Finite Sample Differentially Private Confidence Intervals*

Vishesh Karwa[†]

Salil Vadhan[‡]

Abstract

We study the problem of estimating finite sample confidence intervals of the mean of a normal population under the constraint of differential privacy. We consider both the known and unknown variance cases and construct differentially private algorithms to estimate confidence intervals. Crucially, our algorithms guarantee a finite sample coverage, as opposed to an asymptotic coverage. Unlike most previous differentially private algorithms, we do not require the domain of the samples to be bounded. We also prove lower bounds on the expected size of any differentially private confidence set showing that our the parameters are optimal up to polylogarithmic factors.

. Introduction

1.1 Overview

Differential privacy (Dwork et al., 2006) is a strong and by now widely accepted definition of privacy for statistical analysis of datasets with sensitive information about individuals. While there is now a rich and flourishing body of research on differential privacy, extending well beyond theoretical computer science, the following three basic goals for research in the area have not been studied in combination with each other:

Differentially private statistical inference: The vast majority of work in differential privacy has studied how well one can approximate statistical properties of the dataset itself, i.e. empirical quantities, rather than inferring statistics of an underlying *population* from which a dataset is drawn. Since the latter is the ultimate goal of most data analysis, it should also be a more prominent object of study in the differential privacy literature.

Conservative statistical inference: An important purpose of statistical inference is to limit the chance that data analysts draw incorrect conclusions because their dataset may not accurately reflect the underlying population, for example due to the sample size being too small. For this reason, classical statistical inference also offers measures of statistical significance such as *p*-values and confidence intervals. Constructing such measures for differentially private algorithms is more complex, as one must also take into account the additional noise that is introduced for the purpose of privacy protection. For this reason, we advocate that differentially private inference should be *conservative*, and err on the side of underestimating statistical significance, even at small sample sizes and for all settings of other parameters.

^{*}Presented at TPDP 2017 and to appear at ITCS 2018.

[†]Department of Statistics, The Ohio State University, karwa.8@osu.edu. Work done while a Post doctoral fellow in the Center for Research on Computation & Society, The Department of Statistics and the School of Engineering and Applied Sciences at Harvard University, supported by the NSF grant CNS-1237235.

[‡]Center for Research on Computation & Society and School of Engineering & Applied Sciences, Harvard University, salil_vadhan@harvard.edu, http://seas.harvard.edu/~salil. Supported by NSF grant CNS-1237235, a Simons Investigator Award, a grant from the Sloan Foundation, and Cooperative Agreement CB16ADR0160001 with the Census Bureau.

Rigorous analysis of the inherent price of privacy: As has been done extensively in the differential privacy literature for empirical statistics, we should also investigate the fundamental "privacy–utility tradeoffs" for (conservative) differentially private statistical inference. This involves both designing and analyzing differentially private statistical inference procedures, as well as proving negative results about the performance that can be achieved, using the best non-private procedures as a benchmark.

In this paper, we pursue all of these goals, using as a case study the problem of constructing a confidence interval for the mean of normal data. The latter is one of the most basic problems in statistical inference, yet already turns out to be nontrivial to fully understand under the constraints of differential privacy. We expect that most of our modeling and methods will find analogues for other inferential problems (e.g. hypothesis testing, Bayesian credible intervals, non-normal data, and estimating statistics other than the mean).

1.2 Confidence Intervals for a Normal Mean

We begin by recalling the problem of constructing a $(1 - \alpha)$ -level confidence interval for a normal mean without privacy. Let X_1, \ldots, X_n be an independent and identically distributed *(iid)* random sample from a normal distribution with an unknown mean μ and variance σ^2 . The goal is to design an estimator I that given X_1, \ldots, X_n , outputs an interval $I(X_1, \ldots, X_n) \subseteq \mathbb{R}$ such that

$$\mathbb{P}\left(I(X_1,\ldots,X_n) \ni \mu\right) \ge 1 - \alpha$$

for all μ and σ . Here $1 - \alpha$ is called the *coverage probability*. Given a desired coverage probability, the goal is minimize the *expected length* of the interval, namely $\mathbb{E}[|I(X_1, \ldots, X_n)|]$.

Known Variance. In the case that variance σ^2 is known (so only μ is unknown), the classic confidence interval for a normal mean is:

$$I(X_1,\ldots,X_n) = \bar{X} \pm \frac{\sigma}{\sqrt{n}} \cdot z_{1-\alpha/2},$$

where \bar{X} is the sample mean and z_a represents the a^{th} quantile of a standard normal distribution.¹ It is known that this interval has the smallest expected size among all $1 - \alpha$ level confidence sets for a normal mean, see for example, Lehmann and Romano (2006). In this case, the length of the confidence interval is fixed and equal to

$$|I(X_1,\ldots,X_n)| = (2\sigma z_{1-\alpha/2})/\sqrt{n} = \Theta\left(\sigma\sqrt{\log(1/\alpha)/n}\right).$$

Unknown Variance. In the case that the variance σ^2 is unknown, the variance must be estimated from the data itself, and the classic confidence interval is:

$$I(X_1,\ldots,X_n) = \bar{X} \pm \frac{s}{\sqrt{n}} \cdot t_{n-1,1-\alpha/2},$$

where s^2 is the sample variance defined by

$$s^{2} = \frac{1}{n-1} \sum_{i=1}^{n} (X_{i} - \bar{X})^{2},$$

and $t_{n-1,a}$ is the a^{th} quantile of a t-distribution with n-1 degrees of freedom (see the appendix for definitions).² Now the length of the interval is a random variable with expectation

$$\mathbb{E}[|I(X_1,\ldots,X_n)|] = \frac{2\sigma}{\sqrt{n}} \cdot k_n \cdot t_{n-1,1-\alpha/2} = \Theta\left(\sigma\sqrt{\log(1/\alpha)/n}\right),$$

for an appropriate constant $k_n = 1 - O(1/n)$. (See Lehmann and Romano (2006).)

¹The proof that this is in fact a $(1 - \alpha)$ -confidence interval follows by observing that $\sqrt{n} \cdot (\bar{X} - \mu)$ has a standard normal distribution, and $[-z_{1-\alpha/2}, z_{1-\alpha/2}]$ covers a $1 - \alpha$ fraction of the mass of this distribution.

²Again the proof follows by observing that $s \cdot (\bar{X} - \mu)$ follows a t distribution, with no dependence on the unknown parameters.

Relation to Hypothesis Tests. In general, including both cases above, a confidence interval for a population parameter also gives rise to hypothesis tests, which is often how the confidence intervals are used in applied statistics. For example, if our null hypothesis is that the mean μ is nonnegative, then we could reject the null hypothesis if the interval $I(X_1, \ldots, X_n)$ does not intersect the positive real line. The significance level of this hypothesis test is thus at least $1 - \alpha$. Minimizing the length of the confidence interval corresponds to being able to reject the alternate hypotheses that are closer to the null hypothesis; that is, when the confidence interval is of length at most β and μ is distance greater than β from the null hypothesis, then the test will reject with probability at least $1 - \alpha$.

1.3 Differential Privacy

Let $\underline{x} = (x_1, \ldots, x_n)$ be a dataset of *n* elements where each $x_i \in \Omega$. In the problems that we consider, $\Omega = \mathbb{R}$ and $\underline{x} \in \mathbb{R}^n$. Two datasets \underline{x} and \underline{x}' , both of size *n*, are called *neighbors* if they differ by one element.

Definition 1 (Differential Privacy, Dwork et al. (2006)). A randomized algorithm $M : \Omega^n \to \Omega_M$ is (ϵ, δ) differentially private if for all neighboring $\underline{x}, \underline{x}' \in \Omega^n$ and for all measurable sets of outputs $S \subseteq \Omega_M$, we have

$$\mathbb{P}\left(M(\underline{x}) \in S\right) \le e^{\epsilon} \cdot \mathbb{P}\left(M(\underline{x}') \in S\right) + \delta,$$

where the probability is over the randomness of M.

Intuitively, this captures privacy because arbitrarily changing one individual's data (i.e. changing a row of \underline{x} to obtain \underline{x}') has only a small effect on the output distribution of M. Typically we think of ϵ as a small constant (e.g. $\epsilon = .01$) while δ should be cryptographically small (in particular, much smaller than 1/n) to obtain a satisfactory formulation of privacy. The case that $\delta = 0$ is often called *pure* differential privacy, while $\delta > 0$ is known as *approximate* differential privacy.

Nontrivial differentially private algorithms are necessarily randomized (as the probabilities in the definition above are taken only over the randomness of M, which is important for the interpretation and composability of the definition), and thus such algorithms work by injecting randomness into statistical computations to obscure the effect of each individual.

The most basic differentially private algorithm is the Laplace Mechanism (Dwork et al., 2006), which approximates an arbitrary a function $f: \Omega^n \to \mathbb{R}$ by adding Laplace noise:

$$M(\underline{x}) = f(\underline{x}) + Z$$
, where $Z \sim \text{Lap}(GS_f/\epsilon)$.

Here $\text{Lap}(\tau)$ is the Laplace distribution with scale τ (which has standard deviation proportional to τ), and GS_f is the global sensitivity of f — the maximum of $|f(\underline{x}) - f(\underline{x}')|$ over all pairs of neighboring datasets $\underline{x}, \underline{x}'$.

In particular, if $f(\underline{x})$ is the *empirical mean* of the dataset \underline{x} and $\Omega = [-B, B]$, then $GS_f = 2B/n$, so $M(\underline{x})$ approximates the empirical mean to within additive error $O(B/\epsilon n)$ with high probability.

1.4 Statistical Inference with Differential Privacy

The Laplace mechanism described above is about estimating a function $f(\underline{x})$ of the dataset \underline{x} , rather than the population from which \underline{x} is drawn, and much of the differential privacy literature is about estimating such empirical statistics. There are several important exceptions, the earliest being the work on differentially private PAC learning (Blum et al. (2005); Kasiviswanathan et al. (2011)), but still many basic statistical inference questions have not been addressed.

However, a natural approach was already suggested in early works on differential privacy. In many cases, we know that population statistics are well-approximated by empirical statistics, and thus we can try to estimate these empirical statistics with differential privacy. For example, the population mean μ for a normal population is well-approximated by the sample mean \overline{X} , which we can estimate using the Laplace mechanism:

$$M(X_1,\ldots,X_n) = X + Z$$
, where $Z \sim \text{Lap}(2B/\epsilon n)$.

On the positive side, observe that the noise being introduced for privacy vanishes linearly in 1/n, whereas \overline{X} converges to the population mean at a rate of $1/\sqrt{n}$, so asymptotically we obtain privacy "for free" compared to the (optimal) non-private estimator \overline{X} .

However, this rough analysis hides some important issues. First, it is misleading to look only at the dependence on n. The other parameters, such as σ , ϵ , and B can be quite significant and should not be treated as constants. Indeed $\sigma/\sqrt{n} \gg B/\epsilon n$ only when $n \gg (B/\epsilon \sigma)^2$, which means that the asymptotics only kick in at a very large value of n. Thus it is important to determine whether the dependence on these parameters is necessary or can be improved. Second, the parameter B is supposed to be a (worst-case) bound on the range of the data, which is incompatible with a modeling the population as following a normal distribution (which is supported on the entire real line). Thus, there have been several works seeking the best asymptotic approximations we can obtain for population statistics under differential privacy, such as Dwork and Lei (2009); Smith (2011); Wasserman and Zhou (2010); Wasserman (2012); Hall et al. (2013); Duchi et al. (2013a,b); Barber and Duchi (2014).

1.5 Conservative Statistical Inference with DP

The works discussed in the previous section focus on providing point estimates for population quantities, but as mentioned earlier, it is also important to be able to provide measures of statistical significance, to prevent analysts from drawing incorrect conclusions from the results. These measures of statistical significance need to take into account the uncertainty coming both from the sampling of the data and from the noise introduced for privacy. Ignoring the noise introduced for privacy can result in wildly incorrect results at finite sample sizes, as demonstrated empirically many times (e.g. Fienberg et al. (2010); Karwa and Slavković (2012, 2016)) and this can have severe consequences. For example, Fredrikson et al. (2014) found that naive use of differential privacy in calculating warfarin dosage would lead to unsafe levels of medication, but of course one should never use any sort of statistics for life-or-death decisions without some analysis of statistical significance.

Since calculating the exact statistical significance of differentially private computations seems difficult in general, we advocate *conservative* estimates of significance. That is, we require that $\mathbb{P}(I(X_1, \ldots, X_n) \ni \mu) \ge 1 - \alpha$, for all values of n, values of the population parameters, and values of the privacy parameter.

