Statistically Valid Inferences from Privacy Protected Data: Supplementary (Online) Appendices*

Georgina Evans† Gary King‡
Margaret Schwenzfeier§ Abhradeep Thakurta¶

June 12, 2022

*The current version of the paper and this Supplementary Appendix are available at GaryKing.org/dp.

†Ph.D. Candidate, Department of Government, Harvard University, 1737 Cambridge Street Cambridge, MA 02138; Georgina-Evans.com, GeorginaEvans@g.harvard.edu.

‡Albert J. Weatherhead III University Professor, Institute for Quantitative Social Science, 1737 Cambridge Street, Harvard University, Cambridge MA 02138; GaryKing.org, King@Harvard.edu.

§Ph.D. Candidate, Department of Government, Harvard University, 1737 Cambridge Street Cambridge, MA 02138; MegSchwenzfeier.com, schwenzfeier@g.harvard.edu.

¶Assistant Professor, Department of Computer Science, University of California Santa Cruz, bit.ly/AbhradeepThakurta, aguhatha@ucsc.edu.

Contents

Appendix A Bounds for Differential Privacy

First write Equation 2 in the text as $\mathcal{N}(t \mid \hat{\theta}, S^2) \leq \delta + e^{\epsilon} \cdot \mathcal{N}(t \mid \hat{\theta} + \Delta, S^2)$, for any t, where Δ is the "sensitivity" of this estimator (the largest change in $M(\text{mean}, D)$ over all possible pairs of datasets that differ by at most one row), where $|\hat{\theta}_D - \hat{\theta}_{D'}| \leq \Delta$ such that $\hat{\theta}_D$ and $\hat{\theta}_{D'}$ denote the estimator computed from D and D', respectively. The censored mean $\hat{\theta}$ has sensitivity $\Delta = 2\Lambda/N$.

A simple expression for S that satisfies the differential privacy standard is

$$S \equiv S(\Lambda, \epsilon, \delta, N) = \frac{\Delta\sqrt{2\ln(1.25/\delta)}}{\epsilon} = \frac{2\Lambda\sqrt{2\ln(1.25/\delta)}}{N\epsilon}, \tag{1}$$

Equation 3 in the text is a special case of this expression with $\delta = 0.0005$.

Note that Equation 1 holds only for $\epsilon \leq 1$ (Dwork and Roth, 2014). In practice, we use the tighter numerical solution by Balle and Wang (2018), which allows 20-30% smaller values of S when $\epsilon \leq 1$ and is still valid for larger values. To summarize: for the Gaussian mechanism, write Equation 2 in the text in terms of the cumulative standard normal density: $\Phi\left(\frac{\Delta}{2\sigma} - \frac{\epsilon\sigma}{\Delta}\right) - e^{\epsilon}\Phi\left(-\frac{\Delta}{2\sigma} - \frac{\epsilon\sigma}{\Delta}\right) \leq \delta$. Then set S to the minimum σ that satisfies this inequality. This numerical calculation exactly calibrates the noise to a given privacy budget and hence minimizes S for a given level of $\{\epsilon, \delta\}$.

An alternative option is to use Zero Concentrated Differential Privacy (zCDP) which is known to have tighter composition properties, therefore allowing for less noise across multiple queries. This is due to the fact that the algorithm is primarily based on the Gaussian mechanism. The fact our algorithm satisfies zCDP also implies that the δ we provide in the paper does not correspond to the probability of "catastrophic failure". A version of the Gaussian mechanism satisfying zCDP is given in Bun and Steinke (2016). This article also demonstrates that zCDP implies approximate differential privacy and shows how to convert to the zCDP parameter, which is immediately computable based on the standard deviation of the Gaussian noise in our paper. (For context, the US Census Bureau also routinely reports the zCDP parameter for the Gaussian mechanism, also for its desirable composition properties.)

Appendix B Privatized Versions of Classical Uncertainty Estimates Are Not Valid for Privatized Estimates

In Section 2.3, we discuss the need for uncertainty estimates of differentially private statistics. We show here that using the same algorithm to privatize and disclose the classical variance estimate is not a solution to the problem.

In a system without differential privacy, let $\hat{\theta}$ be a point estimate of a quantity of interest θ in an unobserved population. Denote by $V(\hat{\theta})$ the (true) variance of $\hat{\theta}$ over repeated (hypothetical, unobserved) samples of datasets drawn with the same data generation process. Consider now a differentially private point estimator (or "mechanism") $\hat{\theta}^{\text{dp}}$ of θ. Although it would be easy to disclose a differentially private estimate of the variance, $\hat{V}(\hat{\theta})^{\text{dp}}$, it is of no direct use, since $\hat{\theta}$ is never disclosed and so an estimate of *its* variance is irrelevant. Indeed, $\hat{V}(\hat{\theta})^{\text{dp}}$ is a biased estimator of the quantity we need, $V(\hat{\theta}^{\text{dp}})$. Since we will show below that $\hat{\theta}^{\text{dp}}$ itself is biased, $V(\hat{\theta}^{\text{dp}})$ would be of no direct use even if it were known.

Appendix C Comparison to Smith (2011)

Our method achieves approximate unbiasedness in the presence of censoring by applying a bias correction procedure. Censoring then is advantageous because less noise is required and no penalty is paid in terms of bias. An alternative approach to obtaining unbiased estimates, which also uses the sample and aggregate approach, avoids censoring altogether by selecting a censoring value Λ outside the data (with high probability) in a differentially private way. Along these lines, Smith (2011) proposes the "Widened Winsorized Mean" algorithm which, after constructing P partition estimates of the quantity of interest, returns estimates of the 0.25 and 0.75 quantiles of the P partition estimates via the Exponential mechanism. The censoring value, Λ, is then constructed by widening the estimated quantiles so as to avoid censoring any of the partitions with high probability. Noise is then added to the aggregated partition estimates, where the variance scales in the censoring bound in order to satisfy differential privacy.

