Differentially Private Population Quantity Estimates via Survey Weight Regularization (Paper Proposal)

Jeremy Seeman, Yajuan Si, and Jerome P. Reiter

December 18, 2023

Abstract

Applying differential privacy (DP) to the release of survey-weighted statistics remains challenging with complex sample survey data, primarily because weighted statistics are more sensitive to individual survey responses than unweighted statistics. However, when weighted and unweighted statistics are similar, privacy-preserving noise regularly dominates any bias corrections for population representation. In this paper, we present a DP method for population quantity estimates by first, privately estimating a hyperparameter that determines how much to regularize or shrink survey weights as a function of privacy loss. By adaptively navigating this three-way trade-off, bias-variance-privacy, for each population quantity estimate, we release statistics whose sensitivity is tailored to the particular relationship between survey weights and response variables. We apply our results to the Panel Study of Income Dynamics (PSID), and show that optimal strategies for releasing DP survey-weighted income estimates require orders-of-magnitude less DP noise. We conclude with discussions on further details for the full paper.

1 Introduction

Privacy protection is crucial for the public release of complex sample survey data. Existing applications of differentially private (DP) [7, 8] methods are typically based on census or administrative data and fail to account for the data collection process. Survey weights are often available to adjust for the unequal probability of individual selection and balance the sample decomposition with the target population. Weighted statistics offer unbiased and consistent estimates for finite population quantities, such as the overall population mean. We extend DP to complex sample surveys in releasing survey-weighted statistics.

Preliminary work at the intersection of DP and survey statistics has focused on synthetic data generation [11], survey statistics under specific sampling designs [13], and privacy interpretations of existing survey sampling methods [4, 10]. Each of these approaches attempts to utilize as much information as possible about the sampling process. However, weighting schemes can cause practical problems for DP, as weighted statistics have significantly larger sensitivities than their unweighted counterparts, hence requiring substantially more noise to provide the equivalent level of DP protections at the sample level of privacy loss [6]. Moreover, while survey weights may theoretically correct for selection bias, weighted and unweighted estimates may be quite similar if the weights are uncorrelated with a particular survey outcome [3]. Additionally, anomalously large survey weights can significantly increase the variability of survey statistics, prompting approaches to smoothing the estimates or regularizing survey weights [9, 2].

Introducing DP into survey inference suggests that the degree of regularization should depend on privacy loss budgets; just as there is a bias-variance-privacy trilemma for mean estimation with independently identically distributed (iid) records [12], similar three-way trade-offs must be

made when analyzing survey-weighted quantities. Our work proposes methods for DP survey-weighted estimates where the optimal degree of regularization depends on the confidential data. In this setting, we must consume privacy loss to estimate this optimal degree of regularization *and* the statistics of interest. Doing so allows us to adaptively consume privacy loss budget for fine-tuning uncertainty quantification when constructing interval estimates and assessing their coverage properties.

2 Methods

We consider a response variable and survey weights $\{(y_i, w_i)\}_{i=1}^n$ lying within bounded intervals $[L_Y, U_Y] \times [L_W, U_W]$, and our goal is to estimate a weighted mean and variance estimates with respect to the inverse-proportional sampling variability, where $w_i \propto \mathbb{P}(i \text{ in sample})^{-1}$. Note that we consider $\sum_{i=1}^n w_i$ to be public information, since survey weights can be arbitrarily scaled by a constant factor, which is often the population size, the sample size, or 1. We will use zero-concentrated DP [5], i.e., for two samples $\{(y_i, w_i)\}_{i=1}^n \sim \{(y_i', w_i')\}_{i=1}^n$ which differ on a single entry and a randomized algorithm of outputs M, we have, for all $\alpha \in (1, \infty)$,

$$d_\alpha \left(M(\{(y_i, w_i)\}_{i=1}^n) || M(\{(y_i', w_i')\}_{i=1}^n) \right) \leq \rho\alpha, \tag{1}$$

where $d_\alpha(\cdot, \cdot)$ is the α-Rényi divergence. Our goal is to use ρ-zCDP to estimate

$$\hat{\theta}_0 \triangleq \frac{\sum_{i=1}^n w_i y_i}{\sum_{i=1}^n w_i}, \qquad \hat{\text{Var}}(\hat{\theta}_0) \triangleq \frac{\sum_{i=1}^n w_i (y_i - \hat{\theta}_\lambda)^2}{\left(\sum_{i=1}^n w_i\right)^2}, \tag{2}$$

the original weighted mean and variance. We propose a linear regularization of the survey weights using parameter $\lambda \in [0, 1]$ given by

$$G_\lambda(w_i) \triangleq (1 - \lambda) w_i + \frac{\lambda}{n} \sum_{i=1}^n w_i, \tag{3}$$

where $\lambda = 0$ corresponds to the original weights and $\lambda = 1$ corresponds to constant weights. Furthermore, λ quantifies the proportion of bias reduced relative to the unweighted statistic; for example, if $\lambda = .25$, it means $\hat{\theta}_{.25}$ reduces 75% of bias from the unweighted statistic. This regularization yields modified weighted mean and variance estimates

$$\hat{\theta}_\lambda \triangleq \frac{\sum_{i=1}^n G_\lambda(w_i) y_i}{\sum_{i=1}^n G_\lambda(w_i)}, \qquad \hat{\text{Var}}(\hat{\theta}_\lambda) \triangleq \frac{\sum_{i=1}^n G_\lambda(w_i)(y_i - \hat{\theta}_\lambda)^2}{\left(\sum_{i=1}^n G_\lambda(w_i)\right)^2}, \tag{4}$$

where $\hat{\theta}_0$ $\hat{\text{Var}}(\hat{\theta}_0)$ are the original, non-DP estimates we hope to approximate. Doing so yields a new weighted mean sensitivity

$$\Delta_\lambda \triangleq \frac{G_\lambda(U_W)\Delta_Y + L_Y(1 - \lambda)\Delta_W}{n G_\lambda(L_W)}. \tag{5}$$

