Mechanisms for Global Differential Privacy
under Bayesian Data Synthesis

Jingchen Hu, Matthew R. Williams and Terrance D. Savitsky

Vassar College, RTI International and U.S. Bureau of Labor Statistics

Abstract: This paper introduces a new method that embeds any Bayesian model used to generate synthetic data and converts it into a differentially private (DP) mechanism. We propose an alteration of the model synthesizer to utilize a censored likelihood that induces upper and lower bounds of $[\exp(-\epsilon/2), \exp(\epsilon/2)]$, where ϵ denotes the level of the DP guarantee. This censoring mechanism equipped with an $\epsilon-$DP guarantee will induce distortion into the joint parameter posterior distribution by flattening or shifting the distribution towards a weakly informative prior. To minimize the distortion in the posterior distribution induced by likelihood censoring, we embed a vector-weighted pseudo posterior mechanism within the censoring mechanism. The pseudo posterior is formulated by selectively downweighting each likelihood contribution proportionally to its disclosure risk. On its own, the pseudo posterior mechanism produces a weaker asymptotic differential privacy (aDP) guarantee. After embedding in the censoring mechanism, the DP guarantee becomes strict such that it does not rely on asymptotics. We demonstrate that the pseudo posterior mechanism creates synthetic data with the highest utility at the price of a weaker, aDP guarantee, while embedding the pseudo posterior mechanism in the proposed censoring mechanism produces synthetic data with a stronger, non-asymptotic DP guarantee at the cost of slightly reduced utility. The

1

perturbed histogram mechanism is included for comparison.

Key words and phrases: asymptotic differential privacy, Bayesian data synthesis, censoring, differential privacy, pseudo posterior mechanism, synthetic data

1. Introduction

This paper focuses on constructing a synthetic data mechanism equipped with a formal privacy guarantee that allows the use of any Bayesian probability model as the data synthesizer. A formal privacy guarantee is quantifiable and attaches to a data generating mechanism independent of the behaviors of putative intruders seeking to re-identify data records; in particular, we focus on the differential privacy (DP) framework (Dwork et al., 2006) to provide a formal privacy guarantee.

Definition 1 (Differential Privacy). Let $D \in \mathbb{R}^{n \times k}$ be a database in input space $\mathcal{D}$. Let $\mathcal{M}$ be a randomized mechanism such that $\mathcal{M} : \mathbb{R}^{n \times k} \to O$. Then $\mathcal{M}$ is ϵ-differentially private if

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

for all possible outputs $O = Range(\mathcal{M})$ under all possible pairs of datasets $D, D' \in \mathcal{D}$ of the same size which differ by only a single row (i.e., Hamming-1 distance).

DP assigns a disclosure risk for a statistic to be released to the public, $f(D)$ (e.g., total employment for a state-industry) of any $D \in \mathcal{D}$ based on the global

2

sensitivity, $\Delta_G = \sup_{D,D' \in \mathcal{D}:\ \delta(D,D')=1} |f(D) - f(D')|$, over the space of databases, $\mathcal{D}$, where $\delta(D, D') = 1$ denotes the Hamming-1 distance such that D differs from D' by a single record. If the value of the statistic, f, expresses a high magnitude change after changes or the removal of a data record in D' (i.e., large Δ_G value), then the mechanism will be required to induce a relatively higher level of distortion to f.

A commonly-used data privacy approach generates synthetic microdata from statistical models estimated on confidential data for proposed release by data disseminators (Rubin, 1993; Little, 1993). We denote such a release mechanism used to generate synthetic data, $\xi(\theta \mid \mathbf{x})$, where $\mathbf{x}$ are the confidential data and θ denotes the parameters used to generate the synthetic data. In particular, we consider the case where $\xi(\theta \mid \mathbf{x})$ is a posterior distribution of a Bayesian hierarchical probability model.

The Exponential Mechanism by McSherry and Talwar (2007) is a popular approach to generating draws of parameters, θ, and associated synthetic data which is differentially private.

Definition 2 (Exponential Mechanism)**.** The Exponential Mechanism releases values of θ from a distribution proportional to

$$\exp\left(\frac{\epsilon u(\mathbf{x}, \theta)}{2\Delta_u}\right) \xi(\theta), \tag{1.1}$$

3

where $u(\mathbf{x}, \theta)$ is a utility function and $\xi(\theta)$ is a base or prior distribution. Let

$$\Delta_u = \sup_{\mathbf{x} \in \mathcal{X}^n} \sup_{\mathbf{x}, \mathbf{y} : \delta(\mathbf{x}, \mathbf{y}) = 1} \sup_{\theta \in \Theta} |u(\mathbf{x}, \theta) - u(\mathbf{y}, \theta)| \qquad (1.2)$$

denote the sensitivity of the utility function, $u(\mathbf{x}, \theta)$, defined globally over $\mathbf{x} = (x_1, \ldots, x_n) \in \mathcal{X}^n$, the $\sigma-$algebra of datasets, $\mathbf{x}$, governed by product measure, P_{θ_0}; $\delta(\mathbf{x}, \mathbf{y}) = \#\{i : x_i \neq y_i\}$ is the Hamming distance between $\mathbf{x}, \mathbf{y} \in \mathcal{X}^n$. Then each draw of θ from the Exponential Mechanism provides an $\epsilon-$DP privacy guarantee, where $\epsilon \leq 2\Delta_u$.

For a Bayesian model utilizing the data log-likelihood as the utility function of the exponential mechanism, Savitsky et al. (2022) demonstrate the exponential mechanism specializes to the model posterior distribution, which provides a straight-forward mechanism from which to draw samples. Dimitrakakis et al. (2017) define a model-based sensitivity, $\sup_{\mathbf{x}, \mathbf{y} \in \mathcal{X}^n : \delta(\mathbf{x}, \mathbf{y}) = 1} \sup_{\theta \in \Theta} |f_\theta(\mathbf{x}) - f_\theta(\mathbf{y})| \leq \Delta$, which is constructed as a Lipschitz bound. They demonstrate a connection between the Lipschitz bound, Δ and $\epsilon \leq 2\Delta$ for each draw of parameters, θ, where $f_\theta(\mathbf{x})$ is the model log-likelihood. The guarantee applies to all databases $\mathbf{x}$, in the space of databases of size n, $\mathcal{X}^n$, and denoted as the posterior mechanism.

Computing a finite $\Delta < \infty$ in practice, however, as acknowledged by Dimitrakakis et al. (2017), is difficult-to-impossible for an unbounded parameter space (e.g., a normal distribution) under simple models. A truncation of the parameter

4

space is required to achieve a finite Δ and the truncation only works for some models to achieve a finite Δ. Even more, parameter truncation becomes intractable analytically or computationally or both for practical models that utilize a multidimensional parameter space. For example, a non-linear mixed effects model could have the number of model parameters on the order of the number of data records, which can be millions, in practice.

The pseudo posterior mechanism of Savitsky et al. (2022) reviewed in Section 2.1 allows achievement of a target Lipschitz value that, in turn, determines a formal privacy guarantee by using likelihood weights, $\alpha_i \in [0, 1]$, to selectively downweight high-risk records. This mechanism achieves a DP guarantee *asymptotically* as the parameter space contracts onto a point. In other words, while the mechanism is flexible, the privacy guarantee is not strict.

This paper introduces a new mechanism that embeds the pseudo posterior mechanism by *censoring* the pseudo likelihood values above or below a threshold to targeted Lipschitz (and hence privacy) guarantee, which allows achievement of a target $\epsilon-$DP guarantee without reliance on asymptotics. Any Bayesian data synthesizing model may be utilized under the censoring mechanism.

The remainder of the paper is organized as the following. Section 2 reviews the pseudo posterior mechanism, followed by a review of the perturbed histogram mechanism that will be used for comparison. The new censoring mechanism is introduced

in Section 3. Section 4 focuses on a series of simulation studies where the mechanisms are applied and their data utilities are compared at equivalent privacy guarantees. Section 5 presents an application to a sample of the Survey of Doctoral Recipients. We end the paper with a few concluding remarks in Section 6.

2. Synthetic Microdata Generating Privacy Mechanisms

In this section, we review the pseudo posterior mechanism proposed by Savitsky et al. (2022) in Section 2.1 and discuss a variant of it in Section 2.2. The perturbed histogram mechanism (Dwork et al., 2006; Wasserman and Zhou, 2010), is briefly reviewed in Section 2.3.