However, as we now demonstrate, the approach in Smith (2011) requires adding far more noise than our method because Λ must be set so large. Our method benefits from censoring because Λ can be set smaller. The result is that our approach applied to finite datasets can have dramatically more efficient inferences. (We can also marshal in our approach the portion of the algorithm in Smith 2011 used for discovering the value of Λ at which censoring would begin. When less information is available from prior research than usual, we use this algorithm but set Λ to a smaller value and correct for the censoring, resulting in less noise and more efficiency.)

To demonstrate this point, consider a simple data generating process where we draw 1000 i.i.d. samples of $n = 10000$ for each value of ϵ from Poisson(5). Our quantity of interest is the population mean, which we estimate in $P = \sqrt{n}$ partitions (as per the advice in Smith (2011)) by the sample mean. To ensure we calibrate both methods with an identical privacy budget, we estimate Λ from the data via the Exponential mechanism in both ('Private Quantile Estimation'), allocating an equal privacy budget to this step for both methods. For our procedure, our target Λ censors about 40% of the partitions with the privacy budget divided equally between the partition mean and the private quantiles.

The results are shown in Figure 1, where we compare the Root Mean Square Error (RMSE) of our estimate ('bias adjusted') with Smith (2011), as the privacy budget, ϵ, varies. The difference is so large that we had to plot the RMSE on the log scale. The figure demonstrates that, at all privacy budgets given, our estimator has substantially lower RMSE. At the lowest privacy budget, $\epsilon = 0.5$, Smith's estimator has a RMSE that is 428 times higher than our estimator.

Appendix D Estimates from Partitions

In Section 3.1, algorithm Step 1a, we compute $\hat{\theta}_p$ from the data in partition p, D_p, via one of two options: (1) Use the same statistical method we would have used if we were able to analyze the entire private dataset. If we make this choice, we must be careful to appropriately scale up this result to the entire dataset when required (a version of "subsampling"; see Politis, Romano, and Wolf 1999). Alternatively, (2) use the general purpose "bag of

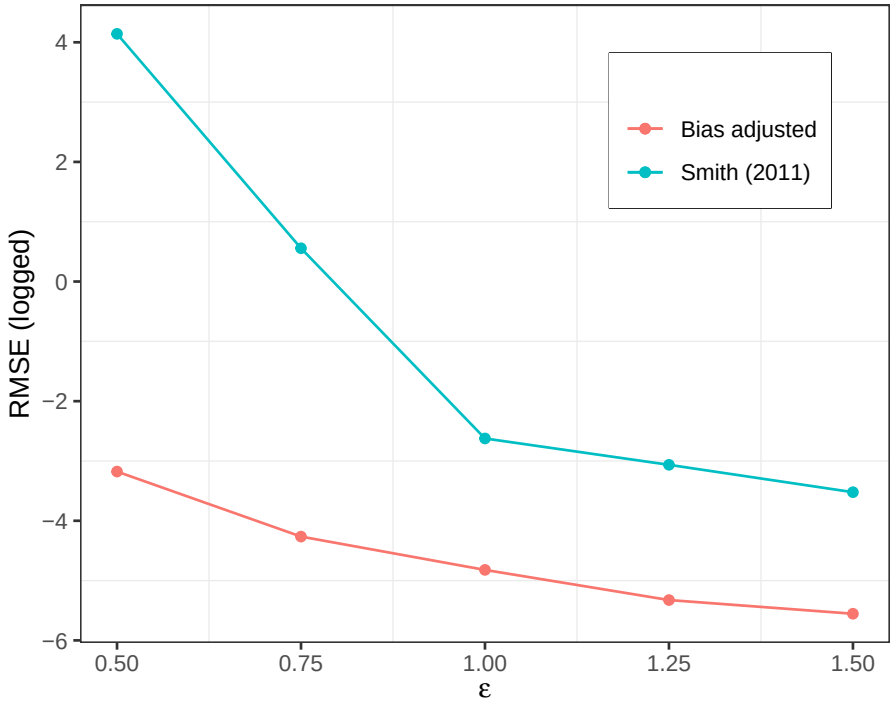

Figure 1: Root Mean Square Error Comparison of our bias adjusted approach to Smith (2011), with the vertical axis on the log scale

little bootstraps" algorithm (Kleiner et al., 2014) so scaling up is automatic.

To implement this optional bag of little bootstraps in partition p, first repeat these steps B times:

1. Simulate bootstrap b (i.e., $b = 1, \ldots, B$) by sampling one weight for each of the n units in partition p as $w_b \equiv \{w_{1,b}, \ldots, w_{n,b}\} \sim \text{Multinomial}(N, \mathbf{1}_n/n)$.

2. Calculate a statistic (an estimate of population value θ) from bootstrapped sample b in partition p: $\hat{\theta}_{p,b} = s(D_p, w_b)$, such as a predicted value, expected value, or classification.

Then summarize the set of B bootstrapped estimates within each partition with a (still unobserved) estimator, which we write generically as $\hat{\theta}_p$. Examples include a mean $\hat{\theta}_p = m(\hat{\theta}_{p,b})$ or the probability of the Democrat winning a majority of the vote, $\hat{\theta}_p = m[\mathbb{1}(\hat{\theta}_{p,b} > 0.5)]$.

Under option (1), the necessary scale factors the researcher would need to derive differ depending on the type of estimator and quantity of interest; some quantities, like the

variance, are a linear function of N, but each partition's estimate is a function of n and so need to be scaled up by a factor of $(N-n)/n$.

The bag of little bootstraps in option (2) is computationally more intensive but allows simpler and more flexible estimation strategies for some quantities, such as the probability that a causal effect is greater than zero, estimated by merely counting the proportion of positive bootstrap estimates that meet selected criteria. It is also far less computationally demanding than the standard bootstrap procedures.

Appendix E Bounding Error in The Expected Value

Let $Z_m \sim \mathcal{N}(\theta, \sigma^2/n)$. Our goal is to show that $\mathbb{E}\left[\hat{\theta}_n^{dp}\right] = \mathbb{E}[c(Z_n, \Lambda)] \pm O(1/\sqrt{n})$, where $n \equiv N/P$ and $c(x,T)$ is a function that censors an input x at $[-T,T]$. We will assume that $\hat{\theta}$ is an asymptotically normal statistic and that $\mathbb{E}[|\hat{\theta}|^3) = \rho < \infty$ (i.e., the estimate has a finite third moment).