We consider three sources of privacy loss, ρ_1, ρ_2, and ρ_2, each spent on sequentially estimating $\hat{\lambda}, \hat{\theta}_{\text{DP}}$, and $\hat{\text{Var}}(\hat{\theta})_{\text{DP}}$, respectively. For the Gaussian mechanism as applied to θ_λ, we have the following mean-square error decomposition,

$$\mathbb{E}[(\hat{\theta}_{\text{DP}} - \hat{\theta})^2] = \lambda^2 (\hat{\theta}_1 - \hat{\theta}_0)^2 + \frac{\Delta_\lambda^2}{2\rho_2}, \tag{6}$$

which is minimized by a data-dependent and privacy-loss-parameter-dependent value of λ, which we call λ_{opt}. This defines the bias-variance-privacy three-way trade-off as a function of the privacy loss for the mean estimation ρ_2.

Next, we outline our procedure in Algorithm 1. In Equation 7, we use the exponential mechanism [14] to select $\hat{\lambda}$ which minimizes this MSE while satisfying ρ_1-zCDP. Second, in

Algorithm 1 DP regularized survey-weighted statistics

Require: $\rho_1, \rho_2, \rho_3 \in (0, \infty)$, $\{(y_i, w_i)\}_{i=1}^n \in \{[L_Y, U_Y] \times [L_W, U_W]\}^n$

Sample $\hat{\lambda}$ from the density $f(\lambda)$ where

$$f(\lambda) \propto \mathbb{1}_{\{\lambda \in [0,1]\}} \exp\left(-\frac{\sqrt{\rho_1}}{\sqrt{2}\left(\frac{U_Y \Delta_W}{n L_W}\right)^2} \left[\lambda^2 (\hat{\theta}_1 - \hat{\theta}_0)^2 + \frac{\Delta_\lambda^2}{2\rho_2} \right] \right) \tag{7}$$

Sample

$$\hat{\theta}_{\mathrm{DP}} \sim N\left(\hat{\theta}_{\hat{\lambda}}, \frac{1}{2\rho_2} \left(\frac{G_{\hat{\lambda}}(U_W)\Delta_Y + L_Y(1-\hat{\lambda})\Delta_W}{n G_{\hat{\lambda}}(L_W)} \right)^2 \right) \tag{8}$$

Sample

$$\hat{\mathrm{Var}}(\hat{\theta})_{\mathrm{DP}} \sim N\left(\hat{\mathrm{Var}}(\hat{\theta}_{\hat{\lambda}}), \frac{1}{2\rho_3} \left(\frac{G_{\hat{\lambda}}(U_W)^2}{n^2 G_{\hat{\lambda}}(L_W)^2} \left| \Delta_Y^2 + (n-1) 2 \Delta_Y \Delta_{\hat{\lambda}} \right| \right)^2 \right) \tag{9}$$

return $\hat{\lambda}, \hat{\theta}_{\mathrm{DP}}, \hat{\mathrm{Var}}(\hat{\theta})_{\mathrm{DP}}$

Equation 8 we use $\hat{\lambda}$ and ρ_2 with the Gaussian mechanism to estimate $\hat{\theta}_{\mathrm{DP}}$ satisfying ρ_2-zCDP; similarly, in Equation 9, we use $\hat{\lambda}$ and ρ_3 to estimate $\hat{\mathrm{Var}}(\hat{\theta})_{\mathrm{DP}}$ to satisfy ρ_3-zCDP. By adaptive composition, the final result is $(\rho_1 + \rho_2 + \rho_3)$-zCDP. The full proof is in Appendix B.

Using $\hat{\lambda}, \hat{\theta}_{\mathrm{DP}}$, and $\hat{\mathrm{Var}}(\hat{\theta})_{\mathrm{DP}}$, we can construct confidence intervals surrounding $\hat{\theta}$, the non-DP estimator, or θ, the true population mean under a theoretical inverse-proportional-sampling distribution. Both cases require accounting for all sources of error in the resulting confidence intervals. Errors in the DP mean and variance estimates depend only on the Gaussian mechanism errors. To account for bias, we leverage the relationship between $\hat{\lambda}, \lambda_{\mathrm{opt}}$, and $B \triangleq (\hat{\theta}_1 - \hat{\theta}_0)^2$ to bound the bias. Note that by using the exponential mechanism, there exists a constant C that depends on $\gamma > 0$ and ρ_1, ρ_2 such that

$$\mathbb{P}(|\hat{\lambda} - \lambda_{\mathrm{opt}}| > \gamma) \leq C(\gamma, \rho_1, \rho_2). \tag{10}$$

Since there exists a monotonic function $M(\cdot)$ where $\lambda_{\mathrm{opt}} = M(B)$, this implies

$$\mathbb{P}(B \in [M^{-1}(\hat{\lambda} + \gamma), M^{-1}(\hat{\lambda} - \gamma)]) \leq C(\gamma, \rho_1, \rho_2). \tag{11}$$

The expressions above can be numerically evaluated using publicly known constants and the DP results. A similar procedure can be performed for potential biases in $\hat{\mathrm{Var}}(\hat{\theta})_{\mathrm{DP}}$. Combining all these yields the confidence intervals of interest.

