implement DP without first deciding (perhaps implicitly) where the DP mechanism starts and which pre-processing steps are taken as invariant. Hence, contrary to commonly-held beliefs, DP does make important

assumptions on the data and on the attacker, because the data custodian's decision impacts both the utility and privacy semantics of the DP-outputted data.

Based on the discussions in the previous sections, we can offer some recommendations on the settings a data custodian might want to choose. Firstly, we advice against the population-level setting (i.e. using the domain $\mathcal{D}_0 = \mathcal{D}_0^{\text{pp}}$). Compared to the frame-level setting ($\mathcal{D}_0 = \mathcal{D}_0^{\text{fr}}$), the only advantage of the population-level setting would be potential amplification gains because the frame could be treated as a random subset of the pseudo-population. However, quantifying the resulting privacy amplification effects seems difficult, if not impossible, in practice. Moreover, for high quality frames the amplification effect should be small since the fraction of the pseudo-population on the frame would be high. On the other hand, using $\mathcal{D}_0^{\text{pp}}$ would always require the DP flavor to be frame-agnostic, implying that the design weights could no longer be treated as fixed. This would potentially increase the sensitivity of the output of interest and would make the computation of the sensitivity challenging in most cases.

Secondly, opting for the frame-level setting ($\mathcal{D}_0 = \mathcal{D}_0^{\text{fr}}$) offers amplification from sampling, but requires a frame agnostic DP flavor, implying that the sampling weights still cannot be treated as fixed. Since previous research has shown that amplification effects tend to be small for many complex sampling designs [Bun et al., 2022] and privacy amplification is only achievable if the nosy neighbor and the journalist scenario discussed in Sections 5.1.3 and 5.1.4 are unrealistic threat models, it seems that the benefits of amplifications are outweighted by the disadvantages of this DP setting.

Thirdly, when using one of the sample-level settings, it seems preferable to work under $\mathscr{D}_{\text{samp|fr}}$, i.e., treating the frame as invariant, as this would allow the design weights to be treated as fixed. These benefits should outweight the fact that treating the frame as invariant will increase the risks by limiting the space of neighboring datasets. These constraints on the possible values of a record $\mathfrak{s}_i$ may be small in practice, although more research is needed to verify this.[3] In principle, the sample-level setting would also offer amplification from nonresponse. However, as discussed previously, quantifying these amplification effects would require knowledge of the true response mechanism.

Finally, we do not see any benefits from starting the data release mechanism only at the responding sample. If the data custodian still prefers to choose this option, we would recommend using $\mathscr{D}_{\text{resp|fr}}$ and not $\mathscr{D}_{\text{resp|samp}}$. Our concern is that treating the target sample as fixed might enforce strong constraints on the possible values of a record $\mathfrak{r}_i$ in some circumstances. Whether one can find examples where this is really the case would be an interesting area for future research.

[3] In the final version of this paper, we will address this question in further detail.

Acknowledgements

JB gratefully acknowledges partial financial support from the Australian-American Fulbright Commission and the Kinghorn Foundation. JD gratefully acknowledges support from Cooperative Agreement CB20ADR0160001 with the Census Bureau. The views expressed in this paper are those of the authors and not those of the U.S. Census Bureau or any other sponsor.

References

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*, pages 308–318, Oct. 2016. doi: 10.1145/2976749.2978318.

J. M. Abowd and M. B. Hawes. Confidentiality protection in the 2020 us census of population and housing. *Annual Review of Statistics and Its Application*, 10:119–144, 2023.

Apple's Differential Privacy Team. Learning with privacy at scale. *Apple Machine Learning Journal*, 1(8), 2017.

J. Bailie and R. Gong. General inferential limits under differential and pufferfish privacy. http://arxiv.org/abs/2401.15491, Jan. 2024.

J. Bailie, R. Gong, and X.-L. Meng. A refreshment stirred, not shaken (I): Building blocks of differential privacy. *In preparation*, 2024+a.

J. Bailie, R. Gong, and X.-L. Meng. A statistical appreciation and assessment of differential privacy. *in preparation*, 2024+b.

B. Balle, G. Barthe, and M. Gaboardi. Privacy amplification by subsampling: Tight analyses via couplings and divergences. *Advances in neural information processing systems*, 31, 2018.