We start by writing the expectation as:

$$\mathbb{E}\left[\hat{\theta}_n^{dp}\right] = \mathbb{E}[c(\hat{\theta}_n, \Lambda)] + \underbrace{\mathbb{E}[V]}_{\text{DP noise}} = \mathbb{E}[c(\hat{\theta}_n^p, \Lambda)]$$

Since $\mathbb{E}[V] = 0$. Let $F_n(x) = \Pr(\hat{\theta}_n^p \leq x)$ and $p_n(x) = F_n'(x)$. Therefore:

$$\mathbb{E}\left[\hat{\theta}_n^{dp}\right] = \int_{-\infty}^{-\Lambda} -\Lambda p_n(x)dx + \int_{-\Lambda}^{\Lambda} x p_n(x)dx + \int_{\Lambda}^{\infty} \Lambda p_n(x)dx$$

$$= -\Lambda F_n(-\Lambda) + \int_{-\Lambda}^{\Lambda} x p_n(x)dx + \Lambda(1 - F_n(\Lambda)) \tag{2}$$

Focusing on the second term, we have that:

$$\int_{-\Lambda}^{\Lambda} x p_n(x)dx = \left[x \int p_n(x)dx\right]_{-\Lambda}^{\Lambda} - \int_{-\Lambda}^{\Lambda} \int p_n(x)dxdx$$

$$= \Lambda F_n(\Lambda) + \Lambda F_n(-\Lambda) - \int_{-\Lambda}^{\Lambda} F_n(x)dx \tag{3}$$

Substituting Eqn. (3) into Eqn. (2), we have that

$$\mathbb{E}\left[\hat{\theta}_n^{dp}\right] = -\Lambda F_n(-\Lambda) + \Lambda F_n(-\Lambda) - \int_{-\Lambda}^{\Lambda} F_n(x)dx + \Lambda(1 - F_n(\Lambda)) + \Lambda F_n(\Lambda)$$

$$= \int_{-\Lambda}^{\Lambda} 1/2 - F_n(x)dx \tag{4}$$

By a similar logic it follows that $\mathbb{E}[c(Z_n, \Lambda)] = \int_{-\Lambda}^{\Lambda} 1/2 - F_n^*(x)dx$, where $F_n^* = \Pr(Z_n \leq x)$. Then, we apply the Berry-Esseen theorem (Feller, 1971, Section 16.5) to Eqn. (4), and denote by C a positive constant, which yields:

$$\int_{-\Lambda}^{\Lambda} 1/2 - F_n(x)dx = \int_{-\Lambda}^{\Lambda} 1/2 - F_n^*(x) \pm \frac{C\rho}{\sigma^3\sqrt{n}}dx$$

$$= \mathbb{E}[c(Z_n, \Lambda)] \pm \frac{2\Lambda C\rho}{\sigma^3\sqrt{n}}$$

Recognizing the alternative formulation of $\mathbb{E}[c(Z_n, \Lambda)]$, we obtain our desired result:

$$E\left(\hat{\theta}_n^{dp}\right) = -\alpha_1\Lambda + (1 - \alpha_2 - \alpha_1)\theta_T + \alpha_2\Lambda \pm O(1/\sqrt{n}), \tag{5}$$

where

$$\theta_T = \theta + \sigma/\sqrt{n} \cdot \left(\frac{\mathcal{N}(-\Lambda \mid \theta, \sigma^2/n) - \mathcal{N}(\Lambda \mid \theta, \sigma^2/n)}{1 - \alpha_2 - \alpha_1}\right).$$

Appendix F Software Design

We recommend data access systems that use our procedures allow a wide range of statistical methods and quantities of interest. Researchers should be able to choose almost any quantity to estimate and almost any statistical model. Given the limited privacy budget, researchers will want to choose which quantities to disclose selectively. For example, instead of logit coefficients, researchers would typically be more interested in reporting relative risks, probabilities, or risk differences (King, Tomz, and Wittenberg, 2000; King and Zeng, 2002). Even regression coefficients are often best replaced by quantities like a

predicted value, the probability a party's candidate wins the election, or a first difference. Software should allow researchers to submit statistical code to be checked and included, since the algorithm can wrap around any legitimate statistical procedure.

Designing the user interface to encourage best statistical practices can be valuable. This is especially so for users unfamiliar with differential privacy. One simple procedure is to provide a simulated dataset (without leaking any privacy from the real dataset) so users can compare the results from runs with and without privacy protections and get a feel for how to do data analysis within the framework.

For estimation, we note that α_2 is bounded to the unit interval, and so we could set $\Lambda_\alpha = 1$ without risk of censoring, but this would overstate this parameter's sensitivity by a factor of two (since we are paying from the privacy budget in anticipation of α being negative, which is impossible). Instead of resolving this issue by changing the expression for censoring and S, we do so more conveniently, without changing the notation (or code) above, by simply reparameterizing as $\beta = \alpha_2 - 0.5$, setting $\Lambda_\beta = 0.5$, estimating and disclosing β, and then solving to obtain our estimate of α_2 before using it to solve our three equations.

Useful approaches also exist for unusual situations where little information about Λ is available; see Appendix C and Liu and Talwar (2018).

Finally, under the topic of "do not try this at home," data providers should understand that a differentially private data access system involves details of implementation not covered here. These include random number generators, privacy budgets, parallelization, security, authentication, and authorization. They also involve avoiding side attacks on the timing of the algorithm, statistical methods that occasionally fail (e.g., due to collinearity in regression or, in logit, perfect discrimination), the privacy budget, and the state of the computer system (e.g., Garfinkel, Abowd, and Powazek, 2018; Haeberlen, Pierce, and Narayan, 2011).