3 Application: The Panel Study of Income Dynamics (PSID)

To demonstrate the methodology above, we apply it to the the Panel Study of Income Dynamics (PSID), a longitudinal survey containing family-level statistics on income sources and other sociodemographic information [17]. We use data published from 2019, consisting of $n = 9420$ families. Our goal is to estimate population quantities describing family income, such as the mean. We apply the cube-root transformation so the resulting values are approximately normally distributed. We plot the relationship between survey weights and family income in Figure 1. Lower-income families are oversampled in the survey, as we see in Figure 2; however, the final composite survey weights are weakly correlated with family income (Spearman's rank correlation of approximately .14).

Below, we present some preliminary results. Figure 3 shows how the theoretical bias-variance-privacy trade-off manifests for estimating the survey-weighted average cube-root income (`inc3`) and proportion of families below the 2019 poverty line (`pov`). For comparison purposes, we also include two hypothetical responses: simulated iid Bernoulli(.5) responses (`bernoulli`) and a copy of the survey weights themselves (`wgt`), representing minimal and maximal correlation between survey weights and responses. We plot the noise-to-signal ratio as the theoretical MSE over the weighted mean estimate on the y-axis, with the regularization parameter λ on the x-axis. We see that as the magnitude of the bias decreases (moving from top left subfigure to bottom right subfigure), the optimal MSE is achieved at larger values of λ_{opt} for the same privacy loss budget ρ_2. Moreover, as ρ_2 decreases, λ_{opt} increases for each response variable under consideration. We see that for reasonably small choices of ρ_2, we tend to reject small λ to optimize the bias-variance trade-off at each fixed ρ_2 value.

Figure 4 shows the sampling distribution of $\hat{\lambda}$ for estimating cube-root income at different values of ρ_1 and ρ_2. By construction, this statistic is the most sensitive by depending on the gap between the weighted and unweighted population means. Therefore we do not sample λ particularly close to the optimal $\hat{\lambda}$ without a large ρ_1. However, even for small values of ρ_1, we can reasonably avoid sampling small values of λ with high probability; for example, when $\rho_1 = .1$, we are highly unlikely to select values of λ below .25, even for smaller ρ_2 values.

4 Further details and discussion

This paper proposal covers preliminary results that theoretically and empirically suggest survey weight regularization drastically reduces the amount of additional noise needed to preserve DP. Our end goal, however, is to produce methods which are operationally feasible while allowing precise uncertainty quantification through the entire decision-making process. While our proposed statistics can admit the construction of valid finite-sample confidence intervals, these theoretical methods may be too conservatively wide or provide excessively large empirical coverage guarantees. Moreover, we want to provide guidance for practitioners on how to incorporate prior information into selecting parameters such as the privacy losses ρ_1, ρ_2, and ρ_3.

We propose two main directions to continue this work for presentation at the NBER workshop. First, we will perform more extensive empirical studies on the PSID data in a few different ways. We will perform our analysis with a wider range of response variables, including those with different response types (for example, binary and ordinal integer-valued responses) and different biases and correlations between survey weights and responses (for example, simulated responses with artificially low and high bias and/or correlation with the survey weights). Additionally, by leveraging weighted finite-population Bayesian bootstrapping, we can compare coverage with respect to estimates of the complete population constructed from the PSID sample. Finally, by incorporating prior information, we can provide guidance on decision-making for selecting the privacy loss budgets and optimally allocating a total privacy loss to each statistic needed.

Second, we want to consider more complex privacy and utility trade-offs between pure DP and relaxations of DP where not all statistics are subject to the same DP protections. While DP theoretically forbids using data-dependent hyperparameters without DP mechanisms, many commonly used DP algorithms and analyses do not adhere to this rule [1, 18], necessarily yielding additional privacy vulnerabilities in practice [16, 15]. It could be the case that certain hyperparameters could substantially improve the end-to-end usefulness of our estimators at a modest expense to privacy risk, but this would require a much more extensive and nuanced privacy analysis than that offered by a naive comparison of privacy loss budgets. Still, such an analysis could be helpful to understand where DP itself fundamentally limits the kinds of statistical validity offered in survey settings, where worst-case data generating scenarios may be highly unrealistic in practice.

References

[1] Martin Abadi, Andy Chu, Ian Goodfellow, H Brendan McMahan, Ilya Mironov, Kunal Talwar, and Li Zhang. Deep learning with differential privacy. In *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security*, pages 308–318, 2016.

[2] Jean-François Beaumont. A new approach to weighting and inference in sample surveys. *Biometrika*, 95(3):539–553, 2008. Publisher: Oxford University Press.

[3] Kenneth A Bollen, Paul P Biemer, Alan F Karr, Stephen Tueller, and Marcus E Berzofsky. Are survey weights needed? A review of diagnostic tests in regression analysis. *Annual Review of Statistics and Its Application*, 3:375–392, 2016. Publisher: Annual Reviews.

[4] Mark Bun, Jörg Drechsler, Marco Gaboardi, Audra McMillan, and Jayshree Sarathy. Controlling privacy loss in sampling schemes: An analysis of stratified and cluster sampling. In *3rd Symposium on Foundations of Responsible Computing (FORC 2022)*, 2022.