For sample sizes that are too small or privacy parameters that are too aggressive, we may achieve this property by allowing the algorithm to sometimes produce an extremely large confidence interval, but that is preferable to producing a small interval that does not actually contain the true parameter which may violate the desired coverage property. Note that what constitutes a sample size that is "too small" can depend on the unknown parameters of the population (e.g. the unknown variance σ^2) and their interplay with other parameters (such as the privacy parameter ϵ).

Returning to our example of estimating a normal mean with known variance under differential privacy, if we use the Laplace Mechanism to approximate the empirical mean (as discussed above), we can obtain a conservative confidence interval for the population mean by increasing the length of classical, non-private confidence interval to account for the likely magnitude of the Laplace noise. More precisely, starting with the differentially private mechanism

$$M(X_1,\ldots,X_n) = X + Z$$
, where $Z \sim \text{Lap}(2B/\epsilon n)$,

the following is a $(1 - O(\alpha))$ -level confidence interval for the population mean μ :

$$I(X_1, \dots, X_n) = M(X_1, \dots, X_n) \pm \left(\frac{\sigma}{\sqrt{n}} \cdot z_{1-\alpha/2} + \frac{B}{\epsilon n} \cdot \log(1/\alpha)\right).$$

The point is that with probability $1 - O(\alpha)$, the Laplace noise Z has magnitude at most $(B/\epsilon n) \cdot \log(1/\alpha)$, so increasing the interval by this amount will preserve coverage (up to an $O(\alpha)$ change in the probability). Again, the privacy guarantees of the Laplace mechanism relies on the data points being guaranteed to lie in [-B, B]; otherwise, points need to be clamped to lie in the range, which can bias the empirical mean and compromise

the coverage guarantee. Thus, to be safe, a user may choose a very large value of B, but then this makes for a much larger (and less useful) interval, as the length of the interval grows linearly with B. Thus, a natural research question (which we investigate) is whether such a choice and corresponding cost is necessary.

Conservative hypothesis testing with differential privacy, where we require that the significance level is at least $1-\alpha$, was advocated by Gaboardi et al. (2016b). Methods aimed at calculating the combined uncertainty due to sampling and privacy (for various differentially private algorithms) were given in Vu and Slavkovic (2009); Williams and McSherry (2010); Karwa and Slavković (2012); Karwa et al. (2015, 2014); Karwa and Slavković (2016); Gaboardi et al. (2016b); Solea (2014); Wang et al. (2015); Kifer and Rogers (2016), but generally the utility of these methods (e.g. the expected length of a confidence interval or power of a hypothesis test) is only evaluated empirically or the conservativeness only holds in a particular asymptotic regime. Rigorous, finite-sample analyses of conservative inference were given in Sheffet (2017) for confidence intervals on the coefficients from ordinary least-squares regression (which can be seen as a generalization of the problem we study to multivariate Gaussians) and in Cai et al. (2017) for hypothesis testing of discrete distributions. However, neither paper provides matching lower bounds, and in particular, the algorithms of Sheffet (2017) only apply for bounded data (similar to the basic Laplace mechanism). In our work, we provide a comprehensive theoretical analysis of conservative differentially private confidence intervals for a normal mean, with both algorithms and lower bounds, without any bounded data assumption.

Before stating our results, we define more precisely the notion of a (conservative) $(1 - \alpha)$ -level confidence set. Let $\mathcal{D} = \{\mathbb{D}_{\theta,\gamma}\}_{\theta \in \Theta, \gamma \in \Gamma}$ be a family of distributions supported on \mathbb{R} where $\theta \in \Theta \subseteq \mathbb{R}$ is a real valued parameter and $\gamma \in \Gamma \subseteq \mathbb{R}^k$ is a vector of *nuisance parameters*. A nuisance parameter is an unknown parameter that is not a primary object of study, but must be accounted for in the analysis. For example, when we consider estimating the mean of a normal distribution with unknown variance, the variance is a nuisance parameter. We write $X_1, \ldots, X_n \stackrel{iid}{\sim} \mathbb{D}_{\theta,\gamma}$ if X_1, \ldots, X_n is an independent and identically distributed random sample from a distribution $\mathbb{D}_{\theta,\gamma}$. We sometimes abuse the notation and write \mathbb{D}_{θ} instead of $\mathbb{D}_{\theta,\gamma}$ when γ is clear from the context.

Definition 2 $((1 - \alpha)$ -level confidence set). Let $\alpha \in (0, 1)$. Let $X_1, \ldots, X_n \stackrel{iid}{\sim} \mathbb{D}_{\theta, \gamma}$ where $\theta \in \Theta \subseteq \mathbb{R}$ and $\gamma \in \Gamma \subseteq \mathbb{R}^k$ is a vector of nuisance parameters. A $(1 - \alpha)$ -level confidence set for θ with sample complexity n is a (possibly randomized) measurable function $I : \mathbb{R}^n \to \mathbb{S}$, where \mathbb{S} is a set of measurable subsets of \mathbb{R} , such that for all $\theta \in \Theta$ and $\gamma \in \Gamma$, we have

$$\mathbb{P}_{\substack{X_1,\ldots,X_n \sim \mathbb{D}_{\theta,\gamma}\\I}} (I(X_1,\ldots,X_n) \ni \theta) \ge 1 - \alpha,$$

where the probability is taken over the randomness of both I and the data X_1, \ldots, X_n .

1.6 Our Results

As discussed above, in this paper we develop conservative differentially private estimators of confidence intervals for the mean μ of a normal distribution with known and unknown variance σ^2 . Our algorithms are designed to be differentially private for all input datasets and they provide $(1 - \alpha)$ -level coverage whenever the data is generated from a normal distribution. Unlike the Laplace mechanism described above and many other differentially private algorithms, we do not make any assumptions on the boundedness of the data. Our pure DP (i.e. $(\epsilon, 0)$ -DP) algorithms assume that the mean μ and variance σ^2 lie in a bounded (but possibly very large) interval, and we show (using lower bounds) that such an assumption is necessary. Our approximate (i.e. (ϵ, δ)) differentially private algorithms do not make any such assumptions, i.e. both the data and the parameters (μ, σ^2) can remain unbounded. We also show that the differentially private estimators that we construct have nearly optimal expected length, up to logarithmic factors. This is done by proving lower bounds on the length of differentially private confidence intervals. A key aspect of the confidence intervals that we construct is their conservativeness — the coverage guarantee holds in finite samples, as opposed to only holding asymptotically. We also show that as $n \to \infty$, the length of our differentially private confidence intervals is at most 1 + o(1)factor larger than length of their non-private counterparts. Let X_1, \ldots, X_n be an independent and identically distributed *(iid)* random sample from a normal distribution with an unknown mean μ and variance σ^2 , where $\mu \in (-R, R)$ and $\sigma \in (\sigma_{\min}, \sigma_{\max})$. Our goal is to construct (ϵ, δ) -differentially private $(1 - \alpha)$ -level confidence sets for μ in both the known and the unknown variance case, i.e. we seek a set $I = I(X_1, \ldots, X_n)$ such that

- 1. $I(X_1, \ldots, X_n)$ is a (1α) -level confidence interval, and
- 2. $I(x_1, \ldots, x_n)$ is (ϵ, δ) -differentially private.
- 3. $\mathbb{E}_{X_1,\ldots,X_n,I}[I(X_1,\ldots,X_n)]$ is as small as possible.

Known Variance: We prove the following result for estimating the confidence interval with the privacy constraint:

Theorem 1.1 (known variance case). There exists an (ϵ, δ) -differentially private algorithm that on input $X_1, \ldots, X_n \stackrel{iid}{\sim} N(\mu, \sigma^2)$ with known σ^2 and unknown mean $\mu \in (-R, R)$ outputs a $(1 - \alpha)$ -level confidence interval for μ . Moreover, if

$$n > \frac{c_1}{\epsilon} \min\left\{\log\left(\frac{R}{\sigma}\right), \log\left(\frac{1}{\delta}\right)\right\} + \frac{c_2}{\epsilon} \log\left(\frac{1}{\alpha}\right),$$

(where c_1 and c_2 are universal constants) then the interval is of fixed width β where

$$\beta \le \max\left\{\frac{\sigma}{\sqrt{n}}\mathcal{O}\left(\sqrt{\log\left(\frac{1}{\alpha}\right)}\right), \frac{\sigma}{\epsilon n} \operatorname{polylog}\left(\frac{n}{\alpha}\right)\right\}$$

Theorem 1.1 asserts that there exists a differentially private algorithm that outputs a fixed width $(1-\alpha)$ -level confidence interval for any n. Moreover, when n is large enough, the algorithm outputs a confidence interval of length β which is non-trivial in the sense that $\beta \ll R$. Specifically, β is a maximum of two terms: The first term is $\mathcal{O}\left(\sigma\sqrt{\log(1/\alpha)/n}\right)$ which is the same as the length of the non-private confidence interval discussed in Section 1.2 up to constant factors. The second term is $\mathcal{O}\left(\sigma/(\epsilon n)\right)$ up to polylogarithmic factors – it goes to 0 at the rate of $\tilde{\mathcal{O}}(1/n)$ which is faster than the rate at which the first term goes to 0. Thus for large n the increase in the length of the confidence interval due to privacy is mild. Note that, unlike the basic approach based on the Laplace mechanism discussed in Section 1.5, the length of the confidence interval has no dependence on the range of the data, or even the range (-R, R) of the mean μ .

The sample complexity required for obtaining a non-trivial confidence interval is the minimum of two terms: $\mathcal{O}((1/\epsilon)\log(R/\alpha\sigma))$ and $\mathcal{O}((1/\epsilon)\log(1/\alpha\delta))$. The dependence of sample complexity on R/σ is only logarithmic. Thus one can choose a very large value of R. Moreover, when $\delta > 0$, we can set $R = \infty$ and hence there is no dependence of the sample complexity on R.

The first term in the length of the confidence interval in Theorem 1.1 hides some constants which can lead to a constant multiplicative factor increase in the length of the differentially private confidence intervals when compared to the non-private confidence intervals. We show that it is possible to eliminate this multiplicative increase and obtain differentially private confidence intervals with only additive increase in the length:

Theorem 1.2. There exists an (ϵ, δ) -differentially private algorithm that on input $X_1, \ldots, X_n \stackrel{iid}{\sim} N(\mu, \sigma^2)$ with known σ^2 and unknown mean $\mu \in (-R, R)$ outputs a $(1 - \alpha)$ -level confidence interval of μ . Moreover, if

$$n > \frac{c}{\epsilon} \min\left\{\log\left(\frac{R}{\sigma}\right), \log\left(\frac{1}{\delta}\right)\right\} + \frac{c}{\epsilon} \log\left(\frac{\log(1/\epsilon)}{\alpha}\right)$$

(where c is a universal constant) then

$$\beta = \frac{2\sigma}{\sqrt{n}} z_{1-\alpha/2} + \frac{\sigma}{\epsilon n} \text{polylog}\left(\frac{n}{\alpha}\right)$$

Theorem 1.2 asserts that with a small change to the sample complexity on n by an additive term of $(1/\epsilon) \cdot \log(1/\epsilon)$, we can achieve an additive increase in the length β of the confidence interval as opposed to a multiplicative increase. Note that the first term in β exactly matches the length of the non-private confidence interval, namely $2\sigma z_{1-\alpha/2}/\sqrt{n}$, while the second term vanishes more quickly as a function of n.³ An important point is that we retain an n rather than an ϵn in the first term, contrary to the common belief that differential privacy has a price of $1/\epsilon$ in sample size. (See Steinke and Ullman (2015) for a proof of such a statement for computing summary statistics of a dataset rather than inference, and Hay et al. (2016) for an informal claim along these lines.)

Unknown Variance: Our (ϵ, δ) -differentially private confidence interval in the unknown variance case is as follows:

Theorem 1.3 (unknown variance case). There exists an (ϵ, δ) -differentially private that on input $X_1, \ldots, X_n \stackrel{iid}{\sim} N(\mu, \sigma^2)$ with unknown mean $\mu \in (-R, R)$ and variance $\sigma^2 \in (\sigma_{\min}, \sigma_{\max})$ always outputs a $(1 - \alpha)$ -level confidence interval of μ . Moreover, if

$$n \ge \frac{c}{\epsilon} \min\left\{ \max\left\{ \log\left(\frac{R}{\sigma_{\min}}\right), \log\left(\frac{\sigma_{\max}}{\sigma_{\min}}\right) \right\}, \log\left(\frac{1}{\delta}\right) \right\} + \frac{c}{\epsilon} \log\left(\frac{\log\left(\frac{1}{\epsilon}\right)}{\alpha}\right),$$

(where c is a universal constant), then the expected length of the interval β is such that

$$\beta \le \max\left\{\frac{\sigma}{\sqrt{n}}\mathcal{O}\left(\sqrt{\log\left(\frac{1}{\alpha}\right)}\right), \frac{\sigma}{\epsilon n} \operatorname{polylog}\left(\frac{n}{\alpha}\right)\right\}$$

As in the known variance case, Theorem 1.3 asserts that there exists an (ϵ, δ) differentially private algorithm that always outputs an $(1-\alpha)$ confidence interval of μ for all n. If n is large enough, the length of the confidence interval is a maximum of two terms, where the first term is same as the length of the non-private confidence interval and the second term goes to 0 at a faster rate.