Appendix G Variance Estimation Derivations

We derive our method of variance estimation outlined in Section 4.2. Our goal is to use the output from our point estimate algorithm to compute $\hat{V}(\tilde{\theta}^{\text{dp}})$. We do this without any additional tax on the privacy budget. Using notation (i) to denote the ith simulation, we write:

$$\hat{\theta}^{\text{dp}}(i), \hat{\alpha}_2^{\text{dp}}(i) \sim \mathcal{N}\left(\begin{bmatrix} \hat{\theta}^{\text{dp}} \\ \hat{\alpha}_2^{\text{dp}} \end{bmatrix}, \begin{bmatrix} \hat{V}(\hat{\theta}^{\text{dp}}) & \widehat{\text{Cov}}(\hat{\alpha}_2^{\text{dp}}, \hat{\theta}^{\text{dp}}) \\ \widehat{\text{Cov}}(\hat{\alpha}_2^{\text{dp}}, \hat{\theta}^{\text{dp}}) & \hat{V}(\hat{\alpha}_2^{\text{dp}}) \end{bmatrix} \right). \tag{6}$$

To implement this procedure we require intermediate quantities $\hat{V}(\hat{\theta}^{\text{dp}})$, $\hat{V}(\hat{\alpha}_2^{\text{dp}})$, and $\widehat{\text{Cov}}(\hat{\alpha}_2^{\text{dp}}, \hat{\theta}^{\text{dp}})$, which we show below can be written as functions of information already disclosed. We plug these into Equation 6 and repeatedly draw $\{\hat{\theta}^{\text{dp}}(i), \hat{\alpha}_2^{\text{dp}}(i)\}$, each time bias correcting via the procedure in Section 4.1 to compute $\tilde{\theta}^{\text{dp}}(i)$. Finally, we compute the sample variance over these simulations to yield our estimate $\hat{V}(\tilde{\theta}^{\text{dp}})$.

We decompose the two variance parameters using the results following Equation 7. The first we write as $\hat{V}(\hat{\theta}^{\text{dp}}) = \hat{V}(\hat{\theta}) + S^2_{\hat{\theta}}$, where $\hat{V}(\hat{\theta})$ is the variance of the mean over P draws from a normal censored at $[-\Lambda, \Lambda]$ (divided by P), and $S^2_{\hat{\theta}}$ is the variance of the differentially private noise. The distribution from which this variance is calculated then is a three component mixture (see Equation 9 in the paper). The first component is a truncated normal with mean θ_T, and bounds $[-\Lambda, \Lambda]$; the two other components are the spikes at Λ and $-\Lambda$. Begin with the following generic formula for the variance of the mean of draws from a 3-component mixture distribution with weights w_i, and component mean and variances of $E[\theta_i]$, σ_i^2 respectively:

$$V(\hat{\theta}) = \frac{1}{P} \cdot \left(\left[\sum_{i=1}^{3} w_i (E[\hat{\theta}_i]^2 + \sigma_i^2) \right] - E[\hat{\theta}]^2 \right) \tag{7}$$

with weights $\mathbf{w} = [(1 - \alpha_1 - \alpha_2), \alpha_2, \alpha_1]$, and with means for the spikes at $E[\hat{\theta}_2] = \Lambda$, and $E[\hat{\theta}_3] = -\Lambda$ and variances $\sigma_2^2 = \sigma_3^2 = 0$. Then, rearranging Equation 9, we write the truncated normal mean as

$$E[\hat{\theta}_1] \equiv \theta_T = \frac{E[\hat{\theta}] - \Lambda(\alpha_1 + \alpha_2)}{1 - \alpha_2 - \alpha_1}. \tag{8}$$

and we express the variance of the truncated normal as

$$\sigma_1^2 = \sigma^2 \left[1 + \frac{\left(\frac{-\Lambda-\theta}{\sigma}\right) Q_1 - \left(\frac{\Lambda-\theta}{\sigma}\right) Q_2}{1-\alpha_2-\alpha_1} - \left(\frac{Q_1 - Q_2}{1-\alpha_2-\alpha_1} \right)^2 \right] \tag{9}$$

where $Q_1 = \frac{1}{\sqrt{2\pi}} \text{exp}\left(-\frac{1}{2} \left(\frac{-\Lambda-\theta}{\sigma}\right)^2 \right)$ and $Q_2 = \frac{1}{\sqrt{2\pi}} \text{exp}\left(-\frac{1}{2} \left(\frac{\Lambda-\theta}{\sigma}\right)^2 \right)$. σ^2 is the variance of the distribution from which partitions are drawn (before censoring).

We now use these results to fill in Equation 7:

$$V(\hat{\theta}) = \frac{1}{P} \cdot \left((1-\alpha_2-\alpha_1)\left(\theta_T + \sigma_1^2\right) + \Lambda^2(\alpha_2+\alpha_1) - E[\hat{\theta}]^2 \right). \tag{10}$$

Finally, our estimator of this variance simply involves plugging in for $\{\hat{\alpha}_1, \hat{\alpha}_2, \tilde{\theta}^{\text{dp}}, \sigma^{\text{dp}}, \hat{\theta}^{\text{dp}}\}$ the values $\{\alpha_1, \alpha_2, \theta, \sigma, E[\hat{\theta}]\}$, respectively.

Next, we decompose the second parameter of the variance matrix of Equation 6 in the same way: $\hat{V}(\hat{\alpha}_2^{\text{dp}}) = V(\hat{\alpha}_2) + S_{\hat{\alpha}}^2$, the first component of which is the variance of the proportion of partitions that are censored (prior to adding noise). We represent whether a partition is censored or not by an indicator variable equal to 1 with probability α_2: If $A_p = \mathbb{1}(\hat{\theta}_p > \Lambda)$, then $\Pr(A_p = 1) = \alpha_2$. Then the sum of iid binary variables is a binomial, with variance $V\left(\sum_{p=1}^P A_p\right) = P\alpha_2(1-\alpha_2)$. Plugging $\hat{\alpha}_2^{\text{dp}}$ into the decomposition yields

$$\hat{V}(\hat{\alpha}) = \frac{1}{P}(1-\hat{\alpha}_2^{\text{dp}})\hat{\alpha}_2^{\text{dp}} + S_{\hat{\alpha}}^2. \tag{11}$$