[5] Mark Bun and Thomas Steinke. Concentrated differential privacy: Simplifications, extensions, and lower bounds. In *Theory of Cryptography Conference*, pages 635–658. Springer, 2016.

[6] Jörg Drechsler. Differential Privacy for Government Agencies—Are We There Yet? *Journal of the American Statistical Association*, 118(541):761–773, 2023. Publisher: Taylor & Francis.

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

[8] Cynthia Dwork, Aaron Roth, and others. The algorithmic foundations of differential privacy. *Found. Trends Theor. Comput. Sci.*, 9(3-4):211–407, 2014.

[9] Andrew Gelman. Struggles with Survey Weighting and Regression Modeling. *Statistical Science*, 22(2):153–164, 2007.

[10] Jingchen Hu, JÖrg Drechsler, and Hang J Kim. Accuracy Gains from Privacy Amplification Through Sampling for Differential Privacy. *Journal of Survey Statistics and Methodology*, 10(3):688–719, 2022. Publisher: Oxford University Press.

[11] Jingchen Hu, Terrance D Savitsky, and Matthew R Williams. Private tabular survey data products through synthetic microdata generation. *Journal of Survey Statistics and Methodology*, 10(3):720–752, 2022. Publisher: Oxford University Press.

[12] Gautam Kamath, Argyris Mouzakis, Matthew Regehr, Vikrant Singhal, Thomas Steinke, and Jonathan Ullman. A Bias-Variance-Privacy Trilemma for Statistical Estimation. *arXiv preprint arXiv:2301.13334*, 2023.

[13] Shurong Lin, Mark Bun, Marco Gaboardi, Eric D Kolaczyk, and Adam Smith. Differentially Private Confidence Intervals for Proportions under Stratified Random Sampling. *arXiv preprint arXiv:2301.08324*, 2023.

[14] Frank McSherry and Kunal Talwar. Mechanism design via differential privacy. In *48th Annual IEEE Symposium on Foundations of Computer Science (FOCS'07)*, pages 94–103. IEEE, 2007.

[15] Shubhankar Mohapatra, Sajin Sasy, Xi He, Gautam Kamath, and Om Thakkar. The role of adaptive optimizers for honest private hyperparameter selection. In *Proceedings of the aaai conference on artificial intelligence*, volume 36, pages 7806–7813, 2022. Issue: 7.

[16] Nicolas Papernot and Thomas Steinke. Hyperparameter tuning with renyi differential privacy. *arXiv preprint arXiv:2110.03620*, 2021.

[17] Survey Research Center at the Institute for Social Research, University of Michigan Ann Arbor. Panel study of income dynamics, public use dataset, 2019.

[18] Florian Tramer and Dan Boneh. Adversarial training and robustness for multiple perturbations. *Advances in neural information processing systems*, 32, 2019.

A Figures

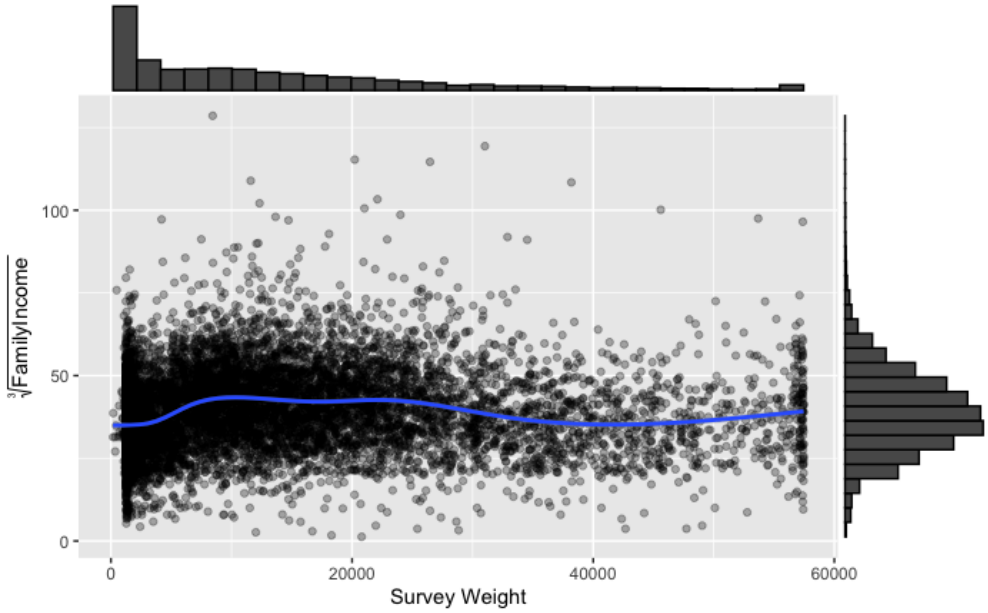


Figure 1: PSID cube-root family income by survey weight, with LOESS smoothing (blue line) and marginal histograms above the axes.