2.1 Pseudo Posterior Mechanism

To guarantee the achievement of a finite $\Delta < \infty$ for any synthesizing model over an unbounded parameter space, Savitsky et al. (2022) propose a *pseudo* posterior mechanism that uses a log-pseudo likelihood with a vector of observation-indexed weights $\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_n) \in [0,1]^n$ where each α_i exponentiates the likelihood contribution, $p(x_i \mid \theta)$, for each record $i \in (1, \ldots, n)$. Each weight, $\alpha_i \in [0,1]$ is set to be inversely proportional to a measure of disclosure risk for record, i, such that the model used to generate synthetic data will be less influenced by relatively high-risk records.

6

The pseudo posterior mechanism of Savitsky et al. (2022) is formulated as

$$\xi^{\boldsymbol{\alpha}(\mathbf{x})}(\theta \mid \mathbf{x}) \propto \prod_{i=1}^{n} p(x_i \mid \theta)^{\alpha_i} \times \xi(\theta), \tag{2.3}$$

where each $\alpha_i \in [0,1]$ serves to downweight the likelihood contribution of record i with $\alpha_i \propto 1/\sup_{\theta \in \Theta}|f_\theta(x_i)|$, and $|f_\theta(x_i)| = |\log p(x_i \mid \theta)|$ is the absolute log-likelihood of record i. The differential downweighting of individual records intends to better preserve utility by focusing the downweighting on high-risk records. High-risk records tend to be those located in the tails of the distribution where the absolute log-likelihood, $|f_\theta(x_i)|$, is highest. The differential downweighting allows the preservation of the high mass portions of the data distribution in the generated synthetic data. The pseudo posterior mechanism sets $\alpha_i = 0$ for any record with a *non-finite* log-likelihood, which ensures a finite $\Delta_{\boldsymbol{\alpha}} = \sup_{\mathbf{x},\mathbf{y} \in \mathcal{X}^n : \delta(\mathbf{x},\mathbf{y})=1} \sup_{\theta \in \Theta} |\alpha(\mathbf{x}) \times f_\theta(\mathbf{x}) - \alpha(\mathbf{y}) \times f_\theta(\mathbf{y})| < \infty$, where $\alpha(\mathbf{x}) = [\alpha_1(\mathbf{x}), \ldots, \alpha_n(\mathbf{x})]$, $f_\theta(\mathbf{x}) = [f_\theta(x_1), \ldots, f_\theta(x_n)]$ and $\alpha(\mathbf{x}) \times f_\theta(\mathbf{x}) = \sum_{i=1}^{n} \alpha_i(\mathbf{x}) f_\theta(x_i)$ that downweights the associated log-likelihood contributions, $f_\theta(x_i)$ with multiplication by $\alpha_i(\mathbf{x}) \leq 1$. We see that $\Delta_{\boldsymbol{\alpha}} \leq \Delta$ since $\alpha_i(\mathbf{x}) = \alpha_i \leq 1$. The $\boldsymbol{\alpha}-$weighted pseudo synthesizer, $\xi^{\boldsymbol{\alpha}(\mathbf{x})}(\theta \mid \mathbf{x})$ satisfies $\epsilon-$DP if the following inequality holds.

Definition 3 (Differential Privacy under the Pseudo Posterior Mechanism).

$$\sup_{\mathbf{x},\mathbf{y} \in \mathcal{X}^n : \delta(\mathbf{x},\mathbf{y})=1} \sup_{B \in \beta_\Theta} \frac{\xi^{\boldsymbol{\alpha}(\mathbf{x})}(B \mid \mathbf{x})}{\xi^{\boldsymbol{\alpha}(\mathbf{y})}(B \mid \mathbf{y})} \leq e^\epsilon, \tag{2.4}$$

where $\xi^{\boldsymbol{\alpha}(\mathbf{x})}(B \mid \mathbf{x}) = \int_{\theta \in B} \xi^{\boldsymbol{\alpha}(\mathbf{x})}(\theta \mid \mathbf{x}) d\theta$.

Definition 3 limits the change in the pseudo posterior distribution over all sets, $B \in \beta_\Theta$ (i.e., β_Θ is the $\sigma-$algebra of measurable sets on Θ), from the inclusion of a single record. Although the pseudo posterior distribution mass assigned to B depends on $\mathbf{x}$, the ϵ guarantee is defined as the supremum over all $\mathbf{x} \in \mathcal{X}^n$ and for all $\mathbf{y} \in \mathcal{X}^n$ which differ by one record (i.e., $\delta(\mathbf{x}, \mathbf{y}) = 1$).

Let $\Delta_{\boldsymbol{\alpha},\mathbf{x}} = \sup_{\delta(\mathbf{x},\mathbf{y})=1} \sup_{\theta \in \Theta} |\alpha(\mathbf{x}) \times f_\theta(\mathbf{x}) - \alpha(\mathbf{y}) \times f_\theta(\mathbf{y})|$ be the Lipschitz bound computed, *locally*, on database $\mathbf{x}$, over all databases, $\mathbf{y}$, at a Hamming-1 distance from $\mathbf{x}$. (We note that the term local in the DP literature can refer to the trust model as in central versus local or the sensitivity calculation as in global versus local. Our use here, however, differs from these two and focus on being local to the database $\mathbf{x}$.) The pseudo posterior mechanism *indirectly* sets the local DP guarantee, $\epsilon_{\mathbf{x}} = 2\Delta_{\boldsymbol{\alpha},\mathbf{x}}$, through the computation of the likelihood weights, $\boldsymbol{\alpha}$. The $\boldsymbol{\alpha}$ may be further scaled and shifted; for example, tuning settings (c_1, c_2) can be used to construct $\tilde{\alpha}_i = c_1 \times \alpha_i + c_2$ and decrease or increase the $\tilde{\boldsymbol{\alpha}}$ weights to achieve a target $\Delta_{\tilde{\boldsymbol{\alpha}},\mathbf{x}}$, which in turn delivers a target $\epsilon_{\mathbf{x}} = 2\Delta_{\tilde{\boldsymbol{\alpha}},\mathbf{x}}$ (Savitsky et al., 2022; Hu et al., 2022).

Savitsky et al. (2022) show that the local $\Delta_{\boldsymbol{\alpha},\mathbf{x}}$ estimated on a confidential database, $\mathbf{x}$, contracts onto the global $\Delta_{\boldsymbol{\alpha}}$ over the space of databases of size n asymptotically in sample size, n; in particular, $\Delta_{\boldsymbol{\alpha},\mathbf{x}}$ becomes arbitrarily close to the global $\Delta_{\boldsymbol{\alpha}}$ such that the pseudo posterior mechanism is asymptotically differentially

private (aDP).

The process of creating synthetic data under the $\boldsymbol{\alpha}-$weighted pseudo posterior mechanism works as follows:

1. For a confidential database $\mathbf{x}$ of n records, one fits a Bayesian model parameterized by θ.

2. Based on the estimated model and the log-likelihood of each record, one calculates a vector of observation-indexed weights $\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_n) \in [0,1]^n$ where $\alpha_i \propto 1/|f_\theta(x_i)|$.

3. Re-estimate the model under the pseudo posterior in Equation (2.3). The local Lipschitz bound, $\Delta_{\boldsymbol{\alpha},\mathbf{x}}$, is computed based on the estimated θ^* under the pseudo posterior. These data satisfy a *local* ($\epsilon_{\mathbf{x}} = 2 \times \Delta_{\boldsymbol{\alpha},\mathbf{x}}$)$-$DP guarantee that contracts on an $\epsilon = \epsilon_{\mathbf{x}}$ global aDP guarantee for n sufficiently large.

4. As a post-processing step, a synthetic dataset $\mathbf{x}^*$ is generated from the pseudo posterior predictive distribution under estimated parameters θ^*.

We label the $\boldsymbol{\alpha}-$weighted pseudo posterior mechanism as "Weighted" in the simulations and application in Sections 4 and 5, as contrasted with an "Unweighted" method that sets all of the $\alpha_i = 1$ to produce a posterior mechanism.

2.2   Weighted-e Pseudo Posterior Mechanism

In addition to the pseudo posterior mechanism with observation-indexed weight α_i, Savitsky et al. (2022) introduce a truncation of each weight, α_i: if its log-likelihood contribution, $\alpha_i \times f_\theta(x_i) > \epsilon/2$, we set final weight, $\alpha_i^* = 0$.