As before the dependence of sample complexity on R/σ_{\min} and $\sigma_{\max}/\sigma_{\min}$ is logarithmic, as opposed to linear. Hence we can set these parameters to a large number. Moreover, when $\delta > 0$, we can set R and σ_{\max} to be ∞ and σ_{\min} to be 0. Thus when $\delta > 0$, there are no assumptions on the boundedness of the parameters.

Finally, along the lines of Theorem 1.2, at the cost of a minor increase in sample complexity, we can obtain a differentially private algorithm that has only additive increase in the length of the confidence interval that is asymptotically vanishing relative to the non-private length. Specifically, we can obtain an interval with length

$$\beta \leq \frac{2\sigma}{\sqrt{n}} \cdot k_n t_{n-1,1-\alpha/2} + \frac{\sigma}{\epsilon n} \operatorname{polylog}\left(\frac{n}{\alpha}\right),$$

where again the first term is exactly the same as in the non-private case (see Section 1.2) and the second term vanishes more quickly as a function of n.

Lower Bounds. We also prove lower bounds on the length of any $(1 - \alpha)$ -level (ϵ, δ) -differentially private confidence set of expected size β :

Theorem 1.4 (Lower bound). Let M be any (ϵ, δ) -differentially private algorithm that on input $X_1, \ldots, X_n \stackrel{iid}{\sim} N(\mu, \sigma^2), \mu \in (-R, R)$ produces a $(1 - \alpha)$ -level confidence set of μ of expected size β . If $\delta < \alpha/2n$, then

$$\beta \ge c \cdot \min\left\{\frac{\sigma}{\epsilon n} \log\left(\frac{1}{\alpha}\right), R\right\}$$

³We note that when the range R of the mean is bounded, as we require for our pure differentially private algorithms, the length of a non-private algorithm can be improved, but the improvement is insignificant in the regime of parameters we are interested in, namely when $R \ge \Omega(\sigma)$. See Theorem 6.1.

Moreover, if $\beta < \sigma < R$, then

$$n \ge c \min\left(\frac{1}{\epsilon} \log\left(\frac{R}{\sigma}\right), \frac{1}{\epsilon} \log\left(\frac{1}{\delta}\right)\right) + \frac{c}{\epsilon} \log\left(\frac{1}{\alpha}\right)$$

where c is a universal constant.

Note that the first lower bound says that we must pay $\Omega(\sigma/(\epsilon n) \cdot \log(1/\alpha))$ in the length of the confidence interval when R is very large. Our algorithms come quite close to this lower bound with an extra factor of polylog (n/α) . The second lower bound shows that the sample complexity required by Theorem 1.1 is necessary to obtain a confidence interval that saves more than a factor of 2 over the trivial interval (-R, R). By setting $\sigma = \sigma_{\min}$, the sample complexity lower bound also matches that of Theorem 1.3 in our parameter regime of interest, namely when $R \ge \Omega(\sigma_{\max})$.

1.7 Techniques

Known Variance Algorithms: Our algorithms for the known variance case (Theorems 1.1 and 1.2) are based on simple Laplace-mechanism-based confidence interval discussed in Section 1.5, except that we calculate a suitable bound *B* based on the data in a differentially private manner, rather than having it be an input provided by a data analyst. Specifically, we give a differentially private algorithm M_{range} that takes *n* real numbers and outputs an interval (which need not be centered at 0) such that for every $\mu \in (-R, R)$, when $X_1, \ldots, X_n \sim \mathcal{N}(\mu, \sigma^2)$, we have:

- 1. With probability at least 1α over X_1, \ldots, X_n and the coins of M_{range} , we have $\{X_1, \ldots, X_n\} \subseteq M_{range}(X_1, \ldots, X_n)$, and
- 2. With probability 1, $|M_{range}(X_1, \ldots, X_n)| \leq O(\sigma \cdot \sqrt{\log(n/\alpha)}).$

Thus, if we clamp all datapoints to lie in the interval $M_{range}(X_1, \ldots, X_n)$ (which will usually have no effect for data that comes from our normal model, by Property 1), we can calculate an approximate mean and thus construct a confidence interval using Laplace noise of scale $O(\sigma \cdot \sqrt{\log(n/\alpha)}/\epsilon n)$.

Now, estimating the range of a dataset with differential privacy is impossible to do with any nontrivial accuracy in the worst case, so we must exploit the distributional assumption on our dataset to construct M_{range} . Specifically, we exploit the following properties of normal data:

- 1. A vast majority of the probability mass of $\mathcal{N}(\mu, \sigma^2)$ is concentrated in an interval of width $O(\sigma)$ around the mean μ .
- 2. With probability at least 1α , all datapoints X_1, \ldots, X_n are at distance at most $O(\sigma \cdot \sqrt{\log(n/\alpha)})$ from μ .

Similar properties hold for many other natural parameterized families of distributions, changing the factor of $\sqrt{\log(n/\alpha)}$ according to the concentration properties of the family.

Given these properties, M_{range} works as follows: we partition the original range (-R, R) (where R might be infinite) into "bins" (intervals) of width $O(\sigma)$, and calculate an differentially private approximate histogram of how many points lie in each bin. By Property 1 and a Chernoff bound, with high probability, the vast majority of our normally distributed data points will be in the bin containing μ or one of the neighboring bins. Existing algorithms for differentially private histograms (Dwork et al., 2006; Bun et al., 2016) allow us to identify one of these heavy bins with probability $1 - \alpha$, provided $n \ge O(\min\{\log(K/\alpha), \log(1/(\delta\alpha))\})/\epsilon$, where $K = O(R/\sigma)$ is the number of bins. After identifying such a bin, Property 2 tells us that we can simply expand the bin by $O(\sigma \cdot \sqrt{\log(n/\alpha)})$ on each side and include all of the datapoints with high probability. This proof sketch gives a $(1 - O(\alpha))$ -level confidence interval, and redefining α yields Theorem 1.1. To obtain, Theorem 1.2, we set parameters more carefully so that the failure probability in estimating the range is much smaller than α , say $\alpha/\text{poly}(n)$, which increases the sample complexity of the histogram algorithms only slightly. This general approach, of finding a differentially private estimate of the range and using it to compute a differentially private mean are inspired by the work of Dwork and Lei (2009). They present an (ϵ, δ) differentially private algorithm to estimate the *scale* of the data. They use the estimate of scale to obtain a differentially private estimate of the median without making any assumptions on the range of the data. Their algorithms require $\delta > 0$. In contrast, our range finding algorithms for Gaussian data work for $\delta = 0$ without making any assumptions on the range of the data, but instead assume that the parameters need to be bounded). Our algorithms also handle the unknown variance case, as discussed below. Also, while the general idea of eliminating the dependence on range of the data is similar, the underlying techniques and privacy and utility guarantees are different.

Unknown Variance Algorithms: For the case of an unknown variance (Theorem 1.3), we begin with the observation that our range-finding algorithm discussed above only needs a constant-factor approximation to the variance σ^2 . Thus, we will begin by calculating a constant-factor approximation to σ^2 in a differentially private manner, and then estimate the range as above. To do this, we consider the dataset of size n/2 given by $|X_1 - X_2|, |X_3 - X_4|, |X_5 - X_6|, \dots, |X_{n-1} - X_n|$. Here each point is distributed as the absolute value of a $\mathcal{N}(0, 2\sigma^2)$ random variable, which has the vast majority of its probability mass on points of magnitude $\Theta(\sigma)$. Thus, if we partition the interval $(\sigma_{\min}, \sigma_{\max})$ into bins of the form $(2^i, 2^{i+1})$ and apply an approximate histogram algorithm, the heaviest bin will give us an estimate of σ to within a constant factor. Actually, to analyze the expected length of our confidence interval, we will need that our estimate of σ is within a constant factor of the true value not only with high probability but also in expectation; this requires a finer analysis of the histogram algorithm, where the probability of picking any bin decays linearly with the probability mass of that bin (so bins further away from μ have exponentially decaying probability of being chosen). Note that this approach for approximating σ also exploits the symmetry of a normal distribution, so that $|X_i - X_{i+1}|$ is likely to have magnitude $\Theta(\sigma)$, independent of μ ; it should generalize to many other common symmetric distribution families. For non-symmetric families, one could instead use differentially private algorithms for releasing threshold functions (i.e. estimating quantiles) at the price of a small dependence on the ranges even when $\delta > 0$. (See Vadhan (2017, Sec. 7.2) and references therein.)

Now, once we have found the range as in the known-variance case, we can again use the Laplace mechanism to estimate the empirical mean to within additive error $\pm O(\sigma \cdot \sqrt{\log(n/\alpha)}/\epsilon n)$. And we can use our constant-factor approximation of σ to estimate the size of the non-private confidence interval to within a constant factor. This suffices for Theorem 1.3. But to obtain the tighter bound, where we only pay an additive increase over the length of the non-private interval, we cannot just use a constant-factor approximation of the variance. Instead, we also use the Laplace mechanism to estimate the sample variance $s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$. Our bound on the range (with clamping) ensures that s^2 has global sensitivity $O(\sigma^2 \cdot \log(n/\alpha))/(n-1)$, and thus can be estimated quite accurately.

Lower Bounds: For our lower bounds (Thm 1.4), we observe that the expected length of a confidence set $M(X_1, \ldots, X_n)$ can be written as

$$\int_{\mu'} \mathbb{P}_{\mu,M} \left(\mu' \in M(X_1, \dots, X_n) \right)$$

where μ' ranges over (-R, R) and the $\mathbb{P}_{\mu,M}$ notation indicates that probability is taken over (X_1, \ldots, X_n) generated according to $\mathcal{N}(\mu, \sigma^2)$ for a particular value of μ , and over the mechanism M. Next, we use the differential privacy guarantee to deduce that

$$\mathbb{P}_{\mu,M}\left(\mu' \in M(X_1,\ldots,X_n)\right) \ge e^{-6\epsilon n \cdot d} \cdot \mathbb{P}_{\mu',M}\left(\mu' \in M(X_1,\ldots,X_n)\right) - 4n\delta \cdot d,$$

where $d \leq \min\{1, |\mu - \mu'|/\sigma\}$ is the total variation distance between $\mathcal{N}(\mu, \sigma^2)$ and $\mathcal{N}(\mu', \sigma^2)$. (This can be seen as a distributional analogue of the "group privacy" property used in "packing lower bounds" for calculating empirical statistics under differential privacy (Hardt and Talwar, 2010; Beimel et al., 2010; Wasserman and Zhou, 2010; Hall et al., 2011), and is also a generalization of the "secrecy of the sample" property of differential privacy (Kasiviswanathan et al., 2011; Smith, 2009; Bun et al., 2015). Finally, we know that $\mathbb{P}_{\mu',M}$ ($\mu' \in M(X_1, \ldots, X_n)$) $\geq 1 - \alpha$ by the coverage property of M, yielding our lower bound (after some calculations).

1.8 Directions for Future Work

The most immediate direction for future work is to close the (small) gaps between our upper and lower bounds. Most interesting is whether the price of privacy in the length of confidence intervals needs to be even additive, as in Theorem 1.1. Our lower bound only implies that the length of a differentially private confidence interval must be at least the *maximum* of a privacy term (namely, the lower bound in Theorem 1.4) and the non-private length (cf. Theorem 6.1), rather than the sum. In particular, when n is sufficiently large, the non-private length is larger than the privacy term, and Theorem 1.4 leaves open the possibility that a differentially private confidence interval. This seems unlikely, and it would be interesting to prove that there must be some price to privacy even if n is very large.

We came to the problem of constructing confidence intervals for a normal mean as part of an effort to bring differential privacy to practice in the sharing of social science research data through the design of the software tool PSI (Gaboardi et al., 2016a), as confidence intervals are a workhorse of data analysis in the social sciences. However, our algorithms are not optimized for practical performance, but rather for asymptotic analysis of the confidence interval length. Initial experiments indicate that alternative approaches (not just tuning of parameters) may be needed to reasonably sized confidence intervals (e.g. length at most twice that of the non-private length) handle modest sample sizes (e.g. in the 1000's). Thus designing practical differentially private algorithms for confidence intervals remains an important open problem, whose solution could have wide applicability.