B. Balle, G. Barthe, and M. Gaboardi. Privacy profiles and amplification by subsampling. *Journal of Privacy and Confidentiality*, 10(1), Jan. 2020. ISSN 2575-8527. doi: 10.29012/jpc.726.

R. A. Bruce and J. R. McDonough. Stress testing in screening for cardiovascular disease. *Bulletin of the New York Academy of Medicine*, 45(12):1288–1305, Dec. 1969. ISSN 0028-7091. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1750554/.

Z. Bu, J. Dong, Q. Long, and S. Weijie. Deep learning with Gaussian differential privacy. *Harvard Data Science Review*, 2(3), July 2020. ISSN 2644-2353, 2688-8513. doi: 10.1162/99608f92.cfc5dd25.

M. Bun, J. Drechsler, M. Gaboardi, A. McMillan, and J. Sarathy. Controlling privacy loss in sampling schemes: An analysis of stratified and cluster sampling. In *Foundations of Responsible Computing (FORC 2022)*, page 24, June 2022.

P. Cuff and L. Yu. Differential privacy as a mutual information constraint. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, pages 43–54, 2016.

S. Das, J. Drechsler, K. Merrill, and S. Merrill. Imputation under differential privacy. http://arxiv.org/abs/2206.15063, July 2022.

B. Ding, J. Kulkarni, and S. Yekhanin. Collecting telemetry data privately. In *Advances in Neural Information Processing Systems*, pages 3571–3580, 2017.

J. Drechsler. Differential privacy for government agencies—Are we there yet? *Journal of the American Statistical Association*, 118(541):761–773, Jan. 2023. ISSN 0162-1459. doi: 10.1080/01621459.2022.2161385. https://doi.org/10.1080/01621459.2022.2161385.

C. Dwork and A. Roth. The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science*, 9(3–4):211–407, 2014.

C. Dwork, K. Kenthapadi, F. McSherry, I. Mironov, and M. Naor. Our data, ourselves: Privacy via distributed noise generation. In S. Vaudenay, editor, *Advances in Cryptology - EUROCRYPT 2006*, Lecture Notes in Computer Science, pages 486–503, Berlin, Heidelberg, 2006a. Springer. ISBN 978-3-540-34547-3. doi: 10.1007/11761679_29.

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

Ú. Erlingsson, V. Pihur, and A. 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.

A. D. Foote, A. Machanavajjhala, and K. McKinney. Releasing earnings distributions using differential privacy: Disclosure avoidance system for post-secondary employment outcomes (pseo). *Journal of Privacy and Confidentiality*, 9(2), 2019. doi: 10.29012/jpc.722. URL https://journalprivacyconfidentiality.org/index.php/jpc/article/view/722.

D. G. Horvitz and D. J. Thompson. A generalization of sampling without replacement from a finite universe. *Journal of the American statistical Association*, 47(260):663–685, 1952.

D. Kifer and A. Machanavajjhala. No free lunch in data privacy. In *Proceedings of the 2011 International Conference on Management of Data - SIGMOD '11*, pages 193–204, Athens, Greece, 2011. ACM Press. ISBN 978-1-4503-0661-4. doi: 10.1145/1989323.1989345.

D. Kifer, J. M. Abowd, R. Ashmead, R. Cumings-Menon, P. Leclerc, A. Machanavajjhala, W. Sexton, and P. Zhuravlev. Bayesian and frequentist semantics for common variations of differential privacy: Applications to the 2020 Census. Technical Report arXiv:2209.03310, Sept. 2022.

S. Lin, M. Bun, M. Gaboardi, E. D. Kolaczyk, and A. Smith. Differentially private confidence intervals for proportions under stratified random sampling. *arXiv preprint arXiv:2301.08324*, 2023.

A. Machanavajjhala, D. Kifer, J. Abowd, J. Gehrke, and L. Vilhuber. Privacy: Theory meets practice on the map. In *Proceedings of the 2008 IEEE 24th International Conference on Data Engineering*, pages 277–286. IEEE Computer Society, 2008.

S. Messing, C. DeGregorio, B. Hillenbrand, G. King, S. Mahanti, Z. Mukerjee, C. Nayak, N. Persily, B. State, and A. Wilkins. Facebook Privacy-Protected Full URLs Data Set. https://doi.org/10.7910/DVN/TDOAPG, 2020.