Setting a likelihood threshold, $\epsilon/2$, reduces the constant of proportionality in the $\mathcal{O}(n^{-1/2})$ contraction of the local Lipschitz bounds to speed convergence onto $\Delta_{\boldsymbol{\alpha}} = \epsilon/2$. This adjustment to the pseudo posterior mechanism induces a rapid contraction of the $\epsilon_{\mathbf{x}}$ computed on the observed (local) database to the global ϵ providing the aDP guarantee. We include this tweak to the pseudo posterior mechanism because it provides an alternative to the censoring mechanism introduced in the next section for achieving a stronger privacy guarantee.

The synthetic data generation process works in the same way with the $\boldsymbol{\alpha}-$weighted pseudo posterior mechanism in Section 2.1, excepts for their slight differences in how the weights are calculated. We label this slightly revised version of the pseudo posterior mechanism as "Weighted-e" in the sequel.

2.3   Perturbed Histogram Mechanism

We next present the commonly-used perturbed histogram mechanism for simulating synthetic microdata that achieves $\epsilon-$DP guarantee (Dwork et al., 2006; Wasserman and Zhou, 2010) as a comparison. For brevity, we include a detailed review in the

Supplementary Materials.

Under the required strong assumption of a *bounded* and continuous variable, one first discretizes it into a histogram with a selected number of bins. One next induces a formal $\epsilon-$DP privacy guarantee into the histogram by adding Laplace noise. Finally, one simulates microdata from the private histogram under $\epsilon-$DP, which is a post-processing step in a similar fashion as generating synthetic data under the pseudo posterior mechanism (given the privacy protected parameter draws) reviewed in Section 2.1. We label the perturbed histogram synthesizer as "PH" in the sequel.

3. Censored (Likelihood) Mechanism

The pseudo posterior mechanism discussed in Section 2.1 solves the important problem of how to turn *any* Bayesian probability model into a data synthesizing mechanism equipped with a formal privacy guarantee. A substantial limitation, however, is that the formal privacy guarantee is asymptotic, not strict. While an asymptotic guarantee may be favorably viewed as "relaxed" in the sense that better utility may be achieved than under a strict guarantee, one may never know with certainty when their database size, n, is sufficiently large such that the asymptotic privacy guarantee de facto becomes strict (or global over the space of databases).

Our new mechanism censors the log-likelihood contribution in a manner that carries over the property of the pseudo posterior mechanism of Savitsky et al. (2022)

that allows any Bayesian probability model to be purposed as a formally private mechanism for generating synthetic data, but is now equipped with a strict $\epsilon-$DP guarantee. It is common practice to construct a Bayesian hierarchical probability model to closely fit the data for the purpose of removing measurement error as is done in small area estimation models estimated on area-indexed survey statistics. Our newly proposed censoring mechanism makes it straightforward to embed this type of closely-fitting model and yet output synthetic data with a strict $\epsilon-$DP guarantee.

The idea of censoring contributions to log-likelihoods has also appeared in Pena and Barrientos (2021) and Cannone et al. (2019) in the DP literature. Both papers are focused on the narrower task of conducting differentially private simple hypothesis testing. Pena and Barrientos (2021) censors a Bayes Factor used to conduct differentially private hypothesis testing between two model alternatives. While they use fully Bayesian model specifications, simple conjugate frameworks are used such that the parameters may be analytically marginalized to compute the Bayes factor. Their example is a nested linear regression under a mixture of g-priors for the regression coefficient. Cannone et al. (2019) performs differentially private hypothesis testing to differentiate distributions P and Q by censoring or "clamping" the log-likelihood ratio.

Our new censoring mechanism, by contrast, outputs differentially private model parameters that facilitate synthetic data generation. Such data may, in turn, be

12

used for many purposes, including hypothesis testing between any number of model parameterization alternatives. Even more, our mechanism may be used to privatize complex hierarchical models where parameter posterior distributions are not analytically marginalized. We note that the DP methods from Pena and Barrientos (2021) and Cannone et al. (2019) are tailored for the specific task of conducting differentially private simple hypothesis testing. Compared to our goal of generating differentially private synthetic data with more flexibility in reusing the outputs for multiple tasks, these specific task oriented DP mechanisms may have the advantage of offering improved utility of the targeted analyses at the expense of less flexibility.

We demonstrate in the sequel that our new censoring mechanism may even embed the pseudo posterior mechanism with the result that one gets the best properties of both mechanisms. The distribution properties of the confidential data are well-preserved in the resulting synthetic data due to the surgical targeting of only high-risk record for likelihood downweighting. Yet, now the resulting synthetic data are equipped with a strict $\epsilon-$DP privacy guarantee.

We begin by laying out the details of how to construct the new censoring mechanism achieving a strict $\epsilon-$DP guarantee as an alternative to relying on the asymptotic contraction of $\theta \in \Theta$ (at θ^*) to achieve an aDP global privacy guarantee for a sufficiently large sample size, n. Our new mechanism censors the log-likelihood at a target threshold, $\epsilon/2$, representing the targeted DP guarantee. The use of censoring intends

to "lock in" any chosen ϵ under an $\epsilon-$DP formal privacy guarantee without relying on asymptotics. The ϵ guarantee is strict. The censored threshold will determine the Lipschitz bound, $\Delta = \epsilon/2$, which will indirectly set the $\epsilon = 2\Delta$ guarantee.

The definition of DP anchors its guarantee to all databases of fixed size n. By contrast, the privacy guarantee offered by the new censoring mechanism is stronger in that the same ϵ guarantee applies to databases of any size, including those $> n$.

We introduce the censoring mechanism by embedding the $\boldsymbol{\alpha}-$weighted pseudo likelihood to construct the following new pseudo likelihood with,

$$p_c^{\boldsymbol{\alpha}}(x_i \mid \theta) = \begin{cases} \exp(\epsilon/2), & p(x_i \mid \theta)^{\alpha_i} > \exp(\epsilon/2), \\ \exp(-\epsilon/2), & p(x_i \mid \theta)^{\alpha_i} < \exp(-\epsilon/2), \\ p(x_i \mid \theta)^{\boldsymbol{\alpha}}, & \text{otherwise}, \end{cases} \tag{3.5}$$

for use in

$$\xi_c^{\boldsymbol{\alpha}}(\theta \mid \mathbf{x}) \propto \prod_{i=1}^{n} p_c^{\boldsymbol{\alpha}}(x_i \mid \theta)\xi(\theta), \tag{3.6}$$

where the subscript c in $p_c^{\boldsymbol{\alpha}}(x_i \mid \theta)$ and $\xi_c^{\boldsymbol{\alpha}}(\theta \mid \mathbf{x})$ stands for the censoring mechanism.

Censoring offers a practical, low-dimensional alternative to truncating the parameter space (as proposed by Dimitrakakis et al. (2017)) to achieve an $\epsilon-$DP guarantee, as truncating the parameter space Θ quickly becomes impractical as the number of parameters grows.

One may imagine a small likelihood value assigned to an extreme observation in the tails. The identity of such an observation is at a relatively high risk of discovery

14

since its value is isolated from those of other records. As a result, the absolute value

of its log-likelihood, $|f_\theta(x_i)| = |\log p(x_i \mid \theta)|$ that determines the Δ and $\epsilon = 2\Delta$, will

be relatively large. Under the censoring mechanism, the large absolute log-likelihood

value will be truncated (in the way shown in Equation (3.5)), which serves to cap

the Lipschitz bound. As compared to a non-censored data distribution over the

parameter space, the censored distribution for each x_i will be relatively flatter or

tempered, removing local features that make detection relatively easier.