As mentioned earlier, we expect that much of the modelling and techniques we develop should also be applicable more widely. In particular, it would be natural to study the estimation of other population statistics, and families of distributions, such as other continuous random variables, Bernoulli random variables, and multivariate families. In particular, a natural generalization of the problem we consider is to construct confidence intervals for the parameters of a (possibly degenerate) multivariate Gaussian, which is closely related to the problem of ordinary least-squares regression (cf. Sheffet (2017)).

Finally, while we have advocated for conservative inference at finite sample size, to avoid spurious conclusions coming from the introduction of privacy, many practical, non-private inference methods rely on asymptotics also for measuring statistical significance. In particular, the standard confidence interval for a normal mean with unknown variance and its corresponding hypothesis test (see Section 1.2) is often applied on non-normal data, and heuristically justified using the Central Limit Theorem. (This is heuristic since the rate of convergence depends on the data distribution, which is unknown.) Is there a criterion to indicate what asymptotics are "safe"? In particular, can we formalize the idea of only using the "same" asymptotics that are used without privacy? Kifer and Rogers (2016) analyze their hypothesis tests using asymptotics that constrain the setting of the privacy parameter in terms of the sample size n (e.g. $\epsilon \ge \Omega(1/\sqrt{n})$), but it's not clear that this relationship is safe to assume in general.

1.9 Organization

The rest of the paper is organized in the following manner. In Section 2, we introduce some preliminary results on DP and techniques such as Laplace mechanism and histogram learners that are needed for our algorithms. In Section 3, we present differentially private algorithms to estimate the range of the data; these algorithms serve as building blocks for estimating differentially private confidence intervals. In Sections 4 and 5, we present (ϵ, δ) differentially private algorithms to estimate an $(1 - \alpha)$ -level confidence interval of μ with known and unknown variance respectively. Section 6 is devoted to lower bounds.

2 Preliminaries

2.1 Notation

We use log to denote natural log to the base e, unless otherwise noted. Random variables are denoted by capital roman letters and their realization by small roman letters. For example X is a random variable and x is its realization. We write $(X_1, \ldots, X_n) \stackrel{iid}{\sim} \mathbb{D}_{\theta,\gamma}$ or equivalently $\underline{X} \sim \mathbb{D}_{\theta}$ or to denote a sample of n independent and identically random variables from the distribution $\mathbb{D}_{\theta,\gamma}$, where $\theta \in \Theta \subseteq \mathbb{R}$ and $\gamma \in \Gamma \subseteq \mathbb{R}^k$ is a vector of nuisance parameters. We sometimes abuse notation and write \mathbb{D}_{θ} instead of $\mathbb{D}_{\theta,\gamma}$. Estimators of parameters θ are denoted by $\hat{\theta} = \hat{\theta}(X_1, \ldots, X_n)$. A differentially private mechanism is denoted by $M(X_1, \ldots, X_n)$ or $M(\underline{X})$.

There are two sources of randomness in our algorithms: The first source of randomness is from the coin flips made by the estimator or algorithm and the dataset is considered fixed. We use the notation of conditioning to denote the probabilities and expectations with respect to the privacy mechanism, when the data is considered to be fixed. Specifically, conditional probability is denoted by $\mathbb{P}(\cdot|X = x)$ and conditional expectation is denoted by $\mathbb{E}[\cdot|X = x]$. The second source of randomness comes from assuming that the dataset is a sample from an underlying distribution $\mathbb{D}_{\theta,\gamma}$. The probability and expectation with respect to this distribution is denoted by $\mathbb{E}_{\underline{X}\sim\mathbb{D}_{\theta}}[\cdot]$, and $\mathbb{P}_{\underline{X}\sim\mathbb{D}_{\theta}}(\cdot)$. While the privacy guarantees are with respect to a fixed dataset, the accuracy guarantees are with respect to both the randomness in the data and the mechanism. In such cases, we state both sources of randomness by writing

$$\mathbb{P}_{\underline{X} \sim \mathbb{D}_{\theta}} \left(M(\underline{X}) \in S \right),$$

where S is any measurable event and the subscripts denote the sources of randomness.

2.2 Differential Privacy

We will present some key properties of differential privacy that we make use of in this paper. One of the attractive properties of Differential Privacy is its ability to compose. To prove the privacy properties of an algorithm, we will rely on the fact that an differentially private algorithm that runs on a dataset and an output of a previous differentially private computation is also differentially private.

Lemma 2.1 (Composition of DP, Dwork et al. (2006)). Let $M_1 : \Omega^n \to \Omega_{M_1}$ be an (ϵ_1, δ_1) differentially private algorithm. Let $M_2 : \Omega^n \times \Omega_{M_1} \to \Omega_{M_2}$ be such that $M_2(\cdot, \omega)$ is an (ϵ_2, δ_2) differentially private algorithm for every fixed $\omega \in \Omega_{M_1}$. Then the algorithm $M(\underline{x}) = M_2(\underline{x}, M_1(\underline{x}))$ is $(\epsilon_1 + \epsilon_2, \delta_1 + \delta_2)$ differentially private.

In many algorithms, we will rely on a basic mechanism for Differential Privacy that works by adding Laplace noise. Let $f: \Omega^n \to \mathbb{R}^k$ be any function of the dataset that we wish to release an approximation of. The *global* sensitivity of f is defined as

$$GS_f = \max_{\underline{x},\underline{x}'neighbors} |f(\underline{x}) - f(\underline{x}')|_1$$

where $|y|_1 = \sum_i |y_i|$ is the l_1 norm of the vector y.

Lemma 2.2 (The Laplace Mechanism, Dwork et al. (2006)). Let $f : \Omega^n \to \mathbb{R}^k$ be a function with global sensitivity at most Δ . The mechanism

$$M(\underline{x}) = f(\underline{x}) + Z$$

is $(\epsilon, 0)$ differentially private, where \underline{x} is the input dataset, Z is a k dimensional random vector where each component of Z is an independent Laplace distribution (defined in 7.2) with mean 0 and scale parameter $b = \Delta/\epsilon$

In many estimators we design, we need a differentially private mechanism for finding a heaviest bin from a (possibly countably infinite) collection of bins, i.e. the bin with the maximum probability mass under the data generating distribution \mathbb{D} . Formally, let B_1, \ldots, B_K be any collection of K disjoint measurable subsets of Ω , which we will refer to as *bins*, where K can be ∞ . The *histogram* of a distribution \mathbb{D} corresponding to the bins

 B_1, \ldots, B_K is given by the vector (p_1, \ldots, p_K) where each $p_k = \underset{X \sim \mathbb{D}}{\mathbb{P}} (X \in B_k)$. In Lemma 2.3, we assert the existence of an differentially private mechanism that on input n iid samples from \mathbb{D} outputs a noisy histogram from which the heaviest bin can be extracted.

Lemma 2.3 (Histogram Learner, following Dwork et al. (2006), Bun et al. (2016), Vadhan (2016)). For every $K \in \mathbb{N} \cup \{\infty\}$, domain Ω , for every collection of disjoint bins B_1, \ldots, B_K defined on Ω , $n \in \mathbb{N}$, $\epsilon, \delta \in (0, 1/n)$, $\beta > 0$ and $\alpha \in (0, 1)$ there exists an (ϵ, δ) -differentially private algorithm $M : \Omega^n \to \mathbb{R}^K$ such that for every distribution \mathbb{D} on Ω , if

1.
$$X_1, \ldots, X_n \stackrel{iid}{\sim} \mathbb{D}, p_k = \mathbb{P}(X_i \in B_k)$$

2. $(\tilde{p}_1, \ldots, \tilde{p}_K) \leftarrow M(X_1, \ldots, X_n), and$

3.

$$n \ge \max\left\{\min\left\{\frac{8}{\epsilon\beta}\log\left(\frac{2K}{\alpha}\right), \frac{8}{\epsilon\beta}\log\left(\frac{4}{\alpha\delta}\right)\right\}, \frac{1}{2\beta^2}\log\left(\frac{4}{\alpha}\right)\right\}$$
(1)

then,

$$\mathbb{P}_{\frac{X}{M}}(|\tilde{p}_{k} - p_{k}| \le \beta) \ge 1 - \alpha \text{ and,}$$
(2)
$$= \left(np_{i} + 2e^{-(\epsilon n/8) \cdot (\max_{k} p_{k})} \quad \text{if } K < 2/\delta \right)$$

$$\mathbb{P}_{\substack{\underline{X}\sim\mathbb{D}\\M}}\left(\arg\max_{k}\tilde{p}_{k}=j\right) \leq \begin{cases} np_{j}+2e^{-(\epsilon n/\delta)\cdot(\max_{k}p_{k})} & \text{if } K<2/\delta\\ np_{j} & \text{if } K\geq2/\delta \end{cases} \tag{3}$$

where the probability is taken over the randomness of M and the data X_1, \ldots, X_n .

Lemma 2.3 asserts the existence of an (ϵ, δ) -differentially private *histogram learner* that takes as input an iid sample and a collection of K disjoint bins, and outputs estimates \tilde{p}_k of p_k , the probability of falling in bin B_k for all k. It has the property that with high probability, the maximum generalization error is at most β . It is important to note that the error is measured with respect to the population (p_k) and not the sample. In particular, the error term includes the noise due to sampling and differential privacy. To bound the generalization error, we follow a standard technique of bounding the generalization error without privacy (i.e. the difference between the sample and the population) and the error introduced for privacy (see for example the equivalence between differentially private query release and differentially private threshold learning in Bun et al. (2015)).

If we take n equal to the RHS of inequality 1 in Lemma 2.3, we obtain a bound on β as a function of n, K, α , and ϵ , namely

$$\beta = \max\left\{\min\left\{\frac{8}{\epsilon n}\log\left(\frac{2K}{\alpha}\right), \frac{8}{\epsilon n}\log\left(\frac{4}{\alpha\delta}\right)\right\}, \sqrt{\frac{1}{2n}\log\left(\frac{4}{\alpha}\right)}\right\}$$
(4)

The last term, $\Theta\left(\sqrt{\log(1/\alpha)/n}\right)$ is the sampling error, which is incurred even without privacy. For privacy, we incur an error that is the *minimum* of two terms: $\mathcal{O}\left(\log(K)/(\epsilon n)\right)$ and $\mathcal{O}\left(\log(1/\delta)/(\epsilon n)\right)$. Note that these two terms vanish linearly in n, faster than the sampling error, which vanishes as $\mathcal{O}\left(1/\sqrt{n}\right)$. Moreover, the dependence on the number of bins K is only logarithmic or in case of $\delta > 0$, even non-existent. When $\delta > 0$, the choice of $K = \infty$ allows us to construct (ϵ, δ) -DP algorithms that have no dependence on the range of the parameters.

We will use the histogram learner to obtain the largest noisy bin from a possibly infinite collection of bins. Hence, apart from a bound on the maximum generalization error, we also need a bound on the probability of picking the wrong bin as the largest bin. Lemma 2.3 asserts that the probability of choosing any bin j as the largest bin is roughly upper bounded by np_j , the expected number of points falling in bin j. As before, this probability is over the sampling and the noise added by the differential privacy mechanism. Note that this bound is useful only when p_j is small, in particular $p_j \ll 1/n$. Hence, the theorem bounds the probability of incorrectly choosing a bin that has very few expected points as the largest bin.

Proof of Lemma 2.3. Let $C_k = \sum_{i=1}^n I(X_i \in B_k)$ be the number of points that fall in bin k and $\hat{p}_k = C_k/n$ be the corresponding proportion of points. The distribution learner operates as follows: When $K < 2/\delta$, it uses an $(\epsilon, 0)$ -DP algorithm, and when $K \ge 2/\delta$, it uses an (ϵ, δ) -DP algorithm to output a noisy histogram.