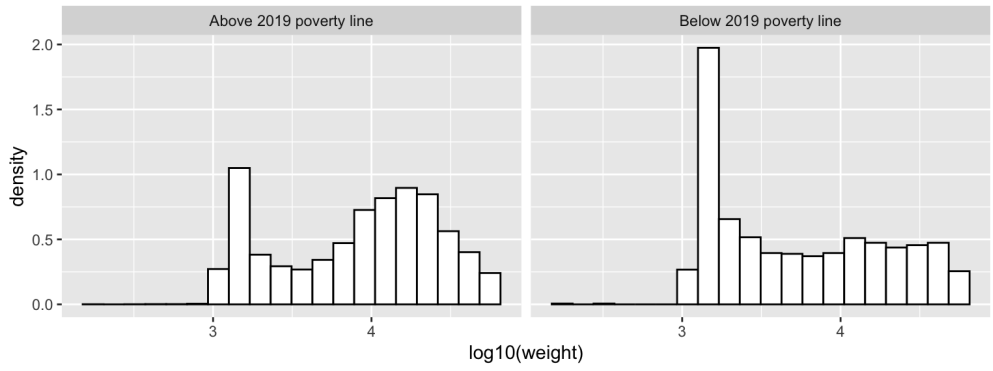


Figure 2: Survey weight histograms for PSID respondent families below and above the 2019 poverty line.

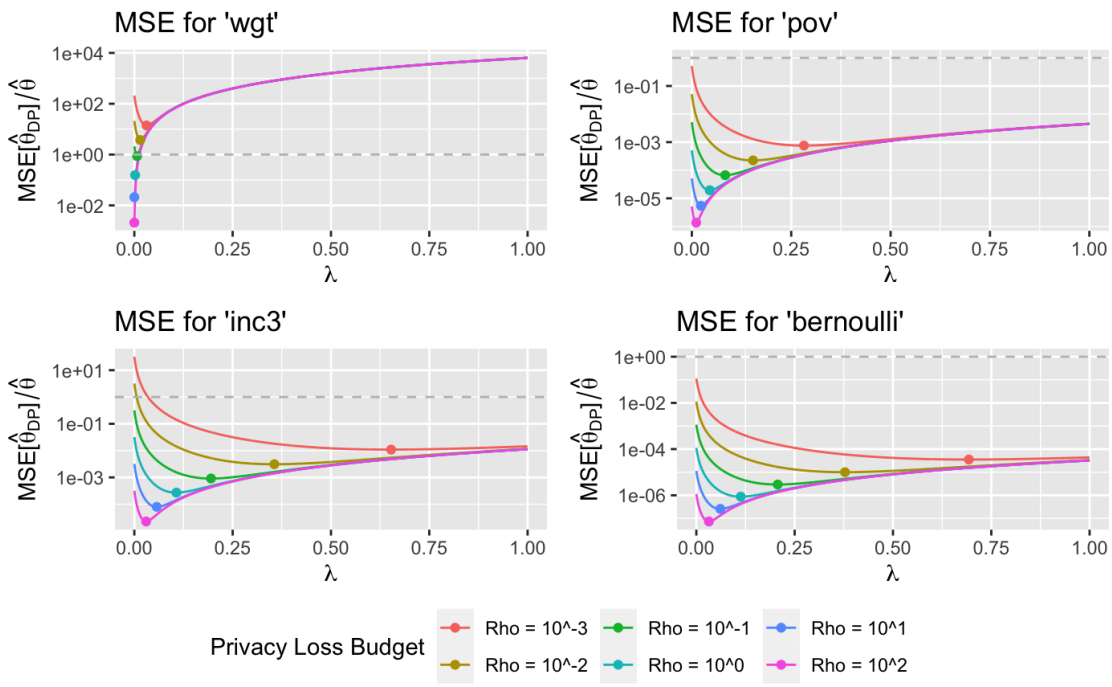


Figure 3: Bias-variance-privacy curves with noise-to-signal ratio on y-axis vs. λ on x-axis for different values of ρ_2 and response variables (subplot titles), ordered from most bias due to correlation between survey weights and responses (top left subplot) to least (bottom right subplot).

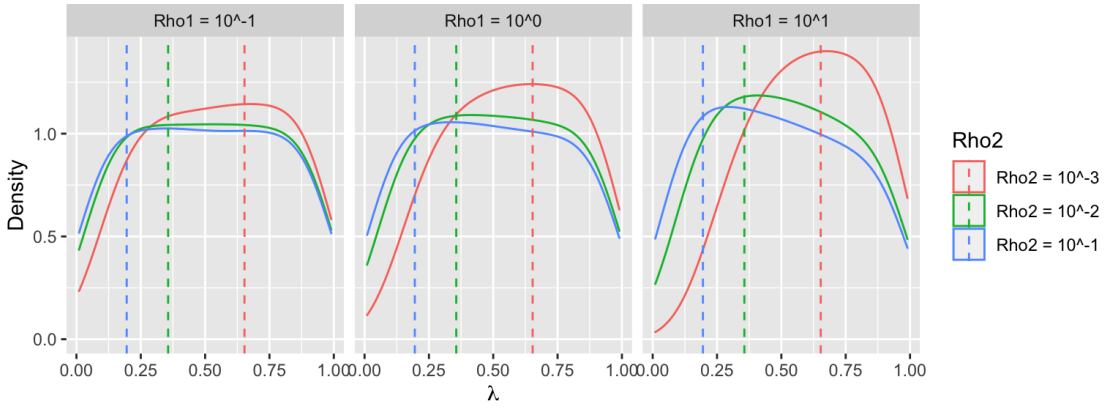


Figure 4: Sampling distributions of $\hat{\lambda}$ for `inc3` at different values of ρ_1 and ρ_2, with λ_{opt} shown in the vertical dashed lines for each value of ρ_2.