The two panels in Figure 1 illustrate the effect of censoring on the resulting

parameter posterior distributions. In this illustration, a right-skewed confidential

dataset with a long tail of 2000 observations is generated from a Beta$(0.5, 2)$ distri-

bution. Our proposed censored mechanism is implemented at $\epsilon = 1.6$. The right

panel of Figure 1 centers each density to compare the relative spread of the prior,

posterior and censored posterior distributions. Compared to the posterior distribu-

tion of a parameter for a given record, the censored distribution is relatively flatter or

tempered. This stems from assigning relatively higher (non-absolute) log-likelihood

values to parameter values in the tails of the distribution. The left panel of Fig-

ure 1 shows that censoring distorts the parameter estimation as compared to the

non-censored posterior. These two panels together shed some light on the utility-

risk trade-off offered by the censoring mechanism: To achieve strict $\epsilon-$DP guarantee

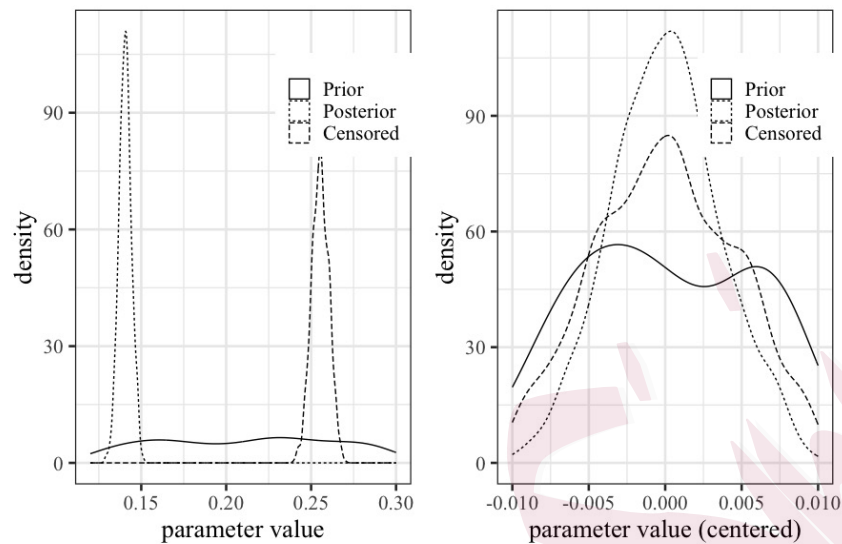censoring induces the distortion into the censored parameter posterior distribution,

Figure 1: Density plots of parameter value under the prior, the posterior, and the censoring method in the raw scale (left) and in the centered scale (right).

which in turn reduces the utility of the resulting synthetic data generated from sampled parameters.

In practice, we do not propose to use a particular dataset for choosing a target ϵ. Instead, one should rely on experience with the class of data (such as data collected on a temporal basis from the same survey instrument). The synthetic data generation process of censored likelihood works in a similar fashion with the $\boldsymbol{\alpha}-$weighted pseudo posterior mechanism in Section 2.1 where one uses estimated θ^{**} under the censored likelihood of Equation (3.5).

In our simulation studies and application in the sequel, we explore both censoring

the data likelihood without any downweighting of the records and also formulate an embedding of the $\alpha-$weighted pseudo likelihood inside the censoring mechanism. We label the former mechanism as "Censor_uw", which stands for censor unweighted, and the latter as "Censor_w", which stands for censor weighted. The Censor_w construction is a mechanism that would be expected to produce better utility preservation of the real data distribution than the former because censoring is a blunt instrument. The downweighting of likelihood contributions for high-risk records would be expected to invoke censoring less often than under censoring of the likelihood.

We next conduct a Monte Carlo simulation study that generates a collection of size n datasets. We develop a distribution over the Lipschitz distributions (over the Monte Carlo iterations) to illustrate and compare the contraction of these local Lipschitz bounds. We also evaluate and compare the utility performances of these mechanisms for an equivalent privacy guarantee with PH as a comparison.

4. Simulation Studies

We describe our simulation design in Section 4.1. Section 4.2 presents privacy comparison results by investigating the distributions of local Lipschitz bounds and Section 4.3 presents utility comparison results (both global and analysis-specific) with PH as a comparison. We introduce additional techniques to fine tune with downscaling in Section 4.4 and implementation guidance for practitioners using our proposed

17

methods are discussed in Section 4.5. As will be seen, the Weighted and the Censor_w synthesizers are the recommended alternatives given their more refined balance of utility-risk trade-off of the resulting synthetic data.

4.1   Simulation design

Our simulation studies focus on creating differentially private synthetic data from a confidential dataset generated from a univariate, continuous, and bounded outcome variable. For $r = 1, \cdots, R = 100$, we simulate a local database $\mathbf{x}_r$ of size $n = 2000$ from Beta(0.5, 3) that produces a right-skewed confidential data distribution with many high-risk records in the long tail. For each local database $\mathbf{x}_r$, we fit the Unweighted (i.e., $\alpha_i = 1$ for every record) and obtain its Lipschitz bound, $\Delta_{\boldsymbol{\alpha}, \mathbf{x}_r}^{uw}$, and create a synthetic dataset for each of the synthesizers. We next generate a synthetic dataset for each of the following privacy targets, $\epsilon \in \{5, 4, 3\}$, under each synthesizer.

As discussed in Sections 2 and 3, the privacy guarantee of each of our four Bayesian data synthesizers is indirectly determined by its calculated Lipschitz bound, after re-estimation of the pseudo posterior. From the implementation point of view, the Weighted-e inputs a target ϵ and updates or changes some of the weights, α_i, from the Weighted method to produce a different pseudo posterior estimator. By contrast, Censor_w also inputs an ϵ target but does not change the weights α_i; rather, it may censor the pseudo likelihood contribution for a subset of units. Censor_uw

18

works in the same way as Censor_w, but inputs the unweighted likelihood.

4.2   Privacy comparison results

Given that the Weighted and the Weighted-e synthesizers achieve a relaxed aDP privacy guarantee while the Censor_uw and the Censor_w synthesizers achieve a scrict $\epsilon-$DP guarantee, we expect to see that the distribution of Lipschitz bounds of the Weighted-e is more concentrated at the target ϵ compared to the Weighted due to its faster contraction rate, whereas the Censor_w and the Censor_uw have distributions of Lipschitz bounds strictly at or below the target $\epsilon/2$. In what follows, we include the type of privacy guarantee (aDP or DP) in a parenthesis for each synthesizer.

Figure 2 shows violin plots of the Lipschitz bound distributions over $R = 100$ replicated datasets under our simulation settings. Each panel represents the Lipschitz bound distributions of each synthesizer under one of the three target $\epsilon \in \{5, 4, 3\}$. Note that since the Weighted (aDP) is constructed without taking into account a target ϵ, it is identical across the three panels. The same applies to the Unweighted in a comparison figure included in the Supplementary Materials. Also note that given the $\epsilon = 2\Delta$ relationship, the scale of the y-axis in Figure 2, which shows the Lipscthiz bounds, is half of the target ϵ value shown at the top of each panel.

As expected, the Weighted-e (aDP) shows a more concentrated distribution of Lipschitz bounds compared to the Weighted (aDP) at every target ϵ. When the

19

distribution of Lipschitz bounds of the Weighted (aDP) is around the target $\epsilon/2$ (i.e., $\epsilon = 5$ under our simulation settings), the faster concentration gained by the Weighted-e (aDP) is minor. As the target ϵ decreases, the distribution of Lipschitz bounds gets more concentrated at the target $\epsilon/2$, particularly obvious when $\epsilon = 3$. In every panel, the Weighted-e (aDP) cannot enforce a strict upper bound of $\epsilon/2$ of the Lipscthiz bounds over the datasets. In certain cases, as for $\epsilon = 3$ in our simulation, some replicates might have Lipschitz bounds much larger than the target $\epsilon/2$, as can be seen in the longer upper tail of the Weighted-e (aDP) for $\epsilon = 3$, indicating a lack of control of the privacy guarantee for smaller ϵ when using the Weighted-e (aDP).