The key idea behind the proof of the existence of a histogram learner is the following. There exist basic $(\epsilon, 0)$ and (ϵ, δ) differentially private mechanisms M with the property that on input $\underline{X} = (X_1, \ldots, X_n)$ and bins B_1, \ldots, B_K , they output $(\tilde{p}_1, \ldots, \tilde{p}_K) = M(\underline{X})$, such that if

$$n \geq \min\left\{\frac{8}{\epsilon\beta}\log\left(\frac{2K}{\alpha}\right), \frac{8}{\epsilon\beta}\log\left(\frac{4}{\alpha\delta}\right)\right\}$$

then for every \underline{x} ,

$$\mathbb{P}_{M}\left(\max_{k}\left|\tilde{p}_{k}-\hat{p}_{k}\right| \geq \beta \mid \underline{X}=\underline{x}\right) \leq \alpha/2$$

That is, with high probability, there is a small difference between the differentially private output \tilde{p}_k and the empirical estimates \hat{p}_k . Moreover, the Dvoretzky-Kiefer-Wolfowitz inequality (Massart, 1990) tells us that with high probability the empirical estimates are close to the population parameters:

$$\mathbb{P}_{\underline{X} \sim \mathbb{D}}\left(\max_{k} |\hat{p}_{k} - p_{k}| > \beta\right) \le 2\exp(-2n\beta^{2})$$

So, if $n \ge (1/2\beta^2) \cdot \log(4/\alpha)$, then $\mathbb{P}(\max_k |\hat{p}_k - p_k| > \beta) \le \alpha/2$. Thus, by a union bound, if

$$n \ge \max\left\{\min\left\{\frac{8}{\epsilon\beta}\log\left(\frac{2K}{\alpha}\right), \frac{8}{\epsilon\beta}\log\left(\frac{4}{\alpha\delta}\right)\right\}, \frac{1}{2\beta^2}\log\left(\frac{4}{\alpha}\right)\right\}$$

then,

$$\frac{\mathbb{P}}{\underline{X}_{M}} \left(\max_{k} |\tilde{p}_{k} - p_{k}| \geq \beta \right) \\
\leq \underbrace{\mathbb{E}}_{\underline{X} \sim \mathbb{D}} \left[\mathbb{P}_{M} \left(\max_{k} |\tilde{p}_{k} - \hat{p}_{k}| \geq \beta \mid \underline{X} = \underline{x} \right) \right] + \underbrace{\mathbb{P}}_{\underline{X} \sim \mathbb{D}} \left(\max_{k} |\hat{p}_{k} - p_{k}| \geq \beta \right) \\
\leq \alpha/2 + \alpha/2.$$

Now we review the two differentially private algorithms we use and also prove the additional claim regarding the probability of any bin being selected as the maximum.

 $(\epsilon, 0)$ case: We will first start with the $(\epsilon, 0)$ differentially private algorithm which we use when $K < 2/\delta$. Consider the Laplace mechanism, given in Lemma 2.2 applied to the *empirical histogram*. The *empirical histogram* of a dataset $\underline{x} = (x_1, \ldots, x_n)$ on bins B_1, \ldots, B_K is given by $f_{B_1,\ldots,B_K}(\underline{x}) = (\hat{p}_1, \ldots, \hat{p}_K)$ Note that the global sensitivity of f_{B_1,\ldots,B_K} is 2/n. Hence we can release the empirical histogram by adding Laplace noise with scale $b = 2/(\epsilon n)$ to each \hat{p}_k . Formally, let $\tilde{p}_k = \hat{p}_k + Z_k$ where $Z_k \sim Lap(0, 2/(\epsilon n))$. If $n \geq 2\log(2K/\alpha)/(\epsilon\beta)$, then,

$$\mathbb{P}_{M}\left(\max_{k} |\tilde{p}_{k} - \hat{p}_{k}| \leq \beta \mid \underline{X} = \underline{x}\right) = \mathbb{P}\left(\max_{k} |Z_{k}| \leq \beta\right) \\
= (1 - \exp\left(-\epsilon n\beta/2\right))^{K} \\
\geq 1 - K \exp\left(-\epsilon n\beta/2\right) \geq 1 - \alpha/2$$

where the second line follows from the tails of a Laplace distribution, see Proposition 7.2. Now, let us prove that

$$\mathbb{P}_{\underline{X} \sim \mathbb{D}}\left(\arg\max_{k} \tilde{p}_{k} = j\right) \le np_{j} + 2e^{-(\epsilon n/8) \cdot (\max_{k} p_{k})}.$$

Let $p_r = \max_k p_k$. Consider the event $\{\arg \max_k \tilde{p}_k = j\}$. Note that \tilde{p}_k is continuous. Hence the probability of ties is 0 and the argmax is unique and well defined. This implies that $\{\tilde{p}_j \ge \tilde{p}_r\}$. In turn, this implies that either, $\hat{p}_j > 0$, or $Z_j \ge p_r/4$ or $\hat{p}_r < p_r/2$ or $Z_r \le -p_r/4$. Hence, by a union bound,

$$\mathbb{P}\left(\arg\max_{k}\tilde{p}=j\right) \leq \mathbb{P}\left(\hat{p}_{j}>0\right) + \mathbb{P}\left(Z_{j}\geq p_{r}/4\right) + \mathbb{P}\left(\hat{p}_{r}< p_{r}/2\right) + \mathbb{P}\left(Z_{r}\leq -p_{r}/4\right)$$
$$\leq 1 - (1-p_{j})^{n} + \frac{1}{2}\exp\left(-\frac{\epsilon np_{r}}{8}\right) + \exp\left(-\frac{np_{r}}{8}\right) + \frac{1}{2}\exp\left(-\frac{\epsilon np_{r}}{8}\right)$$
$$\leq np_{j} + 2\exp\left(-\frac{\epsilon n}{8}\cdot\max_{k}p_{k}\right),$$

where we have used the Chernoff bound (Proposition 7.1), and a tail bound for a Laplace random variable from (Proposition 7.2).

 (ϵ, δ) Case: For the case that $K > 2/\delta$, we will use an (ϵ, δ) -differentially private algorithm, called *stability-based* histogram (Bun et al., 2016) to estimate \hat{p}_k , which removes the dependence on K by allowing for K to be infinity. Specifically, the algorithm on input x_1, \ldots, x_n and B_1, \ldots, B_K , where K is possibly ∞ runs as follows:

- 1. Let $\hat{p}_k = \sum_{i=1}^n I(x_i \in B_k)/n$
- 2. If $\hat{p}_k = 0$, set $\tilde{p}_k = 0$
- 3. If $\hat{p}_k > 0$,
 - (a) Let $\tilde{p}_k = \hat{p}_k + Z_k$, where $Z_k \sim Lap(0, 2/(\epsilon n))$.
 - (b) Let $t = 2\log(2/\delta)/(\epsilon n) + (1/n)$
 - (c) If $\tilde{p}_k < t$, set $\tilde{p}_k = 0$
- 4. Output $\tilde{p}_k, k = 1, \ldots, K$

A proof of (ϵ, δ) differential privacy of this algorithm can be found in Theorem 3.5 of Vadhan (2016). For utility, we will show that if

$$n \ge \frac{8}{\epsilon\beta} \log\left(\frac{4}{\delta\alpha}\right)$$

then $\mathbb{P}(|\tilde{p}_k - \hat{p}_k| > \beta \mid \underline{X} = \underline{x}) \leq \alpha/2$. Note that for any k such that $\hat{p}_k > 0$, we have,

$$\mathbb{P}\left(\left|\tilde{p}_{k}-\hat{p}_{k}\right|>\beta\mid\underline{X}=\underline{x}\right) = \mathbb{P}\left(\left|\tilde{p}_{k}-\hat{p}_{k}\right|>\beta,\hat{p}_{k}+Z_{k}>t\mid\underline{X}=\underline{x}\right) + \mathbb{P}\left(\left|\tilde{p}_{k}-\hat{p}_{k}\right|>\beta,\hat{p}_{k}+Z_{k}\beta,\hat{p}_{k}+Z_{k}>t\mid\underline{X}=\underline{x}\right) + \mathbb{P}\left(\hat{p}_{k}>\beta,\hat{p}_{k}+Z_{k}\beta\right) + \mathbb{P}\left(\hat{p}_{k}>\beta,\hat{p}_{k}+Z_{k}\beta\right) + \mathbb{P}\left(Z_{k}$$

where the last line holds if $\beta > 2t$ and if

$$n \ge \frac{4}{\epsilon\beta} \log\left(\frac{4}{\alpha\delta}\right)$$

For $\beta > 2t$, we need

$$n \ge \frac{4}{\epsilon\beta} \log\left(\frac{2}{\delta}\right) + \frac{2}{\beta}.$$

Since $\alpha < 1$, it suffices to have

$$n \ge \frac{8}{\epsilon\beta} \log\left(\frac{4}{\delta\alpha}\right).$$

There are at most n points such that $\hat{p}_k > 0$. Also, for the points where $\hat{p}_k = 0$, we have $\tilde{p}_k = 0$. Hence for such points, $\mathbb{P}(|\tilde{p}_k - \hat{p}_k| > \beta \mid \underline{X} = \underline{x}) = 0$. Thus by a union bound we have,

$$\mathbb{P}\left(\max_{k} |\tilde{p}_{k} - \hat{p}_{k}| > \beta \mid \underline{X} = \underline{x}\right) \le n\alpha\delta/2 < \alpha/2$$

since $\delta < 1/n$.

Finally, we need to prove that $\mathbb{P}(\arg \max_k \tilde{p}_k = j) \leq np_j$ where the probability is over the randomness of the data and the mechanism. In the stability based algorithm, if $\tilde{p}_k = 0$ for all k, the largest bin is undefined, and we set $\arg \max_k \tilde{p}_k$ as \perp . On the other hand, if $\arg \max_k \tilde{p}_k = j$, then $\hat{p}_j > 0$. Hence $\mathbb{P}(\arg \max \tilde{p}_k = j) \leq \mathbb{P}(\hat{p}_j > 0) = 1 - (1 - p_j)^n \leq np_j$.

3 A differentially private estimate of the range of Gaussian random variables

In this section, we present differentially private algorithms to estimate the range of an iid sample from a Gaussian distribution with known and unknown variances. These algorithms serve as building blocks for estimating the confidence intervals, but may also be of independent interest.

Let X_1, \ldots, X_n be iid samples from a normal distribution with mean μ and variance σ^2 . Our algorithms don't make any assumption on the boundedness of the data. The $(\epsilon, 0)$ DP algorithms assume that μ and σ^2 lie in a bounded domain and the (ϵ, δ) algorithms make no boundedness assumptions. Under these assumptions, the algorithms below output a range with the guarantee that with high probability, the range includes all the data points, and if the sample size is large enough, the estimated range is at most a constant factor larger than the true range. We will begin with the simpler case of estimation of range when the variance is known, followed by the case when the variance is unknown.

3.1 Estimation of range with known variance

Theorem 3.1. For every $n \in \mathbb{N}$, $\sigma, \epsilon, \delta > 0$, $\alpha \in (0, 1/2)$, $R \in (0, \infty]$, there is a w > 0 and an (ϵ, δ) -differentially private algorithm $M : \mathbb{R}^n \to \mathbb{R} \times \mathbb{R}$ such that whenever $\mu \in (-R, R)$ and

$$n \ge c \min\left\{\frac{1}{\epsilon}\log\left(\frac{R}{\sigma\alpha}\right), \frac{1}{\epsilon}\log\left(\frac{1}{\delta\alpha}\right)\right\},\$$

(where c is a universal constant), if X_1, \ldots, X_n are iid Gaussian random variables with mean μ and variance $\sigma_0^2 \leq \sigma^2$ (where σ^2 is known) and

$$(X_{\min}, X_{\max}) \leftarrow M(X_1, \ldots, X_n),$$

we have:

$$\mathbb{P}_{\substack{\underline{X} \sim N(\mu, \sigma_0^2) \\ M}} (\forall i \ X_{\min} \le X_i \le X_{\max}) \ge 1 - \alpha.$$

and with probability 1,

$$|X_{\max} - X_{\min}| = w = \mathcal{O}\left(\sigma\sqrt{\log(n/\alpha)}\right)$$

Theorem 3.1 asserts the existence of a differentially private algorithm that takes as input n iid samples from a normal distribution and outputs a range $[X_{\min}, X_{\max}]$ with a guarantee that with high probability all points are included in the range. Note that when $\delta = 0$, n needs to be at least $c(1/\epsilon) \log(R/(\sigma \alpha))$. Hence, when $\delta = 0$, R needs to be finite i.e., we need μ to be bounded. On the other hand, when $\delta > 0$, R can be ∞ . Hence Theorem 3.1 asserts that when $\delta > 0$, there is no dependence on the range of μ . Moreover, the range of the data $|X_{\max} - X_{\min}|$ estimated by the algorithm is within a constant factor of the true range. This is because, it is well known that for Gaussian data, $\mathbb{E}[\max_i X_i] = \Theta(\sigma\sqrt{\log n})$, and in fact, $\max_i X_i \ge \sigma\sqrt{\log(n/\alpha)}$ with probability $\Omega(\alpha)$.

Proof of Theorem 3.1. Consider the following algorithm.

Algorithm 1 Differentially Private estimate of range with known variance

Input: $X_1, \ldots, X_n, \epsilon, \alpha, R \in (0, \infty], \sigma$.

Output: An (ϵ, δ) differentially private estimate of the range of X_1, \ldots, X_n .