B Method derivations

Recall that we consider a response variable and survey weights $\{(y_i, w_i)\}_{i=1}^n$ lying within bounded intervals $[L_Y, U_Y] \times [L_W, U_W]$, and our goal is to estimate a weighted mean and variance estimates with respect to the inverse-proportional sampling variability, where $w_i \propto \mathbb{P}(i \text{ in sample})^{-1}$. Note that we consider $\sum_{i=1}^n w_i$ to be public information, since survey weights can be arbitrarily scaled by a constant factor, which is often the population size, the sample size, or 1. We will use zero-concentrated DP [5], i.e., for two samples $\{(y_i, w_i)\}_{i=1}^n \sim \{(y_i', w_i')\}_{i=1}^n$ which differ on a single entry and a randomized algorithm of outputs M, we have, for all $\alpha \in (1, \infty)$,

$$d_\alpha \left(M(\{(y_i, w_i)\}_{i=1}^n) || M(\{(y_i', w_i')\}_{i=1}^n) \right) \leq \rho\alpha, \tag{12}$$

where $d_\alpha(\cdot, \cdot)$ is the α-Rényi divergence. Our goal is to use ρ-zCDP to estimate

$$\hat{\theta}_0 \triangleq \frac{\sum_{i=1}^n w_i y_i}{\sum_{i=1}^n w_i}, \qquad \hat{\mathrm{Var}}(\hat{\theta}_0) \triangleq \frac{\sum_{i=1}^n w_i (y_i - \hat{\theta}_\lambda)^2}{\left(\sum_{i=1}^n w_i\right)^2}, \tag{13}$$

the original weighted mean and variance. Note that we will use $\boldsymbol{w}$ and $\boldsymbol{y}$ to refer to the vectors of weights and responses, respectively; also note that, when appropriate, we will treat functions of the constants listed above as constants themselves.

Theorem 1. *Algorithm 1 satisfies $(\rho_1 + \rho_2 + \rho_3)$-zCDP.*

Proof. The proof outline is as follows: first, we show sampling from Equation 7 satisfies ρ_1-zCDP. Second, we show that sampling from Equations 8 and 9 satisfy ρ_2-zCDP and ρ_3-zCDP, respectively. Let these release mechanisms be defined as randomized algorithms $M_1(\boldsymbol{y}, \boldsymbol{w}), M_2(M_1(\boldsymbol{y}, \boldsymbol{w}), \boldsymbol{y}, \boldsymbol{w})$, and $M_3(M_1(\boldsymbol{y}, \boldsymbol{w}), \boldsymbol{y}, \boldsymbol{w})$. By adaptive composition, releasing

$$\left(\hat{\lambda}, \hat{\theta}_{\mathrm{DP}}, \hat{\mathrm{Var}}(\hat{\theta})_{\mathrm{DP}}\right) = (M_1(\boldsymbol{y}, \boldsymbol{w}), M_2(M_1(\boldsymbol{y}, \boldsymbol{w}), \boldsymbol{y}, \boldsymbol{w}), M_3(M_1(\boldsymbol{y}, \boldsymbol{w}), \boldsymbol{y}, \boldsymbol{w})) \tag{14}$$

satisfies (ρ_1, ρ_2, ρ_3)-zCDP. It remains to show that each individual component satisfies zCDP.

First, we demonstrate that sampling $\hat{\lambda}$ from Equation 7 satisfies $\sqrt{2\rho}$-DP [7], and hence ρ-zCDP. To do this, we first calculate the effect of λ on the downstream estimation of $\hat{\theta}_0$. Without survey weight regularization, assuming without loss of generality that adjacent datasets differ on the first record (y_1, w_1),

$$\begin{aligned}
\Delta_0 &\triangleq \sup_{(\boldsymbol{y}, \boldsymbol{w}) \sim_M (\boldsymbol{y}', \boldsymbol{w}')} \left| \hat{\theta}_0(\boldsymbol{y}, \boldsymbol{w}) - \hat{\theta}_0(\boldsymbol{y}', \boldsymbol{w}') \right| \\
&= \sup_{(\boldsymbol{y}, \boldsymbol{w}) \sim_M (\boldsymbol{y}', \boldsymbol{w}')} \left| \frac{\sum_{i=1}^n y_i w_i}{\sum_{i=1}^n w_i} - \frac{\sum_{i=1}^n y_i' w_i'}{\sum_{i=1}^n w_i'} \right| \\
&= \sup_{(\boldsymbol{y}, \boldsymbol{w}) \sim_M (\boldsymbol{y}', \boldsymbol{w}')} \left| \frac{w_1 y_1 + \sum_{i=2}^n y_i w_i}{w_1 + \sum_{i=2}^n w_i} - \frac{w_1' y_1' + \sum_{i=2}^n y_i w_i}{w_1' + \sum_{i=2}^n w_i} \right| \\
&\leq \left| \frac{U_W U_Y + \sum_{i=2}^n L_W L_Y}{U_W + \sum_{i=2}^n L_W} - \frac{L_W L_Y + \sum_{i=2}^n L_Y W_Y}{L_W + \sum_{i=2}^n L_W} \right| \\
&\leq \left| \frac{U_W U_Y + \sum_{i=2}^n L_W L_Y}{L_W + \sum_{i=2}^n L_W} - \frac{L_W L_Y + \sum_{i=2}^n L_Y W_Y}{L_W + \sum_{i=2}^n L_W} \right| \\
&\leq \frac{1}{n L_W} \left| U_W U_Y - L_W L_Y \right| \\
&= \frac{1}{n L_W} \left| U_W U_Y - U_W L_Y + U_W L_Y - L_W L_Y \right| \\
&\leq \frac{1}{n L_W} \left[U_W |U_Y - L_Y| + L_Y |U_W - L_W| \right] \\
&= \frac{1}{n L_W} \left[U_W \Delta_Y + L_Y \Delta_W \right]
\end{aligned} \tag{15}$$