The Censor_w (DP) and the Censor_uw (DP), by contrast, achieve the strict upper bound of $\epsilon/2$ for the Lipschitz bounds at every target ϵ, as expected. Among the three pairs of the Censor_w (DP) and the Censor_uw (DP), the Censor_w (DP) shows a clear mass of replicates whose Lipschitz bounds are below $\epsilon/2$ when $\epsilon = 5$. Recall that the censored likelihood approach directly truncates the pseudo likelihood of each record into $[\exp(-\epsilon/2), \exp(\epsilon/2)]$. The difference here is that the Censor_w (DP) invokes pseudo likelihood censoring of the Weighted (aDP), while the Censor_uw (DP) does so of the Unweighted. Therefore, the Weighted (aDP) downweights high-risk records which subsequently brings down the Lipshitz bound compared to the Unweighted (excluded in Figure 2 for ease of reading, the Unweighted shows Lipschitz bounds ranging from around 7.5 to 15, whereas the Weighted (aDP) Lipschitz bounds
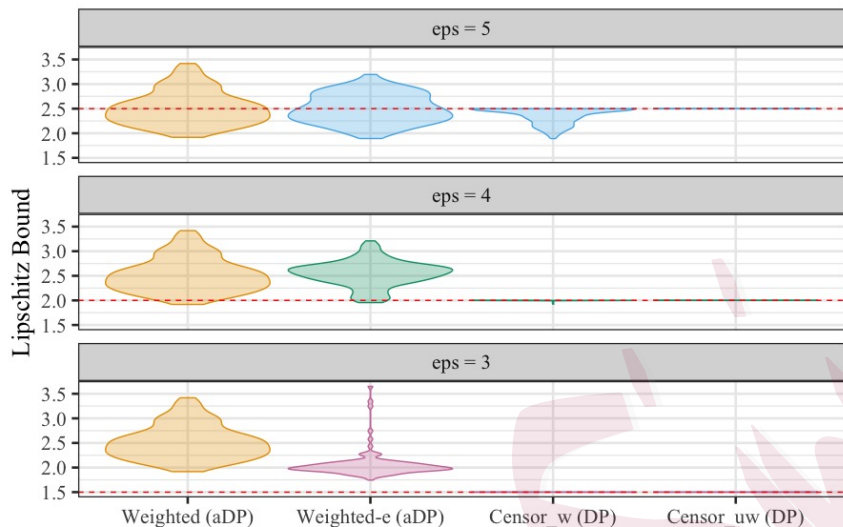
Figure 2: Violin plots of Lipschitz bounds over $R = 100$ replicates under the Weighted (aDP), the Weighted-e (aDP), the Censor_w (DP), and the Censor_uw (DP), with ϵ values of $\{5, 4, 3\}$. A dashed horizontal line at $\epsilon/2$ is included in each panel.

range from around 2 to 3.5; see the comparison figure in the Supplementary Materials), and the same is happening for the Censor_w (DP) compared to the Censor_uw (DP). When $\epsilon = 5$, there may be databases that produce $\epsilon_{\mathbf{x}_r} < 5$ under the Weighted (aDP) method such that few if any records have their likelihood values censored in those databases under Censor_w (DP). By contrast, since Censor_uw (DP) embeds the Unweighted posterior synthesizer, every database will require many records to be censored to enforce the $\epsilon/2 = 2.5$ target Lipschitz bound.

To further illustrate the comparison, Table 1 presents the summaries of the

		Min	Q1	Median	Mean	Q3	Max	sd
$\epsilon = 5$	Weighted-e (aDP)	0	0	0	29	55	166	48.80
	Censor_w (DP)	0	0	0	110	260	341	129.59
	Censor_uw (DP)	196	236	244	247	255	296	17.44
$\epsilon = 4$	Weighted-e (aDP)	0	130	219	195	272	364	103.48
	Censor_w (DP)	0	404	425	408	440	549	92.56
	Censor_uw (DP)	340	398	419	419	435	481	26.73
$\epsilon = 3$	Weighted-e (aDP)	400	536	574	573	618	684	60.62
	Censor_w (DP)	655	715	741	741	763	831	35.66
	Censor_uw (DP)	737	787	806	809	829	888	32.85

Table 1: Summaries of the number of records (out of $n = 2000$) receiving truncated weight at $\alpha_i = 0$ in Weighted-e (aDP) and censored likelihood at $\epsilon/2$ in Censor_w (DP) and Censor_uw (DP). The number of Monte Carlo simulations is $R = 100$.

numbers of invocations of truncation in Weighted-e (aDP) and the numbers of invocations of censoring in Censor_w (DP) and Censor_uw (DP) for each of the target $\epsilon = \{5, 4, 3\}$ over the $R = 100$ Monte Carlo datasets. The summaries include the minimum, Q1, median, mean, Q3, and the maximum of the distribution for invocations of censoring over the $R = 100$ replicates. We also include the standard deviation. Focusing on the three pairs of the Censor_w (DP) and the Censor_uw (DP), we can

clearly see that for every target ϵ, there are more invocations of censoring in the Censor_uw (DP) than in the Censor_w (DP) in terms of the means, resulted from the aforementioned downweighting effects of the Weighted (aDP) in the Censor_w (DP). Moreover, as ϵ decreases, the numbers of invocations of censoring increase for both the Censor_w (DP) and the Censor_uw (DP). This is expected since a smaller range of $[\exp(-\epsilon/2), \exp(\epsilon/2)]$ is used in censoring pseudo likelihood as ϵ decreases, resulting in larger numbers of invocations of censoring.

Focusing on the number of invocations of truncating $\alpha_i = 0$ for Weighted-e (aDP) in Table 1, we observe that same with the censoring, as ϵ decreases, the number of invocations of truncation increases. The number of invocations of truncation of the Weighted-e (aDP) is overall smaller than the number of invocations of censoring for the Censor_w (DP). Yet, despite the reduced number of invocations of censoring over the collection of Monte Carlo datasets, Censor_w (DP) produces the same censored ϵ privacy guarantee as does Censor_uw (DP).

4.3   Utility comparison results

We consider both global utility that describes preservation of the confidential data distribution and analysis-specific utility of the generated synthetic data. We choose the empirical CDF (ECDF) global utility measure, which evaluates whether the distributions for the confidential dataset and the synthetic dataset can be discrim-

inated from each other (Woo et al., 2009). There are two specific measures based on ECDF: 1) the maximum record-level absolute difference (max-ECDF); and 2) the average record-level squared difference (avg-ECDF). Both are calculated as distances between the two ECDFs of the confidential data and the synthetic data. Each measure is non-negative, and a smaller value indicates higher similarity between the two ECDFs, suggesting higher global utility. Violin plots showing the distributions of the max-ECDF for all proposed methods are displayed in Figure 3, and those of the avg-ECDF are in Figure 4, across $R = 100$ Monte Carlo simulations. As before, each panel represents a target $\epsilon = \{5, 4, 3\}$.

As evident in Figures 3 and 4, the Weighted (aDP) achieves the lowest max-ECDF and avg-ECDF across all synthesizers. These are expected since the Weighted (aDP) is not influenced by a target ϵ and has the overall highest Lipschitz bounds (i.e., highest risks), as illustrated in Figure 2.

Among synthesizers focused on achieving a global privacy target, the Censor_w (DP) has the highest utility: It expresses the lowest max-ECDF distribution for every target ϵ in Figure 3 and also the lowest avg-ECDF distribution for target $\epsilon = \{4, 3\}$ in Figure 4. Censor_w (DP) embeds the Weighted (aDP) method inside the censoring mechanism, which downweights high-risk likelihood contributions, resulting in less invocations of the censored likelihood. Although censoring is invoked relatively less often (fixing an ϵ) under embedding of the Weighted (aDP) method, it nevertheless
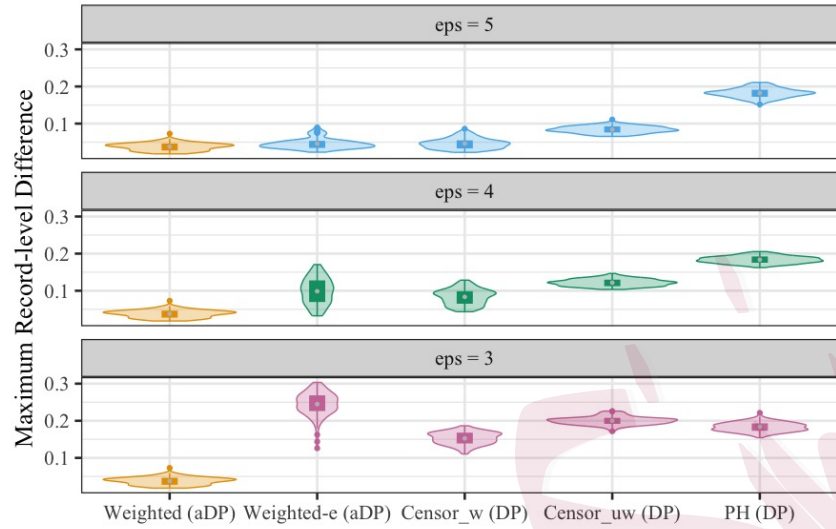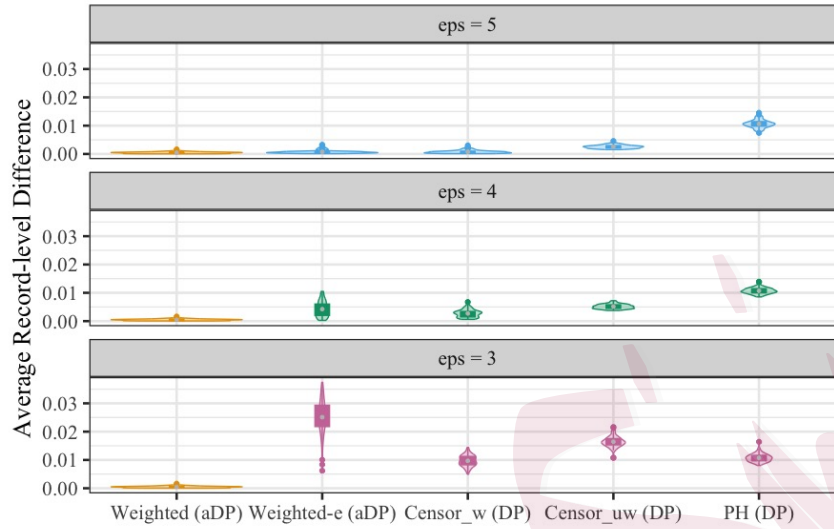
24