- 1: Let $r = \lceil R/\sigma \rceil$. Divide $[-R \sigma/2, R + \sigma/2]$ into 2r + 1 bins of length at most σ each in the following manner a bin B_j equals $((j 0.5)\sigma, (j + 0.5)\sigma]$, for $j \in \{-r, \ldots, r\}$.
- 2: Run the histogram learner of Lemma 2.3 with privacy parameters (ϵ, δ) and bins B_{-r}, \ldots, B_r on input X_1, \ldots, X_n to obtain noisy estimates $\tilde{p}_{-r}, \ldots, \tilde{p}_r$. Let the largest noisy bin be \hat{l} , i.e.

$$\hat{l} = \operatorname*{argmax}_{j=-r,\dots,r} \tilde{p}_j$$

3: Output (X_{\min}, X_{\max}) , where

$$X_{\min} = \sigma \hat{l} - 4\sigma \sqrt{\log\left(\frac{n}{\alpha}\right)}, X_{\max} = \sigma \hat{l} + 4\sigma \sqrt{\log\left(\frac{n}{\alpha}\right)}$$

Let $\sigma^* \leq \sigma$ be the standard deviation of X_1 . We will prove the following two claims: Claim 1: If

$$n \ge c \min\left\{\frac{1}{\epsilon}\log\left(\frac{R}{\sigma\alpha}\right), \frac{1}{\epsilon}\log\left(\frac{1}{\delta\alpha}\right)\right\},\$$

then with probability at least $1 - \alpha/2$, we have

$$|\mu - \hat{l}\sigma| \le 2\sigma.$$

Claim 2: With probability at least $1 - \alpha/2$, we have

$$\left\{ \forall i : |X_i - \mu| \le \sigma^* \sqrt{2 \log\left(\frac{4n}{\alpha}\right)} \right\}.$$

By Claims 1, 2, a union bound and the fact that $\sigma^* \leq \sigma$, we have, with probability at least $1 - \alpha$, for all *i*,

$$|X_i - \hat{l}\sigma| \le |X_i - \mu| + |\hat{l}\sigma - \mu| \le \sigma^* \sqrt{2\log\left(\frac{4n}{\alpha}\right)} + 2\sigma \le 4\sigma \sqrt{\log\left(\frac{n}{\alpha}\right)}$$

which gives the result. We will now prove the two claims.

Proof of Claim 1 Let B_j be the j^{th} bin and $p_j = \mathbb{P}(X_i \in B_j)$. Let $p_{(1)} \ge p_{(2)} \ge p_{(3)} \ge \ldots$ be sorted p_j 's and let $j_{(1)}, j_{(2)}, \ldots$ be the corresponding bins, the tie breaking rule will be described below. Consider the event $E = \{\hat{l} = j_{(1)} \text{ or } j_{(2)}\}$. We will first show that bins with largest and second largest mass are adjacent to each other, and μ always lies in a bin with the largest mass. This will imply that under the event $E, |\hat{l} - \mu| \le 2\sigma$.

We begin by finding the bins with the largest, second largest and third largest mass. Let $\Phi(\cdot)$ denote the cdf of a standard normal distribution. Note that

$$p_j = \mathbb{P} \left(X_i \in \left((j - 0.5)\sigma, (j + 0.5)\sigma \right] \right)$$
$$= \Phi \left(\frac{(j + 0.5)\sigma - \mu}{\sigma^*} \right) - \Phi \left(\frac{(j - 0.5)\sigma - \mu}{\sigma^*} \right)$$
$$= f \left(j \frac{\sigma}{\sigma^*} - \frac{\mu}{\sigma^*} \right),$$

where $f(\gamma) = \Phi\left(\gamma + \frac{\sigma}{2\sigma^*}\right) - \Phi\left(\gamma - \frac{\sigma}{2\sigma^*}\right)$. One can verify that $f(\gamma)$ is maximized at $\gamma = 0$, is symmetric and decreasing with $|\gamma|$. Hence maximizing p_j on integers amounts to minimizing $|\gamma| = |j(\sigma/\sigma^*) - \mu/\sigma^*|$, or equivalently minimizing $|j - \mu/\sigma|$. So we are sorting the integers j according to their distance from μ/σ . Thus we can take $j_{(1)} = \lceil \mu/\sigma \rfloor$ where $\lceil x \rfloor$ is x rounded to the nearest integer, where we break ties by rounding down.

Notice that $\mu \in ((j_{(1)} - 0.5)\sigma, (j_{(1)} + 0.5)\sigma] = B_{j_{(1)}}$. To determine $j_{(2)}$ and $j_{(3)}$, we consider the cases whether $j_{(1)} = \lceil \mu/\sigma \rceil$ or $j_{(1)} = \lfloor \mu/\sigma \rfloor$.

- 1. If $j_{(1)} = \lceil \mu/\sigma \rceil$, then we can take $j_{(2)} = \lceil \mu/\sigma \rceil 1$ and $j_{(3)} = \lceil \mu/\sigma \rceil + 1$.
- 2. If $j_{(1)} = \lfloor \mu/\sigma \rfloor$, then we can take $j_{(2)} = \lfloor \mu/\sigma \rfloor + 1$ and $j_{(3)} = \lfloor \mu/\sigma \rfloor 1$.

Notice that in either case, $|j_{(1)} - j_{(2)}| = 1$, which implies that the largest and the second largest bins are always adjacent to each other, and $|j_{(3)} - \mu/\sigma| = |j_{(1)} - \mu/\sigma| + 1$. Hence we have,

$$g = p_{(1)} - p_{(3)} = f\left(\left|j_{(1)} - \frac{\mu}{\sigma}\right|\frac{\sigma}{\sigma^*}\right) - f\left(\left|j_{(3)} - \frac{\mu}{\sigma}\right|\frac{\sigma}{\sigma^*}\right)$$
$$\geq f\left((1/2) \cdot (\sigma/\sigma^*)\right) - f\left((3/2) \cdot (\sigma/\sigma^*)\right)$$

(Since $f(x) - f(x + \sigma/\sigma^*)$ is decreasing for $x \in [0, \infty)$ and $|j_{(1)} - \mu/\sigma| \le 1/2$),

$$= \Phi(\sigma/\sigma^*) - \Phi(0) - \Phi(2\sigma/\sigma^*) + \Phi(\sigma/\sigma^*)$$

$$\ge 2\Phi(1) - \Phi(0) - \Phi(2)$$

(Since $2\Phi(x) - \Phi(0) - \Phi(2x)$ is a decreasing function for $x \in (0, \infty)$ and $\sigma/\sigma^* \leq 1$

 ≥ 0.1 (By explicit calculation.)

Hence under the event $E = \{\hat{l} = j_{(1)} \text{ or } j_{(2)}\}, |\hat{l} - \mu| \leq 2\sigma.$

Next we will lower bound the probability of event E. If $\max_k |\tilde{p}_k - p_k| < g/2$, then the largest noisy bin \hat{l} is either $j_{(1)}$ or $j_{(2)}$. Since $\mu \in B_{(j_{(1)})}$, and the bin width is σ , we have $|\hat{l}\sigma - \mu| \leq 2\sigma$. Applying Lemma 2.3, with $\beta = g$, we get, if

$$n \ge \mathcal{O}\left(\min\left\{\frac{\log\left(\frac{K}{\alpha}\right)}{g\epsilon}, \frac{\log\left(\frac{1}{\alpha\delta}\right)}{g\epsilon}\right\}\right)$$

then with probability at least $1 - \alpha/2$, we have $\max_k |\tilde{p}_k - p_k| < g/2$. Here $K = 2R/\sigma$ is the number of bins.

Proof of Claim 2: From Proposition 7.2 due to Gaussian tails we have that for all *i*,

$$\mathbb{P}(|X_i - \mu| > c) \le 2e^{-c^2/2{\sigma^*}^2}$$

By applying a union bound, we get,

$$\mathbb{P}\left(\exists i | X_i - \mu| \ge c\right) \le 2ne^{-c^2/2{\sigma^*}^2}$$

Letting $c = \sigma^* \sqrt{2 \log (4n/\alpha)}$ gives the desired result.

3.2 An estimate of variance

In this section, we will construct a differentially private estimator of the variance of data from a Gaussian distribution with unknown variance. This estimate will be used for finding the range of a sample with unknown variance.

Theorem 3.2. For every $n \in \mathbb{N}$, $\sigma_{\min} < \sigma_{\max} \in [0, \infty]$, $\epsilon, \delta > 0$, $\alpha \in (0, 1/2)$, there is an (ϵ, δ) -differentially private algorithm $M : \mathbb{R}^n \to [0, \infty)$ such that if X_1, \ldots, X_n are iid Gaussian random variables with mean μ and with variance $\sigma^2 \in (\sigma_{\min}^2, \sigma_{\max}^2)$, and

$$\hat{\sigma} \leftarrow M(X_1, \ldots, X_n),$$

we have:

1. High probability bound: If

$$n \ge c \min\left\{\frac{1}{\epsilon} \log\left(\frac{\log\left(\frac{\sigma_{\max}}{\sigma_{\min}}\right)}{\alpha}\right), \frac{1}{\epsilon} \log\left(\frac{1}{\delta\alpha}\right)\right\},\$$

(where c is a universal constant),

$$\mathbb{P}_{\substack{\underline{X} \sim N(\mu, \sigma^2) \\ M}} \left(\sigma \le \hat{\sigma} \le 8\sigma \right) \ge 1 - \alpha$$

2. Expectation bound: If

$$n \ge c \min\left\{\frac{1}{\epsilon}\log\left(\frac{\sigma_{\max}}{\sigma_{\min}}\right), \frac{1}{\epsilon}\log\left(\frac{1}{\delta\alpha}\right)\right\},\$$

then

$$\mathbb{E}_{\substack{\underline{X} \sim N(\mu, \sigma^2) \\ M}} \left[\hat{\sigma}^2 \right] \le \sigma^2 \cdot \left(c_1 + c_2 \log^2(n) \cdot \alpha \right)$$

for some universal constants c_1 and c_2 .

As before, note that when $\delta > 0$, σ_{\max} can be set to ∞ and σ_{\min} can be set to 0, and the required sample complexity remains finite. However, when $\delta = 0$, we need σ_{\min} and σ_{\max} to be bounded away from 0 and ∞ respectively. Theorem 3.2 asserts that with high probability the estimate of σ is not too far from the true σ . Moreover, we can also obtain a bound on the expected value of $\hat{\sigma}^2$ with an increase in sample complexity (to a logarithmic dependence on $\sigma_{\min}/\sigma_{\max}$ rather than doubly-logarithmic). For instance, if we set $\alpha = \alpha_0/\log^2 n$, we get $\mathbb{E}[\hat{\sigma}^2] = \mathcal{O}(\sigma^2)$. Reduction of $\log^2 n$ in α amounts to an extra requirement that $n \ge c \cdot \log^2 n/\epsilon$, i.e., $n \ge \tilde{\mathcal{O}}(1/\epsilon)$. The upper bound on $\mathbb{E}[\hat{\sigma}^2]$ will be used to derive the upper bound on the expected length of the width of the confidence interval in Theorem 5.1.

Proof of Theorem 3.2. We claim that Algorithm 2 given below is (ϵ, δ) differentially private with the required properties.

Algorithm 2

Input: $X_1, \ldots, X_n, \epsilon, R, \sigma_{\min}, \sigma_{\max}, \alpha$.

Output: An (ϵ, δ) differentially private approximate estimate $\hat{\sigma}$ of the standard deviation of X_1, \ldots, X_n . 1: If

$$n < c \min\left\{\frac{1}{\epsilon} \log\left(\frac{\log\left(\frac{\sigma_{\max}}{\sigma_{\min}}\right)}{\alpha}\right), \frac{1}{\epsilon} \log\left(\frac{1}{\delta\alpha}\right)\right\},\$$

output \perp .

- 2: Divide the positive half of the real line into bins of exponentially increasing length. The bins are of the form B_j = (2^j, 2^{j+1}] for j = j_{min},..., j_{max}. where j_{max} = ⌈log₂ σ_{max}\rceil + 1 and j_{min} = ⌊log₂ σ_{min}⌋ 2.
 3: Let Y_i = X_{2i} X_{2i-1} for i = 1,..., |n/2|
- 4: Run the histogram learner of Lemma 2.3 with budget (ϵ, δ) and bins $B_{j_{\min}}, \ldots, B_{j_{\max}}$ on input $|Y_1|, \ldots, |Y_{\lfloor n/2 \rfloor}|$ to obtain noisy estimates $\tilde{p}_{j_{\min}}, \ldots, \tilde{p}_{j_{\max}}$. Let the largest noisy bin be \hat{l} , i.e.

$$\hat{l} = \operatorname*{argmax}_{j} \tilde{p}_{j}$$

5: Output $\hat{\sigma} = 2^{\hat{l}+2}$

Proof of Part (1): Let us start by proving Part (1). The proof will be based on the following two informal steps:

Step 1: The bins with largest or second largest probability mass are always the bin containing σ or the bin next to the bin containing σ .

Step 2: Under the right conditions, the largest noisy bin \hat{l} is the either the bin with largest or the second largest mass.