Finally, we derive the covariance:

$$\begin{aligned}
\text{Cov}(\hat{\theta}^{\text{dp}}, \hat{\alpha}_2^{\text{dp}}) &= \text{Cov}(\hat{\theta}, \hat{\alpha}_2) \qquad \text{(noise is additive and independent)} \\
&= \text{Cov}\left(\frac{1}{P}\sum_{p=1}^P c(\hat{\theta}_p, \Lambda), \frac{1}{P}\sum_{p=1}^P A_p \right) \\
&= \frac{1}{P}\text{Cov}\left(c(\hat{\theta}_1, \Lambda), A_1 \right) \qquad (\hat{\theta}_p \text{ and } A_p \text{ are iid over } p) \\
&= \frac{1}{P}\left\{ E[c(\hat{\theta}_1, \Lambda)A_1] - E[c(\hat{\theta}_1, \Lambda)E(A_1)] \right\} \\
&= \frac{1}{P}\left\{ E[c(\hat{\theta}_1, \Lambda) \mid A_1 = 1) - E[c(\hat{\theta}_1, \Lambda) \mid A_1 = 0] \right\} \alpha_2(1-\alpha_2)
\end{aligned} \tag{12}$$

where $E[c(\hat{\theta}_1, \Lambda) \mid A_1 = 1] = \Lambda$, and $E[c(\hat{\theta}_1, \Lambda) \mid A_1 = 0] = \theta_T$, the mean of the truncated normal mean component of the censored normal. We thus use Equation 8 and plug estimates into Equation 12:

$$\text{Cov}(\hat{\theta}^{\text{dp}}, \hat{\alpha}_2^{\text{dp}}) = \frac{1}{P}\left(\Lambda - \frac{\hat{\theta}^{\text{dp}} - \hat{\alpha}_2\Lambda + \hat{\alpha}_1\Lambda}{1-\alpha_2-\alpha_1} \right) \hat{\alpha}_2(1-\hat{\alpha}_2). \tag{13}$$

Appendix H Additional Simulations

In this appendix, we provide additional simulation evidence to demonstrate the performance of our method in finite samples. First, we convey the robustness of our method to an alternative data generating process — and one that is a harder test of our procedure due to the skewed nature of the data, making it harder for the Central Limit Theorem to be approximated by a fixed sample size. Specifically, we draw $X_i \sim \text{Exp}(2)$ for $i = 1, ..., 200,000$, and $Y_i \sim \text{Bern}(\pi_i)$ where $\pi = \frac{\exp(\alpha+\beta X_i)}{\exp(\alpha+\beta X_i)+1}$ and $\alpha = \beta = 0.25$. Our quantity of interest is β.

We simulate 250 data sets for a range of privacy budgets (quantified by ϵ), and divide the data set into $P = 100$ partitions, yielding a sample size of 2000 per partitions. We set λ such that 30% of partitions were censored on average.

Our results are shown in Figures 2 which show that our estimator, $\tilde{\theta}$, correct for the bias in the censored estimate, $\hat{\theta}$, across a range of privacy budgets. Our standard error estimate align approximately with the true standard deviation of the estimate, although slightly conservative on average for smaller values of ϵ.

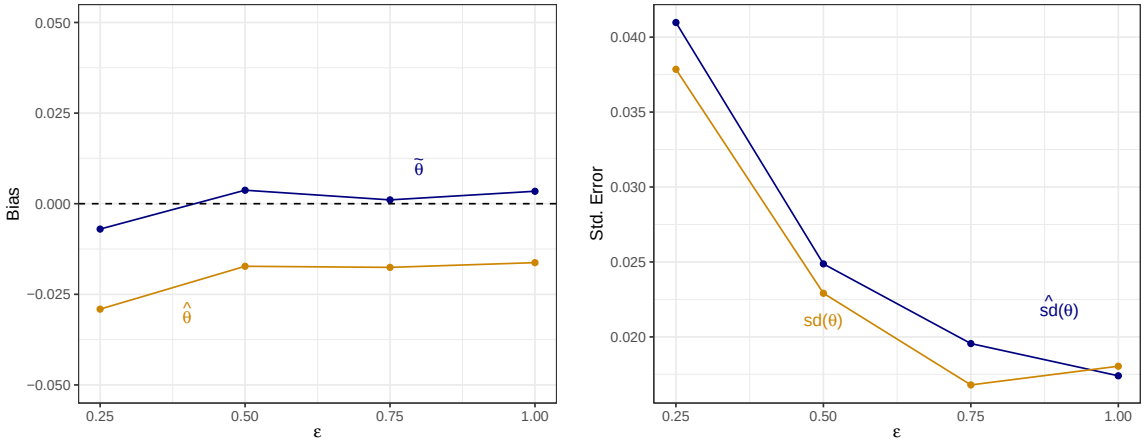

Figure 2: Performance in Non-normal Data

Second, we show in Figure 3 how our procedure performs across a variety of partition sizes (P) for a fixed privacy budget, sample size and censoring level, using the DGP used in our main simulations. When P is small, we see that both our bias-corrected estimate ($\tilde{\theta}^{\text{dp}}$) and the estimate without corrections ($\hat{\theta}^{\text{dp}}$) are biased. Our estimate is biased because

the noise is high with low P and because, when P is low, our estimate of the censoring level is extremely noisy (note however that in simulations where we use the bag-of-little-bootstraps to estimate the proportion of censored partitions, we perform more favorably with low P, since the asymptotics of this estimator are in n rather than P). However, when P reaches 200, our procedure corrects the bias from censoring very effectively. Even when P is 500 (and so the partition level sample size is only $n = 20$) our procedure appears to perform well for this DGP. We have found that such a small sample size can be less optimal for other estimators and in skewed data however, such as the example above. As we discuss in the paper, analysts should consider the trade-off between noise and the partition-level sample size when choosing P.