P. J. Phillips. Oral glucose tolerance testing. *Australian Family Physician*, 41(6): 391–393, June 2012. ISSN 0300-8495. https://www.racgp.org.au/afp/2012/june/oral-glucose-tolerance-testing.

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

D. B. Rubin. Causal inference using potential outcomes: Design, modeling, decisions. *Journal of the American Statistical Association*, 100(469):322–331, Mar. 2005. ISSN 0162-1459, 1537-274X. doi: 10.1198/016214504000001880.

M. C. Tschantz, S. Sen, and A. Datta. SoK: Differential privacy as a causal property. In *2020 IEEE Symposium on Security and Privacy (SP)*, pages 354–371. IEEE, 2020.

Uber Security. Uber releases open source project for differential privacy. https://medium.com/uber-security-privacy/differential-privacy-open-source-7892c82c42b6, 2017.

U.S. Bureau of Labor Statistics. Current Population Survey: Concepts. https://www.bls.gov/opub/hom/cps/concepts.htm, Apr. 2018.

US Census Bureau. Protecting the confidentiality of America's statistics: Adopting modern disclosure avoidance methods at the Census Bureau. https://www.census.gov/newsroom/blogs/research-matters/2018/08/protecting_the_confi.html, Aug. 2018.

U.S. Census Bureau. Design and methodology: Current Population Survey. *Technical Paper 77*, 2019.

US Census Bureau. Disclosure avoidance protections for the American Community Survey. https://www.census.gov/newsroom/blogs/random-samplings/2022/12/disclosure-avoidance-protections-acs.html, Dec. 2022.

U.S. Census Bureau. List of Surveys. https://www.census.gov/programs-surveys/surveyhelp/list-of-surveys.html, 2023. Accessed: 12/12/2023.

R. Valliant, J. A. Dever, and F. Kreuter. *Practical tools for designing and weighting survey samples*. Springer, 2 edition, 2018.

A Proofs

Proof of Lemma 2.7. Let $T \in \mathcal{M}(\mathcal{X}, \mathscr{D}', d_{\mathcal{D}_0}, d_{\mathrm{Pr}}, \varepsilon_{\mathcal{D}'})$. Take some $\mathcal{D} \in \mathscr{D}$ and some $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{D}$. Then

$$\begin{aligned} d_{\mathrm{Pr}}(\mathsf{P}_{\mathfrak{d}}, \mathsf{P}_{\mathfrak{d}'}) &\leq \inf\{\varepsilon_{\mathcal{D}'} d_{\mathcal{D}_0}(\boldsymbol{x}, \boldsymbol{x}') : \mathcal{D}' \in \mathscr{D}' \text{ s.t. } \boldsymbol{x}, \boldsymbol{x}' \in \mathcal{D}'\} \\ &\leq \inf\{\varepsilon_{\mathcal{D}'} d_{\mathcal{D}_0}(\boldsymbol{x}, \boldsymbol{x}') : \mathcal{D}' \in \mathscr{D}' \text{ s.t. } \mathcal{D} \subset \mathcal{D}'\} \\ &= \varepsilon_{\mathcal{D}} d_{\mathcal{D}_0}(\boldsymbol{x}, \boldsymbol{x}'), \end{aligned}$$

where

$$\varepsilon_{\mathcal{D}} = \inf\{\varepsilon_{\mathcal{D}'} : \mathcal{D}' \in \mathscr{D}' \text{ s.t. } \mathcal{D} \subset \mathcal{D}'\}.$$

□

Proof of Lemma 2.11. Let $\mathcal{D}' \in \overline{\mathscr{D}}$. Then there exists some $\mathcal{D} \in \mathscr{D}$ and $\boldsymbol{x} \in \mathcal{D}$ such that $\mathcal{D}' = \mathcal{D} \cap [\boldsymbol{x}]$. Since every $\boldsymbol{x}', \boldsymbol{x}'' \in [\boldsymbol{x}]$ are connected, it follows that every $\boldsymbol{x}', \boldsymbol{x}'' \in \mathcal{D}'$ are also connected. This proves that $\overline{\mathscr{D}}$ is complete.