Let us start with some simple facts. Since $\sigma \in (\sigma_{\min}, \sigma_{\max})$, there exists a bin B_l with label $l \in (\lfloor \log_2 \sigma_{\min} \rfloor - 1, \lceil \log_2 \sigma_{\max} \rceil)$ such that $\sigma \in (2^l, 2^{(l+1)}] = B_l$. Hence we can write $\sigma = 2^{l+c}$, for some $c \in (0, 1]$. Next, note that $Y_i \sim N(0, 2\sigma^2)$ for $i = 1, \ldots, \lfloor n/2 \rfloor$. Following the statement of Lemma 2.3, define

$$p_j = \mathbb{P}\left(|Y_i| \in B_j\right)$$

Sort the p_j 's as $p_{(1)} \ge p_{(2)} \ge p_{(3)} \ge \dots$ and let $j_{(1)}, j_{(2)}, j_{(3)}, \dots$ be the corresponding bins, where the tie breaking rules will be specified below.

Claim 1: The bins corresponding to the largest and second largest mass $p_{(1)}, p_{(2)}$ are $(j_{(1)}, j_{(2)}) \in \{(l, l-1), (l, l+1), (l+1, l)\}$.

Claim 2: $p_{(1)} - p_{(3)} > \frac{1}{100}$.

Deferring the proof of these claims, let us prove the Part (1). Consider the following event

$$E = \{\hat{l} \in (j_{(1)}, j_{(2)})\}.$$

Under E, from Claim 1, it follows that $\hat{\sigma} = 2^{\hat{l}+2}$ will be $2^{l+1}, 2^{l+2}$, or 2^{l+3} . Since $\sigma = 2^{l+c}$, for some $c \in (0, 1]$, we have, $\sigma \leq \hat{\sigma} \leq 8\sigma$. To finish the proof of Part (1), we need to lower bound the probability of E. This follows from the properties of the histogram learner of Lemma 2.3. Let $g = p_{(1)} - p_{(3)}$. If $\max_i |\tilde{p}_i - p_i| < g/2$, then the largest noisy bin is either $p_{(1)}$ or $p_{(2)}$. From Lemma 2.3 this happens with probability at least $1 - \alpha$ if

$$n \ge \mathcal{O}\left(\min\left\{\frac{\log\left(\frac{K}{\alpha}\right)}{g\epsilon}, \frac{\log\left(\frac{1}{\alpha\delta}\right)}{g\epsilon}\right\}\right)$$

where $K = c \cdot \log_2 (\sigma_{\max} / \sigma_{\min})$ is the number of bins. We will now prove the two claims. For any $\Delta \in \mathbb{R}$, define

$$q(\Delta) = \phi\left(2^{\Delta+0.5}\right) - \phi\left(2^{\Delta-0.5}\right)$$

and note that,

$$p_j = 2\left(\Phi\left(\frac{2^{j+1}}{\sqrt{2}\sigma}\right) - \Phi\left(\frac{2^j}{\sqrt{2}\sigma}\right)\right)$$
$$= 2q\left(j-l-c\right)$$

where $\Phi(\cdot)$ is the cdf of a standard normal distribution.

Proof of Claim 1. One can verify that when $\Delta \in \mathbb{R}$,

$$\frac{dq(\Delta)}{d\Delta} = g(\Delta) \left(1 - \frac{1}{2} \exp\left(3 \cdot 2^{2(\Delta - 1)}\right) \right)$$
(5)

where $g(\Delta) > 0 \ \forall \Delta$. Let $d = 1 + 0.5 \log_2((\log 2)/3)$. The derivative is 0 when $\Delta = d$. $q(\Delta)$ is increasing when $\Delta < d$ and decreasing when $\Delta > d$, and the maximum occurs at $\Delta = d$. Since $p_j = 2q(j - l - c)$, and $j \in \{j_{\min}, \ldots, j_{\max}\}$, the largest value of p_j occurs at $j = l + \lceil c + d \rceil$ or $l + \lceil c + d \rceil - 1$. Since $c \in (0, 1]$ and $d \approx -0.056$, $\lceil c + d \rceil \in \{0, 1\}$ and the possible values of j where the maximum occurs are either l or l - 1 or l+1. One can verify that for all $c \in (0, 1]$, $p_l > p_{l-1}$ (For example, note that $p_l - p_{l-1} = 2q(-c) - q(-1-c)$ and q(-c) - q(-1-c) is a monotonic function for $c \in (0, 1]$ and positive for c = 0 and c = 1). Hence the maximum occurs at either l or l + 1.

Now let us find the second largest bin. Note that due to the concavity of $q(\Delta)$, the second largest bin must be adjacent to the largest bin. When the largest bin is l, the second largest bin can be either l-1 or l+1. Similarly, when the largest bin is l+1, the second largest bin can be either l or l+2. One can verify that for all $c \in (0,1]$, $p_{l+2} < p_l$. Hence the largest and the second largest bins $(j_{(1)}, j_{(2)})$ are always the pairs (l, l-1), (l, l+1), (l+1, l) as desired.

Proof of Claim 2. Note that

$$p_{(1)} \ge p_l = 2q(-c) = 2\left[\Phi(2^{0.5-c}) - \Phi(2^{-0.5-c})\right]$$

Also for any k_1, k_2 , we have $p_{(3)} \leq \max_{k \neq k_1, k_2} p_k$. When $c \in (0, 1/2]$, we will use $k_1 = l, k_2 = l - 1$ and when $c \in (1/2, 1]$ we will use $k_1 = l, k_2 = l + 1$. This gives us the following:

Case 1: When $c \in (0, 1/2]$,

$$p_{(3)} \le \max_{t \in \mathbb{Z} \setminus \{0, -1\}} 2q(t-c)$$

By inspecting the derivative (equation 5), one can verify that the maximum occurs when t = 1. Hence we have,

$$p_{(1)} - p_{(3)} \ge \min_{c \in (0, 1/2]} 2 \left[\phi(2^{1/2-c}) - \phi(2^{-1/2-c}) - \phi(2^{3/2-c}) + \phi(2^{1/2-c}) \right]$$

By taking the derivative one can verify that this function is decreasing in c and the minimum occurs at c = 1/2and $g \ge 0.0139$.

Case 2: When $c \in (1/2, 1]$,

$$p_{(3)} \le \max_{t \in \mathbb{Z} \setminus \{0,1\}} 2q(t-c)$$

By inspecting the derivative (equation 5), one can verify that the maximum occurs when t = 2. Hence we have,

$$g = p_{(1)} - p_{(3)}$$

$$\geq \min_{c \in (1/2, 1]} 2 \left[\phi(2^{1/2-c}) - \phi(2^{-1/2-c}) - \phi(2^{5/2-c}) + \phi(2^{3/2-c}) \right]$$

By taking the derivative one can verify that this function is decreasing in c and the minimum occurs at c = 1and $g \ge 0.091$. In both cases, $g \ge 0.01$

Proof of Part (2): Next we need to show that when

$$n \ge (c/\epsilon) \cdot \min\{\log\left(\sigma_{\max}/\sigma_{\min}\right), \log(1/\alpha\delta)\},\tag{6}$$

we have, $\mathbb{E}\left[\hat{\sigma}^2\right] \leq \sigma^2 \cdot (c_1 + c_2 \log^2(n) \cdot \alpha)$. Note that under the Assumption 6 on *n*, the results of Part (1) apply, since this sample complexity is larger than what is needed for the high probability bound. Let $q_i = \mathbb{P}\left(\arg\max_j \tilde{p}_j = i\right)$. Recall that $\sigma = 2^{l+c}$ for some $c \in (0, 1]$.

$$\mathbb{E}\left[\hat{\sigma}^{2}\right] = \sum_{i=j_{\min}}^{i=j_{\max}} (2^{i+2})^{2} \cdot \mathbb{P}\left(\arg\max_{j} \tilde{p}_{j} = i\right)$$
$$= 16\left(\sum_{i \le l+3} 2^{2i}q_{i} + \sum_{l+3 < i \le l+\log_{2}(\log n)} 2^{2i}q_{i} + \sum_{i>l+\log_{2}(\log n)} 2^{2i}q_{i}\right)$$
$$= 16\left(A + B + C\right)$$

Now we will bound each of these three terms:

$$A = \sum_{i \le l+3} 2^{2i} q_i \le 2^{2(l+3)} \cdot \left(\sum_{i \le l+3} q_i\right) \le 64\sigma^2 \cdot 1$$
$$B = \sum_{l+3 < i \le l+\log \log n} 2^{2i} q_i \le 2^{2(l+\log \log n)} \cdot \left(\sum_{l+3 < i < l+\log \log n} q_i\right)$$
$$\le \sigma^2 \log^2 n \cdot \mathbb{P}\left(\hat{\sigma} > 8\sigma\right) \le \sigma^2 \log^2 n \cdot \alpha$$

where we have used Part (1) to bound $\mathbb{P}(\hat{\sigma} > 8\sigma)$ by α . To bound the last summand C we will consider two cases. First note that if $X \sim N(0, \sigma^2)$, we have.

$$p_i = \mathbb{P}\left(|X| \in (2^i, 2^{i+1}]\right) < \mathbb{P}\left(|X| > 2^i\right) \le 2\exp(-2^{2i}/2\sigma^2) \le 2\exp(-2^{2(i-l-1)}/2).$$

Let $K = j_{\text{max}} - j_{\text{min}}$ be the number of bins.

Case 1: $\delta \geq 2/K$ From Lemma 2.3, when $\delta \geq 2/K$,

$$q_i \le np_i \le 2n \exp(-2^{2(i-l)-3})$$

Hence

$$\begin{split} C &= \sum_{i>l+\log\log n} 2^{2i} q_i < 2n \cdot \left(\sum_{i>l+\log\log n} 2^{2i} \exp(-2^{2(i-l)-3})\right) \\ &\leq 2n\sigma^2 \left(\sum_{t>\log\log n} 2^{2t} \exp(-2^{2t-3})\right) \\ &\leq 2n\sigma^2 \int_{\log n}^{\infty} k^2 \exp(-k^2/8) dk \\ &\leq \sigma^2, \text{ for sufficiently large } n, \end{split}$$

and c_3 is a fixed positive constant. Combing the three terms, we get:

$$\mathbb{E}\left[\hat{\sigma}^{2}\right] \leq \sigma^{2}\left(c_{1}+c_{2}\log^{2}(n)\alpha+c_{3}\right)$$

 ${\sf Case \ 2:} \ \delta < 2/K \quad {\rm From \ Lemma \ 2.3, \ when \ } \delta < 2/K,$

$$q_i \le np_i + \exp\left(-\Omega\left(-\epsilon n \max_j p_j\right)\right) \le 2n \exp\left(-2^{2(i-l)-3}\right) + \exp\left(-\Omega\left(-\epsilon n\right)\right),$$

since $\max_j p_j = p_{(1)} > 1/100$ from the proof of Part (1), Claim 2. We get:

$$\begin{split} C &= \sum_{i>l+\log\log n} 2^{2i} q_i \\ &\leq \sum_{i>l+\log\log n} 2^{2i} \exp\left(-2^{2(i-l)-3}\right) + \sum_{l+\log\log n < i \le j_{\max}} 2^{2i} \exp\left(-\Omega\left(-\epsilon n\right)\right) \\ &\leq \sigma^2 + \sigma_{\max}^2 \exp\left(-\Omega\left(-\epsilon n\right)\right), \end{split}$$

for all sufficiently large n. Hence,

$$\mathbb{E}\left[\hat{\sigma}^{2}\right] \leq \sigma^{2} \left(c_{1} + c_{2}\log^{2}(n)\alpha + c_{3} + e^{-\Omega(\epsilon n)} \cdot \frac{\sigma_{\max}^{2}}{\sigma_{\min}^{2}}\right)$$
$$\leq \sigma^{2} \left(c_{1} + c_{2}\log^{2}(n) \cdot \alpha\right)$$

if

$$n \ge \frac{c}{\epsilon} \log\left(\frac{\sigma_{\max}}{\sigma_{\min}}\right)$$

Hence in both cases, we have:

$$\mathbb{E}[\hat{\sigma}^2] \le \sigma^2 \left(c_1 + c_2 \log^2(n) \cdot \alpha \right)$$

22

3.3 Estimation of range with unknown variance

In this section we will present an algorithm that estimates the range of data from a normal distribution when both the mean and the variance are unknown.