Figure 3: Violin plots of max-ECDF utility over $R = 100$ replicates, for the Weighted (aDP), the Weighted-e (aDP), the Censor_w (DP), the Censor_uw (DP), and the PH (DP), with ϵ values of $\{5, 4, 3\}$.

comes equipped with an $\epsilon-$DP guarantee because the censoring procedure is applied to all datasets. The Weighted-e (aDP) method, by contrast, seems to have an even more concentrated and lower distribution when target $\epsilon = 5$. However, as ϵ decreases, the Weighted-e (aDP)'s performance starts to deteriorate rapidly, which indicates that it lacks not only the control of the privacy guarantee (shown in Figure 2) but also the control of utility preservation. The Censor_uw (DP) performs consistently worse than the Censor_w (DP) at every target ϵ value, an expected result since the Censor_uw (DP) invokes censoring for a larger number of records than the Censor_w

Figure 4: Violin plots of avg-ECDF utility over $R = 100$ replicates, for the Weighted (aDP), the Weighted-e (aDP), the Censor_w (DP), the Censor_uw (DP), and the PH (DP), with ϵ values of $\{5, 4, 3\}$.

(DP) (shown in Table 1) which should result in lower utility. The PH (DP), a non-Bayesian data synthesizer included for comparison, clearly performs worse than the almost all proposed Bayesian synthesizers. Overall, all synthesizers (except for the Weighted (aDP) which is not affected by the target ϵ value) have lower utility as ϵ decreases.

Figures 5 displays violin plots of the distributions of the median statistic across $R = 100$ Monte Carlo simulations for $\epsilon = 5$. (Results of the 15th and the 90th quantile statistics and all three statistics for $\epsilon = \{4, 3\}$ tell a similar story and are
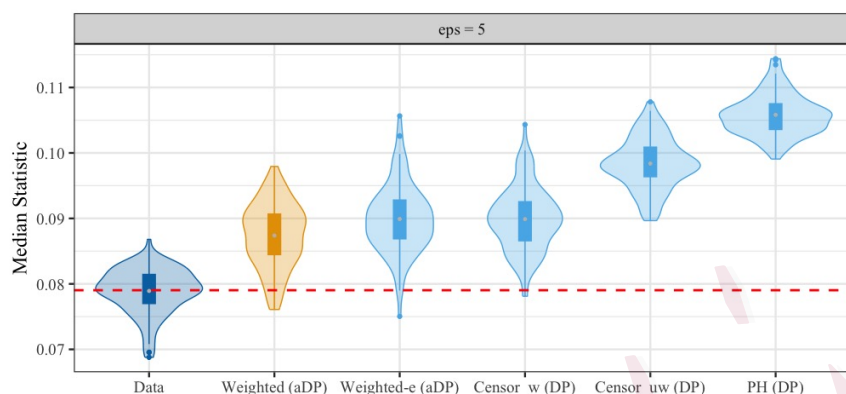
26

Figure 5: Violin plots of median over $R = 100$ replicates for the Weighted (aDP), the Weighted-e (aDP), the Censor_w (DP), the Censor_uw (DP), and the PH (DP), at $\epsilon = 5$. A dashed horizontal line at the analytical median from Beta(0.5, 3) is included.

in the Supplementary Materials.) The distributions from the confidential data are included and labeled as "Data" for comparison. A dashed horizontal line marking the analytical value from Beta(0.5, 3), the generating distribution, is included in each plot. The closer the resulting violin plot from a data synthesizer is to the one for Data and the dashed horizontal line, the higher the utility is for that statistic.

As with the global utility evaluation, the Weighted (aDP) has overall the highest utility, at the price of achieving an aDP guarantee. Among the remaining synthesizers, the Censor_w (DP) once again shows the highest utility performance for the median statistic. The Weighted-e (aDP) overall has better utility performance on

27

these analysis-specific utility measures as compared to the global utility measures discussed earlier. The Censor_uw (DP) performs worse than the Censor_w (DP) in all statistics because censoring is invoked on more data records in every dataset.

4.4   Fine tuning with downscaling

As Figures 3 and 4 of global utility have shown, as ϵ decreases from 5 to 4 or 3, the utility performances of Censor_w (DP) deteriorates because more invocations of censoring across records occur as ϵ decreases. Further scaling and shifting of the weights $\boldsymbol{\alpha}$ in the Weighted (aDP) method may be used in order to achieve a target privacy guarantee for any confidential dataset; for example, by setting (c_1, c_2) to $\tilde{\boldsymbol{\alpha}} \in [0,1]$ (as in $\tilde{\alpha}_i = c_1 \times \alpha_i + c_2$). Instead of using $c_1 = 1$ as we have done in Sections 4.2 and 4.3, we use a scaling factor $c_1 < 1$ that downscales the privacy weights $\boldsymbol{\alpha}$. The weight scaling allows us to roughly target a specific ϵ and censoring is then used to make that target global. This procedure reduces the number of invocations of censoring while preserving a strict $\epsilon-$DP guarantee.

We focus on $\epsilon = 4$ for illustration. We set $c_1 = 0.8$ for the results that follow. Since only the Weighted-e (aDP) and the Censor_w (DP) depend on the Weighted (aDP) method whereas Censor_uw (DP) embeds the Unweighted method, we expect to see effects of downscaling on the Weighted (aDP), the Weighted-e (aDP), and the Censor_w (DP) synthesizers. Therefore, we focus our comparison on these three
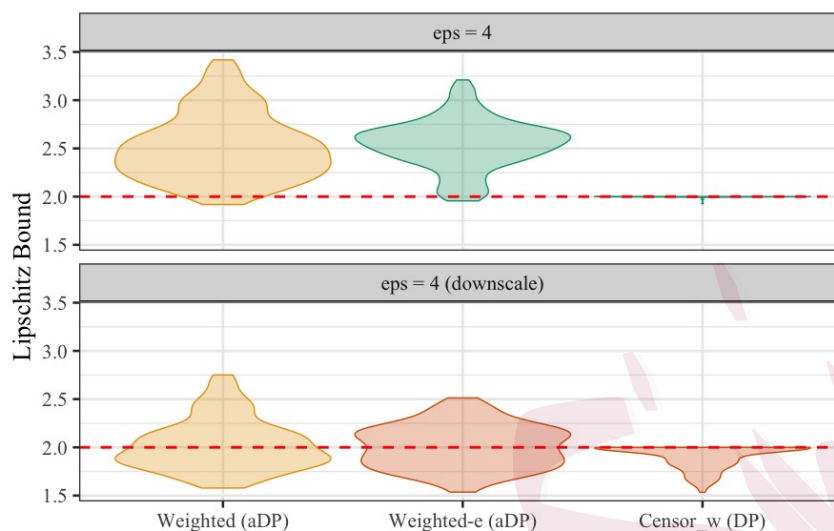
Figure 6: Violin plots of Lipschitz bounds over $R = 100$ replicates under the Weighted (aDP), the Weighted-e (aDP), and the Censor_w (DP), at $\epsilon = 4$, without downscaling (top) and with downscaling (bottom). A dashed horizontal line at $\epsilon/2$ is included.

synthesizers.