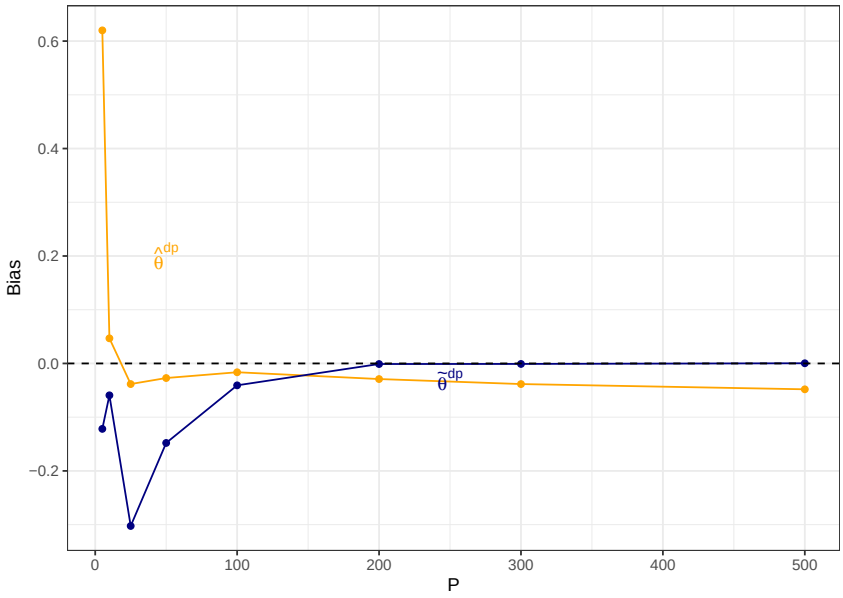

Figure 3: Performance across different values of P (number of data partitions) for a fixed privacy budget ($\epsilon = 1$) and sample size ($N = 100k$).

Appendix I A "True Negative" Example

This appendix provides an example of where our procedure fails, followed by an explanation for why it fails along with recommendations for how to avoid the problem. We do this by further analysis of the data in the example on the "Effect of Affirmative Action on Bureaucratic Performance in India" in Section 6 of the paper.

To do this, we extend the analysis by including state level fixed effects, of which there are 25 in total. Including state-level (or other geography-based) fixed effects is a common data analysis procedure across political science, designed to control for unmeasured confounders that are collinear with state geography.

We present our results for the same quantity of interest in the text from this alternative specification in Figure 4. With the fixed effects, the true effect is about zero with a relatively narrow confidence interval (see the bar at the left of the figure). Our estimate (presented in the middle of the graph) suggests instead that the effect is negative and significantly different from zero.

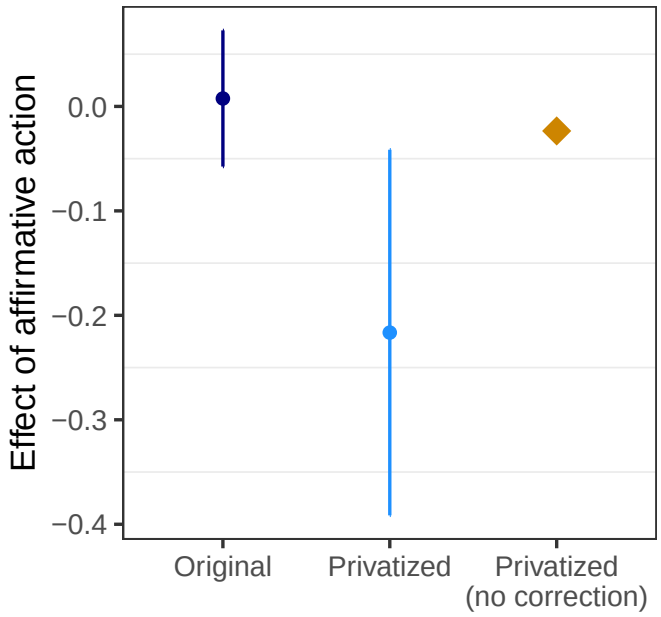

Figure 4: Biased Results

The cause of the problem here is the combination of a relatively small sample size ($n = 2,047$), a relatively large number of parameters (27), along with too many disjoint partitions ($P = 150$). Each one of these is no problem, but all three together result in some partitions containing zero or very few observations from some states, thus making it either impossible to run the model with all the indicator variables, or leading to extremely imprecise estimates. We can see the imprecision reflected in vastly larger confidence intervals. In this instance, the mean partition estimate does not correspond to the estimate

on the full sample, and censoring and DP noise further increases the error.

Technically, if at least one partition had no observations from one of the states, the entire procedure would fail to produce an answer at all because of collinearity with the constant term. However, our code follows the coding conventions of most statistical analysis programs by dropping state indicators when no states appear in that sample, which makes sense statistically since controlling for confounding is not necessary for constants.

Appendix J Proportion of Observations Effectively Lost to Privacy Protection

We first define L, described in Section 7 as the proportion of observations effectively lost due to privacy protective procedures, and then show how to estimate it.

Denote $\hat{\theta}_N$ as the estimator we would calculate if the data were not private and $\tilde{\theta}_N^{\text{dp}}$ as our estimator — each based on N observations. Then we set as our goal estimating N^* (with $N^* < N$) such that $V(\hat{\theta}_{N^*}) = V(\tilde{\theta}_N^{\text{dp}})$. Because $V(\hat{\theta}_{N^*}) \propto 1/N^*$ and $V(\tilde{\theta}_N^{\text{dp}}) \propto 1/N$, we can write $V(\hat{\theta}_{N^*}) = N \cdot V(\hat{\theta}_N)/N^* = V(\tilde{\theta}_N^{\text{dp}})$. We then write the proportionate (effective) loss in observations due to the privacy protective procedures L as

$$L = \frac{N - N^*}{N} = 1 - \frac{V(\hat{\theta}_N)}{V(\tilde{\theta}_N^{\text{dp}})}. \tag{14}$$

We estimate the numerator of the second term as $\hat{\sigma}^2_{\text{dp}}/P$, where $\hat{\sigma}^2_{\text{dp}}$ in the numerator, and the whole denominator, are outputs from our bias correction and variance estimation algorithms (Section 4). So when a dataset has N observations, but is being provided through a differentially private mechanism, this is the equivalent to the researcher having only $LN < N$ observations and no privacy protective procedures. The final estimator is then simply:

$$\hat{L} = 1 - \frac{\hat{\sigma}^2_{\text{dp}}/P}{V(\tilde{\theta}^{\text{dp}})}. \tag{15}$$

This estimator summarizes the effect of the differentially private mechanism and the privacy parameters (ϵ, δ, and Λ).