Suppose that $T \in \mathcal{M}(\mathcal{X}, \overline{\mathscr{D}}, d_{\mathcal{D}_0}, d_{\mathrm{Pr}}, \varepsilon_{\mathcal{D}'})$. Take some $\mathcal{D} \in \mathscr{D}$ and some $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{D}$. We wish to show that

$$d_{\mathrm{Pr}}(\mathsf{P}_{\mathfrak{d}}, \mathsf{P}_{\mathfrak{d}'}) \leq \varepsilon_{\mathcal{D}} d_{\mathcal{D}_0}(\boldsymbol{x}, \boldsymbol{x}'). \tag{A.1}$$

We may assume without loss of generality that $d_{\mathcal{D}_0}(\boldsymbol{x}, \boldsymbol{x}') < \infty$. Define $\mathcal{D}' = \mathcal{D} \cap [\boldsymbol{x}]$. Since $\mathcal{D}' \in \overline{\mathscr{D}}$ and $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{D}'$, we know that

$$d_{\mathrm{Pr}}(\mathsf{P}_{\mathfrak{d}}, \mathsf{P}_{\mathfrak{d}'}) \leq \varepsilon_{\mathcal{D}'} d_{\mathcal{D}_0}(\boldsymbol{x}, \boldsymbol{x}').$$

(A.1) then follows by observing that $\varepsilon_{\mathcal{D}'} \leq \varepsilon_{\mathcal{D}}$.

Suppose that $T \in \mathcal{M}(\mathcal{D}_0, \mathscr{D}, d_{\mathcal{D}_0}, d_{\mathrm{Pr}}, \varepsilon_{\mathcal{D}})$. Take some $\mathcal{D}' \in \overline{\mathscr{D}}$ and some $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{D}'$. Then

$$\begin{aligned} d_{\mathrm{Pr}}(\mathsf{P}_{\mathfrak{d}}, \mathsf{P}_{\mathfrak{d}'}) &\leq \inf\{\varepsilon_{\mathcal{D}} d_{\mathcal{D}_0}(\boldsymbol{x}, \boldsymbol{x}') : \mathcal{D} \in \mathscr{D} \text{ s.t. } \boldsymbol{x}, \boldsymbol{x}' \in \mathcal{D}\} \\ &\leq \inf\{\varepsilon_{\mathcal{D}} d_{\mathcal{D}_0}(\boldsymbol{x}, \boldsymbol{x}') : \mathcal{D} \in \mathscr{D} \text{ s.t. } \mathcal{D}' \subset \mathcal{D}\} \\ &= \varepsilon_{\mathcal{D}'} d_{\mathcal{D}_0}(\boldsymbol{x}, \boldsymbol{x}'). \end{aligned}$$

□

Proof of Theorem 4.3. Let $\mathcal{D} \in \mathscr{D}$ and $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{D}$. The density of $\mathsf{P}_{\mathfrak{d}}(T \in \cdot)$ is

$$f_{\boldsymbol{x}}(t) = (2\Delta_q([\boldsymbol{x}]_{\mathscr{D}}))^{-k} \exp\left(-\frac{\|t - q(\boldsymbol{x})\|_1}{\Delta_q([\boldsymbol{x}]_{\mathscr{D}})}\right).$$

Thus,

$$\begin{aligned} d_{\mathrm{MULT}}(\mathsf{P}_{\mathfrak{d}}, \mathsf{P}_{\mathfrak{d}'}) &= \sup_{t \in \mathbb{R}} \left| \ln \frac{f_{\boldsymbol{x}}(t)}{f_{\boldsymbol{x}'}(t)} \right| \\ &= \sup_{t \in \mathbb{R}} \left| \frac{\|t - q(\boldsymbol{x}')\|_1 - \|t - q(\boldsymbol{x})\|_1}{\Delta_q([\boldsymbol{x}]_{\mathscr{D}})} \right| \\ &\leq \varepsilon d_{\mathcal{D}_0}(\boldsymbol{x}, \boldsymbol{x}'), \end{aligned}$$

where the first line follows by Proposition 38 of Bailie and Gong [2024], the second because $[\boldsymbol{x}]_{\mathscr{D}} = [\boldsymbol{x}']_{\mathscr{D}}$ and the third by the reverse triangle inequality. □