To evaluate the privacy guarantee and utility performances without and with downscaling, we generate $R = 100$ replicate datasets in a Monte Carlo simulation study. Figure 6 shows with downscaling, the Lipschitz bounds are lower and more concentrated than those without downscaling, indicating a better control of the privacy guarantee for these three synthesizers. Figure 7 illustrates the effects of downscaling on the avg-ECDF based global utility measure (results of the max-ECDF are included in the Supplementary Materials). As can be seen, the violin plots of
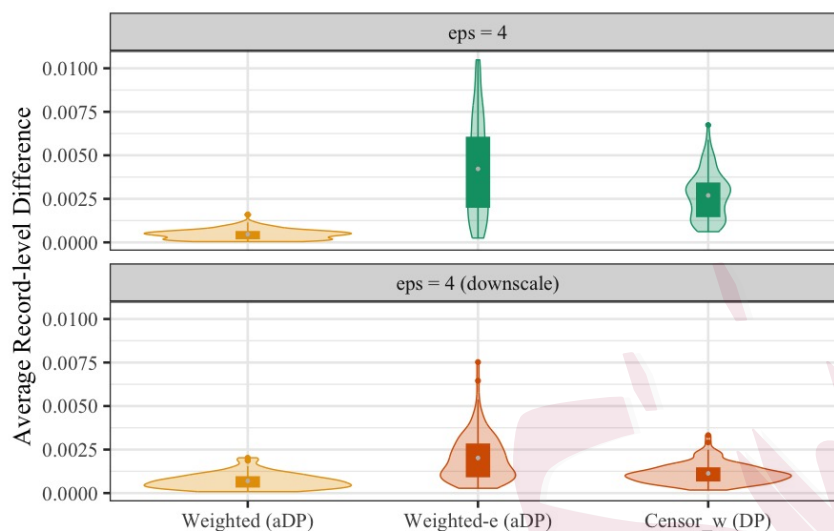
29

Figure 7: Violin plots of avg-ECDF utility over $R = 100$ replicates, for the Weighted (aDP), the Weighted-e (aDP), and the Censor_w (DP) at $\epsilon = 4$, without downscaling (top) and with downscaling (bottom).

the Weighted (aDP), the Weighted-e (aDP), and the Censor_w (DP) are all shifted to lower values under downscaling, indicating higher utility. Moreover, as shown in Figure 8 for the 15th quantile statistic utility, downscaling also improves this and other analysis-specific utility measures (available in the Supplementary Materials).

4.5   Implementation guidance

We next describe a recommended procedure to achieve some target ϵ for a collection of time-indexed datasets (such as the monthly release of total employment for each
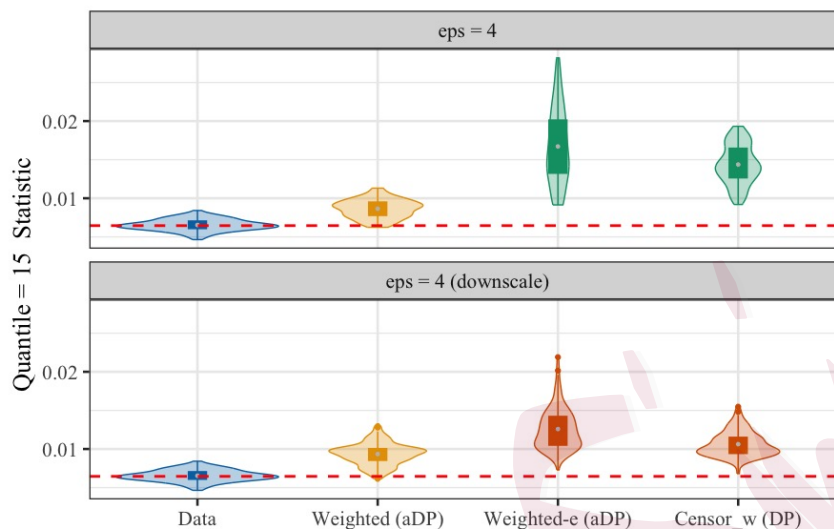
Figure 8: Violin plots of the 15th quantile over $R = 100$ replicates for the Weighted (aDP), the Weighted-e (aDP), and the Censor_w (DP) at $\epsilon = 4$, without downscaling (top) and with downscaling (bottom). A dashed horizontal line at the analytical 15th quantile from Beta(0.5, 3) is included.

state) that uses censoring to acheive a strict $\epsilon-$DP guarantee. We note that our implementation suggestions apply when using any reference dataset with relaxed aDP guarantees, not necessarily those that rely on regularly scheduled releases and temporal autocorrelation.

Using the current period database to set the privacy ϵ for any mechanism can risk leaking a small amount of information about the overall sensitivity, Δ, of the database under a strict $\epsilon-$DP guarantee. Under the pseudo posterior mechanism

equipped with a relaxed, asymptotic DP guarantee, however, the risk of disclosure about the sensitivity of the database is very low. The reason for the low risk is because the record-indexed weights under the pseudo posterior are set to be inversely proportional to their by-record sensitivities such that the resulting mechanism overall sensitivity (and thus local ϵ) after weighting is nearly unrelated to the underlying sensitivity of the local database. We see this by the increased concentration of the distribution of by-record sensitivities after weighting from before weighting for the pseudo posterior mechanism (Savitsky et al., 2022). (We remind the reader that the resulting weights are not published).

We recommend utilizing the censoring mechanism with the pseudo posterior to achieve a strict $\epsilon-$DP guarantee, though we propose to use the sensitivity and utility measures of a *historical* (rather than the current) database to set the global ϵ under the censoring mechanism in an abundance of caution. In practice, the owner of the confidential data (such as a government statistical agency) publishes a database of size n on a regular, periodic basis (e.g., the publication of monthly employment and unemployment by the U.S. Bureau of Labor Statistics).

In the first period of the year, we recommend that the owner set the privacy guarantee ϵ by first constructing $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)$ as outlined in Section 2.1, and re-estimating the model under the pseudo posterior mechanism to compute the associated Δ_α. This gives an asymptotic aDP guarantee of $2\Delta_\alpha$.

32

Next, we recommend to shift and scale the weights to an $\tilde{\boldsymbol{\alpha}}$ that produces a target ϵ on the first period dataset under an aDP guarantee. This may take repeated model re-estimations to evaluate the utility for each shifting and scaling of the weights to assess the risk-utility trade-off needed to pick the target $\tilde{\boldsymbol{\alpha}}$.

In future months, the *same* ϵ would be used to censor the pseudo likelihood contributions with the weights scaled and shifted by the same (c_1, c_2) as used in the first month. In this way, the many computations used to set (c_1, c_2, ϵ) are performed only once. These settings will likely result in fewer invocations of censoring to produce synthetic data in future months under an $\epsilon-$DP guarantee such that the resulting utility of the synthetic data would be expected to be similar to the first month.

5.   Application to Survey of Doctoral Recipients

In this application section, we apply the Weighted (aDP), the Weighted-e (aDP), the Censor_w (DP), the Censor_uw (DP), and the PH (DP) synthesizers to a sample of the Survey of Doctoral Recipients. Our sample comes from the public use file published for 2017. It contains information on salary, gender, age (grouped in 5-year intervals), and the number of working weeks for $n = 1601$ survey respondents who have positions at a 4-year college or university in the field of mathematics and statistics. Our Unweighted model is a beta regression, where the outcome variable salary is scaled into $[0, 1]$ and gender (categorical with 2 levels), age (categorical with 9
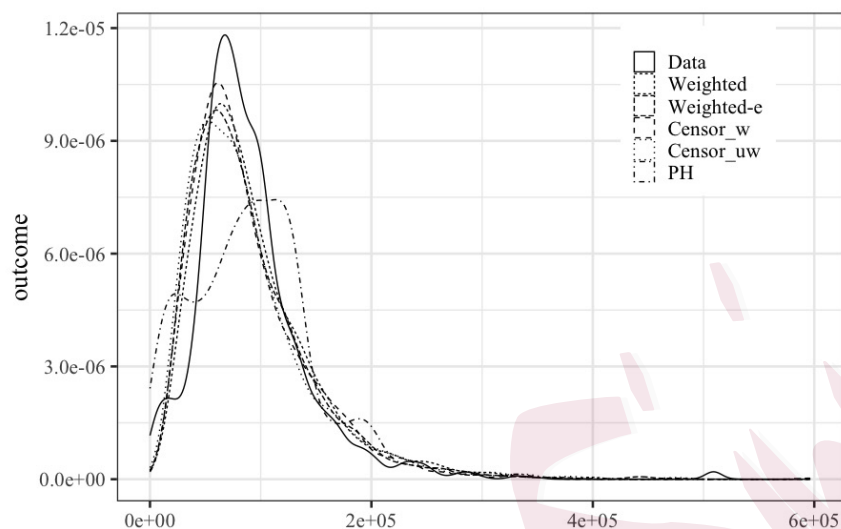
33

Figure 9: Density plots of the confidential data sample (Data) and the synthetic data samples from the Weighted (aDP), the Weighted-e (aDP), the Censor_w (DP), the Censor_uw (DP), and the PH (DP) synthesizers at $\epsilon = 5$.

levels), and weeks (numerical ranging from 2 weeks to 52 weeks) as publicly available predictors. The target privacy budget is $\epsilon = 5$, which is used in our implementation of the perturbed histogram synthesizer for comparison. For every synthesizer, one synthetic dataset is simulated.