Appendix K When Privacy Procedures Obscure All Relevant Information

All privacy protective procedures are designed to destroy or hide information by making it more difficult to draw certain inferences from confidential data. These are worthwhile to protect individual privacy and to ensure that data which might not otherwise be accessible at all are in fact available to researchers. However, with the noise and censoring used in differential privacy, some inferences will be so uncertain that no substantive knowledge can be learned. In even more extreme situations, our bias correction procedures, which rely on some information passing through the differential privacy filters, would have no leverage left to do their work. In this appendix, we develop a *rule of thumb* that suggests when privacy protected data analysis becomes like trying to get blood from a stone: $\max(\alpha_1, \alpha_2) > 0.6$ or $\epsilon P < 100$ (also, if $\epsilon P \gg 100$ then $\max(\alpha_1, \alpha_2)$ could be even larger before a problem occurs). If an analysis is implicated by this rule of thumb, then it is best to rerun the analysis with more partitions, use more of the privacy budget, or adjust Λ. If none of these are possible, then the only options are to negotiate with the data provider for a larger privacy budget allocation, collect more data, or abandon inquiry into this particular quantity of interest.

Recall that we attempt to choose Λ in order that each $\hat{\theta}_p \in [-\Lambda, \Lambda]$. We then keep this interval fixed and study the distribution of the mean $\hat{\theta} = \frac{1}{P}\sum_{p=1}^{P} \hat{\theta}_p$, which has a variance P times smaller than the distribution of $\hat{\theta}_p$. Now consider the unusual edge case where so much noise is added that $|\hat{\theta}^{\text{dp}}| \gg \Lambda$ (in contrast to a small deviation, which has little consequence). In this extreme situation, using $\hat{\theta}^{\text{dp}}$ as a plug-in estimator for $E(\hat{\theta}^{\text{dp}})$ no longer works because no values of θ and σ^2 can be logically consistent with it, given Λ; in some ways, such a result even nonsensically suggests that $\sigma^2 < 0$.

In this situation, we could simply stop and declare that no reasonable inference is possible and, if we do, we wind up with an analogous rule of thumb. However, to build intuition for this rule, we now show what happens if we try to accommodate this edge case computationally. Thus, if $\hat{\theta}^{\text{dp}} > \Lambda$ we learn that $e > 0$ (where e is the differentially private error defined in Equations 4-5), and so we replace $\hat{\theta}^{\text{dp}}$ with $\check{\theta}^{\text{dp}} \equiv \hat{\theta}^{\text{dp}} - S\frac{\sqrt{2}}{\sqrt{\pi}}$, where

the second term is $E(e|\hat{\theta}^{\text{dp}} > \Lambda) = E(e|e > 0)$. This adjustment makes the system of equations (and the resulting $\tilde{\theta}^{\text{dp}}$) possible, at the cost of some (third order) bias. We now derive our rule of thumb by showing how to bound this bias by appropriately choosing ϵ, P, and Λ.

For simplicity, we study the dominant case of one-sided censoring ($\alpha_1 = 0$), which enables us to solve the bias correction equations algebraically rather than numerically; the results are not very different for two-sided censoring. Thus, begin with the facts, including α_2 in Equation 8 and

$$\hat{\theta}^{\text{dp}} = (1 - \hat{\alpha}_2^{\text{dp}}) \left[\theta - \frac{\frac{\sigma}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(\frac{\Lambda - \hat{\theta}^{\text{dp}}}{\sigma} \right)^2 \right)}{(1 - \hat{\alpha}_2^{\text{dp}})} \right] + \hat{\alpha}_2^{\text{dp}} \Lambda. \tag{16}$$

Then solve these equations for θ, which we label $\tilde{\theta}^{\text{dp}}$ as above, and show, conditional on α_2, that $\tilde{\theta}^{\text{dp}}$ is a linear function of $\hat{\theta}^{\text{dp}}$:

$$\tilde{\theta}^{\text{dp}} = \hat{\theta}^{\text{dp}} \left(\frac{1}{B} \right) + \Lambda \left(\frac{B - 1}{B} \right), \tag{17}$$

where $B = (1 - \hat{\alpha}_2^{\text{dp}}) + \frac{\sqrt{2} e^{-T^2/2}}{2T\sqrt{\pi}}$ and $T = \sqrt{2} \cdot \text{erf}^{-1}[2(1 - \hat{\alpha}_2^{\text{dp}}) - 1]$.

Note that if we apply our bias correction (in Section 4.1) using the exact version of $E(\hat{\theta})$ (and α_2) as an input, we would find $\tilde{\theta}^{\text{dp}} = \theta$. We are therefore interested in the discrepancy $d = E(\check{\theta}^{\text{dp}}) - E(\hat{\theta})$, which we write as

$$\begin{aligned}
d =& \left[(1 - \Pr(\hat{\theta}^{\text{dp}} > \Lambda)) \int_{-\infty}^{\Lambda} \frac{t \mathcal{N}(t|\hat{\theta}, S^2)}{(1 - \Pr(\hat{\theta}^{\text{dp}} > \Lambda))} dt \right. \\
&\left. + \Pr(\hat{\theta}^{\text{dp}} > \Lambda) \int_{\Lambda}^{\infty} \frac{\left(t - S\frac{\sqrt{2}}{\sqrt{\pi}} \right) \mathcal{N}(t|\hat{\theta}, S^2)}{\Pr(\hat{\theta}^{\text{dp}} > \Lambda)} dt \right] - E[\hat{\theta}] \\
=& \left[E[\hat{\theta}^{\text{dp}}] - \Pr(\hat{\theta}^{\text{dp}} > \Lambda) S \frac{\sqrt{2}}{\sqrt{\pi}} \right] - E[\hat{\theta}] \\
=& - S \frac{\sqrt{2}}{\sqrt{\pi}} \times \Pr(\hat{\theta}^{\text{dp}} > \Lambda) \\
=& - \frac{2\Lambda\sqrt{2\ln(1.25/\delta)}}{\epsilon P} \frac{\sqrt{2}}{\sqrt{\pi}} \times \Pr(\hat{\theta}^{\text{dp}} > \Lambda),
\end{aligned} \tag{18}$$

where $\Pr(\hat{\theta}^{\text{dp}} > \Lambda) = \int_{\Lambda}^{\infty} \mathcal{N}(t|\hat{\theta}, S^2) dt$ has a maximum value of 0.5. As a result, the

maximum value of the discrepancy is

$$\max(d) = -\frac{2\Lambda\sqrt{\ln(1.25/\delta)/\pi}}{\epsilon P}. \tag{19}$$

Making use of Equation 17, we write the maximum possible bias in $\tilde{\theta}^{\text{dp}}$ as a function of the maximum possible bias in $\check{\theta}^{\text{dp}}$. Thus,

$$E[\tilde{\theta}^{\text{dp}}] - \theta \leq \left(\frac{1}{B}\right) \cdot \max(d) \tag{20}$$

which shows that the bias depends on $1/B$, which itself is a deterministic function of α_2.