By instead replacing $L_W \mapsto G_\lambda(L_W)$ and $U_W \mapsto G_\lambda(U_W)$, we can alternatively consider the sensitivity of the estimate derived from regularized weights, i.e.

$$\begin{aligned}\Delta_\lambda &\triangleq \frac{1}{nG_\lambda(L_W)}[G_\lambda(U_W)\Delta_Y + L_Y(G_\lambda(U_W) - G_\lambda(L_W))] \\ &= \frac{1}{nG_\lambda(L_W)}\left[G_\lambda(U_W)\Delta_Y + L_Y(1-\lambda)\Delta_W\right] \\ &= \frac{1}{n((1-\lambda)L_W + \lambda\overline{w})}\left[((1-\lambda)U_W + \lambda\overline{w})\Delta_Y + L_Y(1-\lambda)\Delta_W\right].\end{aligned} \tag{16}$$

Using $\hat{\theta}_\lambda$ introduces a bias of the form

$$\begin{aligned}\text{Bias}(\lambda) &\triangleq \hat{\theta}_0(\boldsymbol{y}, G_\lambda(\boldsymbol{w})) - \hat{\theta}_0(\boldsymbol{y}, \boldsymbol{w}) \\ &\triangleq \frac{\triangleq \sum_{i=1}^n((1-\lambda)w_i + \lambda\overline{w})y_i}{\sum_{i=1}^n(1-\lambda)w_i + \lambda\overline{w}} - \frac{\sum_{i=1}^n w_i y_i}{\sum_{i=1}^n w_i} \\ &= \frac{1}{n\overline{w}}\left[(1-\lambda)\sum_{i=1}^n w_i y_i - \lambda\overline{w}\sum_{i=1}^n y_i - \sum_{i=1}^n w_i y_i\right] \\ &= \lambda\left[\frac{1}{n}\sum_{i=1}^n y_i - \frac{\sum_{i=1}^n w_i y_i}{\sum_{i=1}^n w_i}\right] \\ &= \lambda(\hat{\theta}_1 - \hat{\theta}_0).\end{aligned} \tag{17}$$

However, the reduction in sensitivity reduces the variance necessary to use the Gaussian mechanism at privacy loss ρ_2. Were we to apply this Gaussian mechanism to $\hat{\theta}_\lambda$, we would now sample from the distribution

$$\hat{\theta}_{\text{DP}} \sim N\left(\hat{\theta}_\lambda, \frac{\Delta_\lambda^2}{2\rho_2}\right), \tag{18}$$

yielding a new theoretical mean-square error of the form

$$\text{MSE}(\hat{\theta}_{\text{DP}}) \triangleq \mathbb{E}\left[(\hat{\theta}_{\text{DP}} - \hat{\theta})^2\right] = \lambda^2\left(\hat{\theta}_1 - \hat{\theta}_0\right)^2 + \frac{\Delta_\lambda^2}{2\rho_2} \tag{19}$$

Since the optimal λ which minimizes Equation 19 is data-dependent, we implement the exponential mechanism [14] to sample an approximation $\hat{\lambda}$ to the optimal value with DP. Specifically, we define

$$\ell(\lambda; \boldsymbol{y}, \boldsymbol{w}) \triangleq \text{MSE}(\hat{\theta}_{\text{DP}}). \tag{20}$$

This gives us a sensitivity

$$\Delta_\ell \triangleq \sup_{(\boldsymbol{y},\boldsymbol{w})\sim_M(\boldsymbol{y}',\boldsymbol{w}')} \sup_{\lambda\in[0,1]} |\ell(\lambda; \boldsymbol{y}, \boldsymbol{w}) - \ell(\lambda; \boldsymbol{y}', \boldsymbol{w}')| \tag{21}$$

$$= \sup_{(\boldsymbol{y},\boldsymbol{w})\sim_M(\boldsymbol{y}',\boldsymbol{w}')} \sup_{\lambda\in[0,1]} \left|\lambda^2\left((\hat{\theta}_1 - \hat{\theta}_0)^2 - (\hat{\theta}_1' - \hat{\theta}_0')^2)\right)\right| \tag{22}$$

$$= \sup_{(\boldsymbol{y},\boldsymbol{w})\sim_M(\boldsymbol{y}',\boldsymbol{w}')} \left|(\hat{\theta}_1 - \hat{\theta}_0)^2 - (\hat{\theta}_1' - \hat{\theta}_0')^2\right| \tag{23}$$

$$\leq \left|\left(\frac{U_Y + (n-1)L_Y}{n} - \frac{U_W U_Y + \sum_{i=2}^n L_W L_Y}{nL_W}\right)^2\right| \tag{24}$$

$$\leq \left|\left(\frac{U_Y L_W + (n-1)L_Y L_W}{nL_W} - \frac{U_W U_Y + \sum_{i=2}^n L_W L_Y}{nL_W}\right)^2\right| \tag{25}$$

$$= \left(\frac{U_Y \Delta_W}{nL_W}\right)^2 \tag{26}$$

By construction, the exponential mechanism density takes the form, with uniform base measure over $\lambda \in [0,1]$,

$$\begin{aligned} f(\lambda) &\propto \mathbb{1}_{\{\lambda\in[0,1]\}} \exp\left(-\frac{\epsilon \ell(\lambda; \boldsymbol{y}, \boldsymbol{w})}{2\Delta_\ell}\right) \\ &\propto \mathbb{1}_{\{\lambda\in[0,1]\}} \exp\left(-\frac{\epsilon}{2\left(\frac{U_Y \Delta_W}{n L_W}\right)^2}\left[\lambda^2(\hat{\theta}_1 - \hat{\theta}_0)^2 + \frac{\Delta_\lambda^2}{2\rho_2}\right]\right) \end{aligned} \tag{27}$$