Figure 9 displays density plots of the confidential data sample and those from the synthetic datasets simulated from the five synthesizers. The PH (DP) evidently has the worst utility performance in this visual check. Among the four Bayesian synthesizers, the Censor_w (DP) seems to capture the main feature of the confidential

34

data the best, with the Weighted (aDP) coming in second, while the Weighted-e (aDP) and the Censor_uw (DP) come in third and fourth, respectively.

Table 2 presents detailed privacy and utility results of our synthesizers. In terms of privacy, we include the calculated Lipschitz bound of the Weighted (aDP), the Weighted-e (aDP), the Censor_w (DP), and the Censor_uw (DP) and their corresponding ϵ values. Note that Data and the PH (DP) have no Lipschitz bounds and therefore coded as NA, and Data has no privacy guarantee which is coded as NA as well. As we know from Sections 2 and 3 and the simulation study results in Section 4.2, the Censor_w (DP) and the Censor_uw (DP) have strict Lipschitz bound at $\epsilon/2$, whereas the Weighted (aDP) and the Weighted-e (aDP) only contract to $\epsilon/2$ but cannot guarantee $\epsilon/2$ (with the Weighted-e (aDP) contracting faster than the Weighted (aDP)). As expected, in our application, the Censor_w (DP) and the Censor_uw (DP) achieve the target $\epsilon = 5$ whereas the Weighted (aDP) and the Weighted-e (aDP) have privacy guarantee 5.24 and 5.22 respectively, both exceed the target privacy budget.

Table 2 also includes global utility and analysis-specific utility results for all synthesizers. For the two global utility measures, max-ECDF and avg-ECDF, the smaller the value, the higher the utility. (Data has no such measures and therefore coded as NA.) For the four analysis-specific utility measures, the mean, the median, the 15th quantile, and the 90th quantile statistics, the closer they are to the value from the Data, the higher the utility. For ease of reading, the value of the best

35

	Data	Weighted (aDP)	Weighted-e (aDP)	Censor_w (DP)	Censor_uw (DP)	PH (DP)
Lipschitz	NA	2.62	2.61	2.50	2.50	NA
Privacy ϵ	NA	5.24	5.22	5.00	5.00	5.00
max-ECDF	NA	**0.0656**	0.1020	<u>0.0968</u>	0.1350	0.1310
avg-ECDF	NA	**0.0011**	<u>0.0025</u>	0.0026	0.0039	0.0057
Mean	91019	<u>92994</u>	**89525**	88581	88840	93654
Median	80000	**80135**	<u>76451</u>	75642	75211	91180
15th Q	51000	**44303**	40574	<u>41142</u>	38107	30108
90th Q	150000	162351	**158037**	<u>158426</u>	162686	163872

Table 2: Privacy and utility results of the Weighted (aDP), the Weighted-e (aDP), the Censor_w (DP), the Censor_uw (DP), and the PH (DP) at $\epsilon = 5$. In the utility rows, the best performing synthesizer is in bold and the second best is underlined. performing synthesizer among the five for each statistic is in bold, and the second best is underlined.

Undoubtedly, the Weighted (aDP) has the highest utility performance, and the Weighted-e (aDP) comes second while the Censor_w (DP) comes third. The inferior results of the Censor_uw (DP) and the PH (DP) are expected: The Censor_uw (DP) embeds the Unweighted synthesizer into censoring, which results in the invocation

of censoring for more records than Censor_w (DP). The PH (DP) does not take advantage of the benefits from modeling and the useful predictor variables available as public information. Although the Weighted-e (aDP) performs slightly better than the Censor_w (DP) in terms of utility in our application, such improvement cannot offset the fact that it provides an asymptotic privacy guarantee.

6. Concluding Remarks

In this paper, we review the pseudo posterior mechanism from (Savitsky et al., 2022) which achieves an asymptotic DP guarantee. To provide a stronger, non-asymptotic DP guarantee, we propose the censoring mechanism that censors the pseudo likelihood of every record within $[\exp(-\epsilon/2), \exp(\epsilon/2)]$. This new mechanism truncates large absolute log-likelihood values, which serves to cap the Lipschitz bound and therefore achieving an $\epsilon-$DP guarantee. It offers a practical, low-dimensional alternative to truncating the parameter space.

Our simulation studies on $[0, 1]$ bounded univariate data under repeated sampling demonstrate the superior utility preservation performance of the pseudo posterior mechanism at the cost of an asymptotic DP guarantee, on the one hand. The censoring mechanism that embeds the pseudo posterior mechanism provides a stronger, non-asymptotic DP guarantee at the price of slightly reduced utility performance, on the other hand. We recommend these two alternatives for data disseminators who

depending on their priority, can choose the preferred strategy between these two.

One interesting line of future work is embedding other weighted synthesizers in censoring to provide a non-asymptotic DP guarantee. Another is to investigate, in an applied setting, how similar do the data distributions at two points in time need to be for their utilities to be similar under the same scaling, shifting, and privacy loss parameters.

Supplementary Materials

Supplementary Materials include a detailed review of the perturbed histogram synthesizer, the Stan script for the censoring method, and additional privacy and utility comparison results from Section 4.

Acknowledgements

in this paper are the authors' and do not necessarily reflect the policies and views of NSF, BLS, DOE, ORAU, or ORISE.

References

Cannone, C. L., G. Kamath, A. McMillan, A. Smith, and J. Ullman (2019). The structure of optimal private tests for simple hypotheses. In *Proceedings of the Fifty-first ACM SIGACT Symposium on Theory of Computing*, pp. 310–321.

Dimitrakakis, C., B. Nelson, Z. Zhang, A. Mitrokotsa, and B. I. P. Rubinstein (2017, January). Differential privacy for bayesian inference through posterior sampling. *Journal of Machine Learning Research 18*(1), 343–381.

Dwork, C., F. McSherry, K. Nissim, and A. Smith (2006). Calibrating noise to sensitivity in private data analysis. In *Proceedings of the Third Conference on Theory of Cryptography*, TCC'06, Berlin, Heidelberg, pp. 265–284. Springer-Verlag.

Hu, J., T. D. Savitsky, and M. R. Williams (2022). Private tabular survey data products through synthetic microdata generation. *Journal of Survey Statistics and Methodology 10*, 720–752.

Little, R. J. A. (1993). Statistical analysis of masked data. *Journal of Official Statistics 9*, 407–426.

McSherry, M. and K. Talwar (2007). Mechanism design via differential privacy. In *Proceedings of the 48th Annual IEEE Symposium on Foundations of Computer Science*, pp. 94–103.

Pena, V. and A. F. Barrientos (2021). Differentially private methods for managing model uncertainty in linear regression models. pp. arXiv:2109.03949.

Rubin, D. B. (1993). Discussion statistical disclosure limitation. *Journal of Official Statistics 9*, 461–468.

Savitsky, T. D., M. R. Williams, and J. Hu (2022). Bayesian pseudo posterior mechanism under asymptotic differential privacy. *Journal of Machine Learning Research 23*, 1–37.

Wasserman, L. and S. Zhou (2010). A statistical framework for differential privacy. *Journal of the American Statistical Association 105*, 375–389.

Woo, M.-J., J. P. Reiter, A. Oganian, and A. F. Karr (2009). Global measures of data utility for microdata masked for disclosure limitation. *Journal of Privacy and Confidentiality 1*(1).

Vassar College, Box 27, 124 Raymond Ave, Poughkeepsie, NY 12604

E-mail: jihu@vassar.edu

RTI International, 3040 East Cornwallis Road, Research Triangle Park, NC 27709

E-mail: mrwilliams@rti.org

U.S. Bureau of Labor Statistics, Office of Survey Methods Research, Suite 5930, 2 Massachusetts Ave NE

Washington, DC 20212

E-mail: Savitsky.Terrance@bls.gov