As shown in Figure 5, which plots this relationship, if censoring (plotted horizontally) is 0.5, then $\tilde{\theta}^{\text{dp}}$ is unbiased. We also see that we can control the maximum value of $1/B$ by controlling the level of censoring. If we follow our rule of thumb and disallow censoring over 60%, then $\max_{0\leq\alpha_2\leq 0.6} |\frac{1}{B}| = 1$.

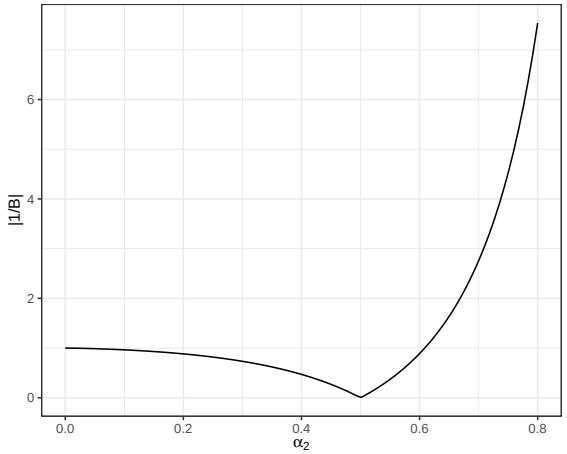

Figure 5: Relationship between $|1/B|$ and Percent Censored

To find the maximum bias under this decision rule, note that if $\Pr(\hat{\theta}^{\text{dp}} > \Lambda)$ is at its maximum, then $\alpha_2 = 0.5$ and $1/B \to 0$. It follows that $\frac{1}{B}\Pr(\hat{\theta}^{\text{dp}} > \Lambda)$ is strictly less than 0.5 and we are able to bound the absolute value of the discrepancy:

$$|E(\tilde{\theta}^{\text{dp}}) - \theta| < \left|\frac{2\Lambda\sqrt{\ln(1.25/\delta)/\pi}}{\epsilon P}\right|. \tag{21}$$

We use this result to show that we have approximately bounded the bias in $\tilde{\theta}^{\text{dp}}$ (if the computational fix is applied) relative to our quantity of interest θ. Since users set Λ on the

scale of their quantity of interest to the range $[-\Lambda, \Lambda]$, the maximum proportionate bias is less than approximately

$$\frac{1}{\Lambda}\left|\frac{2\Lambda\sqrt{\ln(1.25/\delta)\pi}}{\epsilon P}\right| = \left|\frac{2\sqrt{\ln(1.25/\delta)\pi}}{\epsilon P}\right|. \tag{22}$$

For example, if we choose, from our rule of thumb, $\epsilon P = 100$ and $\delta = 0.01$, then this evaluates to 0.03, a small proportionate bias. Of course, this is the upper bound; the actual bias is likely to be a good deal smaller than even this small bound in most applications.

References

Balle, Borja and Yu-Xiang Wang (2018): "Improving the gaussian mechanism for differential privacy: Analytical calibration and optimal denoising". In: *International Conference on Machine Learning (ICML), arXiv:1805.06530*.

Bun, Mark and Thomas Steinke (2016): "Concentrated differential privacy: Simplifications, extensions, and lower bounds". In: *Theory of Cryptography Conference*. Springer, pp. 635–658.

Dwork, Cynthia and Aaron Roth (2014): "The algorithmic foundations of differential privacy". In: *Foundations and Trends in Theoretical Computer Science*, no. 3–4, vol. 9, pp. 211–407.

Feller, W. (1971): *An Introduction to Probability Theory and Its Applications, 3rd Edition*. Wiley.

Garfinkel, Simson L, John M Abowd, and Sarah Powazek (2018): "Issues encountered deploying differential privacy". In: *Proceedings of the 2018 Workshop on Privacy in the Electronic Society*. ACM, pp. 133–137.

Haeberlen, Andreas, Benjamin C Pierce, and Arjun Narayan (2011): "Differential Privacy Under Fire." In: *USENIX Security Symposium*.

King, Gary, Michael Tomz, and Jason Wittenberg (Apr. 2000): "Making the Most of Statistical Analyses: Improving Interpretation and Presentation". In: *American Journal of Political Science*, no. 2, vol. 44, pp. 341–355. URL: bit.ly/makemost.

King, Gary and Langche Zeng (2002): "Estimating Risk and Rate Levels, Ratios, and Differences in Case-Control Studies". In: *Statistics in Medicine*, vol. 21, pp. 1409–1427. URL: bit.ly/estrrCC.

Kleiner, Ariel, Ameet Talwalkar, Purnamrita Sarkar, and Michael I Jordan (2014): "A scalable bootstrap for massive data". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, no. 4, vol. 76, pp. 795–816.

Liu, Jingcheng and Kunal Talwar (2018): "Private Selection from Private Candidates". In: *CoRR*, vol. abs/1811.07971. arXiv: 1811.07971. URL: http://arxiv.org/abs/1811.07971.

Politis, Dimitris N, Joseph P Romano, and Michael Wolf (1999): *Subsampling*. Springer Science & Business Media.

Smith, Adam (2011): "Privacy-preserving statistical estimation with optimal convergence rates". In: *Proceedings of the forty-third annual ACM symposium on Theory of computing*. ACM, pp. 813–822.