By replacing $\epsilon \mapsto \sqrt{2\rho_1}$ and using the fact that any ϵ-DP algorithm is $\frac{\epsilon^2}{2}$-zCDP, we yield the first desired result.

Next, using $\hat{\lambda}$ satisfying ρ_1-zCDP, we can then sample $\hat{\theta}_{\mathrm{DP}}$ using

$$\hat{\theta}_{\mathrm{DP}} \sim N\left(\hat{\theta}_{\hat{\lambda}}, \frac{\Delta_{\hat{\lambda}}^2}{2\rho_2}\right), \tag{28}$$

which after substitution, agrees with Equation 8.

Finally, we perform a similar analysis for the $\hat{\mathrm{Var}}(\hat{\theta})_{\mathrm{DP}}$, here reusing $\hat{\lambda}$ from above. Note that while we could sample a new $\hat{\lambda}$ which optimizes the bias-variance-privacy trade-off specifically for $\hat{\mathrm{Var}}(\hat{\theta})_{\mathrm{DP}}$, we instead reuse $\hat{\lambda}$ for two reasons. First, we do not want to incur additional privacy loss costs. Second, because $\hat{\theta}_{\mathrm{DP}}$ is less sensitive than $\hat{\mathrm{Var}}(\hat{\theta})_{\mathrm{DP}}$, we prefer the original $\hat{\lambda}$ so we have a higher probability of selecting smaller $\hat{\lambda}$ to prevent underestimating $\hat{\mathrm{Var}}(\hat{\theta})_{\mathrm{DP}}$. We show this in the original weighted setting, where

$$\begin{aligned} \Delta_{\hat{\mathrm{Var}}(\theta)} &\triangleq \sup_{(\boldsymbol{y},\boldsymbol{w})\sim_M(\boldsymbol{y}',\boldsymbol{w}')} \left|\hat{\mathrm{Var}}(\theta)(\boldsymbol{y},\boldsymbol{w}) - \hat{\mathrm{Var}}(\theta)(\boldsymbol{y}',\boldsymbol{w}')\right| \\ &= \sup_{(\boldsymbol{y},\boldsymbol{w})\sim_M(\boldsymbol{y}',\boldsymbol{w}')} \left|\frac{\sum_{i=1}^n w_i^2(y_i-\hat{\theta}_0)^2}{\left(\sum_{i=1}^n w_i\right)^2} - \frac{\sum_{i=1}^n w_i^{2'}(y_i'-\hat{\theta}_0')^2}{\left(\sum_{i=1}^n w_i'\right)^2}\right| \\ &= \sup_{(\boldsymbol{y},\boldsymbol{w})\sim_M(\boldsymbol{y}',\boldsymbol{w}')} \left|\frac{w_1^2(y_1-\hat{\theta}_0)^2 + \sum_{i=2}^n w_i^2(y_i-\hat{\theta}_0)^2}{w_1^2 + \left(\sum_{i=1}^n w_i\right)^2 + w_1\sum_{i=2}^n w_i} - \frac{w_1^{2'}(y_1'-\hat{\theta}_0')^2 + \sum_{i=2}^n w_i^2(y_i-\hat{\theta}_0')^2}{w_1^{2'} + \left(\sum_{i=1}^n w_i\right)^2 + w_1'\sum_{i=2}^n w_i}\right| \\ &\le \frac{1}{n^2L_W^2}\left|\left(w_1^2(y_1-\hat{\theta}_0)^2 - w_1^{2'}(y_1'-\hat{\theta}_0')^2\right) + \sum_{i=2}^n w_i^2\left((y_i-\hat{\theta}_0)^2 - (y_i-\hat{\theta}_0')^2\right)\right| \\ &\le \frac{1}{n^2L_W^2}\left[\left|\left(w_1^2(y_1-\hat{\theta}_0)^2 - w_1^{2'}(y_1'-\hat{\theta}_0')^2\right)\right| + \left|\sum_{i=2}^n w_i^2\left((y_i-\hat{\theta}_0)^2 - (y_i-\hat{\theta}_0')^2\right)\right|\right] \\ &\le \frac{1}{n^2L_W^2}\left[U_W^2(U_Y-L_Y)^2 + U_W^2\sum_{i=2}^n\left|2y_i - (\hat{\theta}_0+\hat{\theta}_0')\right|\left|\hat{\theta}_0-\hat{\theta}_0'\right|\right] \\ &\le \frac{U_W^2}{n^2L_W^2}\left|\Delta_Y^2 + (n-1)2\Delta_Y\Delta_0\right|. \end{aligned} \tag{29}$$

Substituting $U_W \mapsto G_{\hat{\lambda}}(U_W)$, $L_W \mapsto G_{\hat{\lambda}}(L_W)$, and $\Delta_0 \mapsto \Delta_{\hat{\lambda}}$ yields Equation 9, completing the proof. □