Theorem 3.3. For every $n \in \mathbb{N}$, $\sigma_{\max} > \sigma_{\min} \in [0, \infty], \epsilon, \delta > 0$, $\alpha \in (0, 1/2)$, $R \in [0, \infty]$, there is an (ϵ, δ) differentially private algorithm $M : \mathbb{R}^n \to (0, \infty) \times \mathbb{R} \times \mathbb{R}$ such that whenever

$$n \ge c \cdot \min\left\{ \max\left\{ \frac{1}{\epsilon} \log\left(\frac{R}{\sigma_{\min}\alpha}\right), \frac{1}{\epsilon} \log\left(\frac{\log(\sigma_{\max}/\sigma_{\min})}{\alpha}\right) \right\}, \frac{1}{\epsilon} \log\left(\frac{1}{\delta\alpha}\right) \right\}, \frac{1}{\epsilon} \log\left(\frac{1}{\delta\alpha}\right) \right\}, \frac{1}{\epsilon} \log\left(\frac{1}{\delta\alpha}\right) \right\}$$

(where c is a universal constant), if X_1, \ldots, X_n are iid Gaussian random variables with mean $\mu \in (-R, R)$ and with variance $\sigma^2 \in (\sigma_{\min}^2, \sigma_{\max}^2)$, and

$$(\hat{\sigma}, X_{\min}, X_{\max}) \leftarrow M(X_1, \dots, X_n)$$

we have:

1.
$$\mathbb{P}_{\substack{\underline{X} \sim N(\mu, \sigma^2) \\ M}} (\forall i \ X_{\min} \leq X_i \leq X_{\max}) \geq 1 - \alpha$$

2.
$$\mathbb{P}_{\substack{\underline{X} \sim N(\mu, \sigma^2) \\ M}} \left(|X_{\max} - X_{\min}| \leq \mathcal{O}\left(\sigma\sqrt{\log(n/\alpha)}\right) \right) \geq 1 - \alpha$$

3.
$$\mathbb{P}_{\substack{\underline{X} \sim N(\mu, \sigma^2) \\ M}} (\sigma \leq \hat{\sigma} \leq 8\sigma) \geq 1 - \alpha$$

Moreover, if

$$n \ge c \cdot \min\left\{\frac{1}{\epsilon} \log\left(\frac{\sigma_{\max}}{\sigma_{\min}}\right), \frac{1}{\epsilon} \log\left(\frac{1}{\delta\alpha}\right)\right\},$$
$$\mathbb{E}_{\underline{X} \sim N(\mu, \sigma^2)} \left[\hat{\sigma}^2\right] \le \sigma^2 \left(c_1 + c_2 \log^2(n) \cdot \alpha\right)$$

then

We briefly comment on Theorem 3.3. As before, when
$$\delta > 0$$
, μ and σ^2 can remain unbounded as there is
no dependence of n on R , σ_{\min} and σ_{\max} . On the other hand, when $\delta = 0$, μ and σ^2 must be bounded, and the
dependence of n is logarithmic on R/σ_{\min} and $\sigma_{\max}/\sigma_{\min}$. Hence even when $\delta = 0$, these parameters can be set
to a large value as the dependence is only logarithmic. Moreover, with high probability, the estimated range is
of the same order as the expected range for Gaussian data, which is $\Theta(\sigma\sqrt{\log n})$

Proof of Theorem 3.3. We first obtain an algorithm that estimates the range by combining Algorithms 2 and 1. The idea is to first find an good estimate $\hat{\sigma}$ of σ using Algorithm 2. By Theorem 3.2, $\hat{\sigma}$ is an upper bound on σ . We then run Algorithm 1 which finds the range of the data when an upper bound on the variance is known.

Algorithm 3 Differentially private estimate of range with unknown variance

Input: $x_1, \ldots, x_n, \epsilon, \alpha, R, \sigma_{\min}, \sigma_{\max}$.

Output: An (ϵ, δ) differentially private estimate of the range of the data x_1, \ldots, x_n . 1: If

$$n < c \cdot \min\left\{ \max\left(\frac{1}{\epsilon} \log\left(\frac{R}{\sigma_{\min}\alpha}\right), \frac{1}{\epsilon} \log\left(\frac{\log(\sigma_{\max}/\sigma_{\min})}{\alpha}\right)\right), \frac{1}{\epsilon} \log\left(\frac{1}{\delta\alpha}\right) \right\},\$$

Output \perp .

- 2: Run Algorithm 2 of Lemma 2.3 with $\epsilon_1 = \epsilon/2$, $\delta_1 = \delta/2$, $\alpha_1 = \alpha/2$, σ_{\min} and σ_{\max} on the data x_1, \ldots, x_n to obtain an estimate of scale $\hat{\sigma}$.
- 3: Run Algorithm 1 with $\epsilon_2 = \epsilon/2$, $\delta_2 = \delta/2$, $\alpha_2 = \alpha/2$, $\sigma = \hat{\sigma}$ and R to get $[X_{\min}, X_{\max}]$.
- 4: Output $\hat{\sigma}$ and $[X_{\min}, X_{\max}]$.

By the composition property of differential privacy given in Lemma 2.1, and the fact that Algorithms 1 and 2 are differentially private (from Theorems 3.1 and 3.2), it follows that Algorithm 3 is (ϵ, δ) differentially private.

Consider the event $E = \{\sigma < \hat{\sigma} < 8\sigma\}$. Under *E*, the results from Theorem 3.1 for the case of range finding when an upper bound on the variance is known can be used. Hence under the event *E*, we have if,

$$n \ge c \min\left\{\frac{1}{\epsilon} \log\left(\frac{R}{\sigma_{\min}\alpha}\right), \frac{1}{\epsilon} \log\left(\frac{1}{\delta\alpha}\right)\right\} \ge c \min\left\{\frac{1}{\epsilon} \log\left(\frac{R}{\hat{\sigma}\alpha}\right), \frac{1}{\epsilon} \log\left(\frac{1}{\delta\alpha}\right)\right\},$$

(1) With probability at least $1 - \alpha/2$,

 $\forall i, X_{\min} \leq X_i \leq X_{\max}, \text{ and }$

(2) With probability 1 we have $|X_{\max} - X_{\min}| = \mathcal{O}\left(\hat{\sigma}\sqrt{\log(n/\alpha)}\right) = \mathcal{O}\left(\sigma\sqrt{\log(n/\alpha)}\right)$ By the properties of the variance estimation algorithm given in Theorem 3.2, if

$$n \ge c \min\left\{\frac{1}{\epsilon} \log\left(\frac{\log_2\left(\frac{\sigma_{\max}}{\sigma_{\min}}\right)}{\alpha}\right), \frac{1}{\epsilon} \log\left(\frac{1}{\delta\alpha}\right)\right\},\$$

we have with probability at least $1 - \alpha/2$,

$$\sigma \leq \hat{\sigma} \leq 8\sigma.$$

The results (1), (2), and (3) follow from a union bound. Moreover, by Theorem 3.2, if

$$n \ge c \min\left\{\frac{1}{\epsilon}\log\left(\frac{\sigma_{\max}}{\sigma_{\min}}\right), \frac{1}{\epsilon}\log\left(\frac{1}{\delta\alpha}\right)\right\},\$$

then $\mathbb{E}\left[\hat{\sigma}^2\right] \leq \sigma^2 \cdot \left(c_1 + c_2 \log^2(n) \cdot \alpha\right)$ for some universal constants c_1 and c_2 .

4 Estimating a confidence interval of a mean with known variance

In this section, we present algorithms to estimate a differentially private confidence interval of a mean of Gaussian distribution when the variance is known. When the variance is fixed and known, the non-private interval is $\bar{X} \pm (\sigma/\sqrt{n})z_{1-\alpha/2}$ where \bar{X} is the sample mean, σ is the known standard deviation and $z_{1-\alpha/2}$ is the $1 - \alpha/2$ quantile of a standard normal distribution. (See section 7.1 for Definitions). Note that the interval is of fixed width namely $2\sigma/\sqrt{n}z_{1-\alpha/2} = \Theta\left((\sigma/\sqrt{n}) \cdot \sqrt{\log(n/\alpha)}\right)$. We will show that one can also obtain a fixed width (ϵ, δ) -DP confidence interval when the variance is known.

Theorem 4.1. Let \mathbb{I} be the set of all possible intervals in \mathbb{R} . For every $n \in \mathbb{N}$, $\sigma, \epsilon, \delta > 0$, $\alpha \in (0, 1/2)$, $R \in (0, \infty]$, there is $\beta > 0$ and an (ϵ, δ) -differentially private algorithm $M : \mathbb{R}^n \to \mathbb{I}$ such that if X_1, \ldots, X_n are iid random variables from $N(\mu, \sigma^2)$, with $\mu \in (-R, R)$ and $I \leftarrow M(X_1, \ldots, X_n)$, then

$$\mathbb{P}_{\substack{\underline{X} \sim N(\mu, \sigma^2) \\ M}} \left(I(X_1, \dots, X_n) \ni \mu \right) \ge 1 - \alpha.$$

Moreover, $|I| = \beta$ and if,

$$n \ge c \min\left\{\frac{1}{\epsilon} \log\left(\frac{R}{\sigma\alpha}\right), \frac{1}{\epsilon} \log\left(\frac{1}{\delta\alpha}\right)\right\},\$$

(where c is a universal constant), then,

$$\beta \leq \max\left\{\frac{\sigma}{\sqrt{n}}\mathcal{O}\left(\sqrt{\log\left(\frac{1}{\alpha}\right)}\right), \frac{\sigma}{\epsilon n} \text{polylog}\left(\frac{n}{\alpha}\right)\right\}.$$

Theorem 4.1 asserts that the (ϵ, δ) -differentially private algorithm produces a $(1 - \alpha)$ -fixed with confidence interval, no matter what the sample size is. Moreover, if n is large enough, the width of confidence interval produced is the maximum of two terms, one of the terms being the width obtained without privacy which is $\Theta\left((\sigma/\sqrt{n}) \cdot \sqrt{\log(1/\alpha)}\right)$, and the other term vanishing linearly in n. We will later show that a width of $\Omega\left(\sigma/(\epsilon n) \cdot \log(1/\alpha)\right)$ is necessary for privacy, (see Theorem 6.2 in Section 6.), so the second term cannot be significantly improved.

Proof of Theorem 4.1. We will show that the following algorithm produces the required confidence interval. First, we obtain a differentially private estimate of range of the data using Theorem 3.1. The data is truncated to lie within this range and the mean of the truncated data is released using the Laplace mechanism calibrated to the estimated range. Finally, the confidence interval is estimated using the noisy truncated mean by explicitly accounting for all sources of randomness. We present this algorithm followed by the proof of it's correctness.

Algorithm 4 Differentially private confidence interval with known variance.

Input: $X_1, \ldots, X_n, \alpha_0, \alpha_1, \alpha_2, \epsilon, \delta, \sigma, R$ **Output:** An $(\epsilon = \epsilon_1 + \epsilon_2, \delta)$ -differentially private $(\alpha = \alpha_0 + \alpha_1 + \alpha_2)$ -level confidence interval of μ . 1: If

$$n < c \min\left\{\frac{1}{\epsilon} \log\left(\frac{R}{\sigma \alpha_2}\right), \frac{1}{\epsilon} \log\left(\frac{1}{\delta \alpha_2}\right)\right\},\,$$

output (-R, R).

2: Run the (ε₂, δ) differentially private range estimation algorithm for known variance (Algorithm 1) on (x₁,...,x_n) to get a (1 - α₂) confident estimate [X_{min}, X_{max}] of the range. Let w₀ = X_{max} - X_{min}.
3: Let

$$Y_i = \begin{cases} X_i & \text{if } X_i \in [X_{\min}, X_{\max}] \\ X_{\max} & \text{if } X_i > X_{\max} \\ X_{\min} & \text{if } X_i < X_{\min} \end{cases}$$

4: Let

$$\tilde{\mu} = \frac{\sum_i Y_i}{n} + Z_1$$

where Z_1 is a Laplace random variable with mean 0 and scale parameter

$$b_1 = \frac{w_0}{\epsilon_1 n}.$$

5: Let

$$w = \frac{\sigma}{\sqrt{n}} z_{1-\alpha_0/2} + b_1 \log\left(\frac{1}{\alpha_1}\right)$$

where $\alpha_0 + \alpha_1 + \alpha_2 = \alpha$, and $z_{1-\alpha_0/2}$ is the $(1 - \alpha_0/2)$ -quantile of a standard normal distribution. 6: Output the interval:

$$\left(\tilde{\mu}-w,\tilde{\mu}+w\right).$$

In Step 1, if

$$n < c \min\left\{\frac{1}{\epsilon}\log\left(\frac{R}{\sigma\alpha_2}\right), \frac{1}{\epsilon}\log\left(\frac{1}{\delta\alpha_2}\right)\right\},\$$

Algorithm 4 outputs (-R, R). This step is trivially (ϵ, δ) -differentially private. Moreover, since it is given that $\mu \in (-R, R)$, the interval is trivially an $(1 - \alpha)$ -level confidence interval. When the algorithm runs beyond Step 1, by composition (Theorem 2.1) and the privacy property of Laplace mechanism (Theorem 2.2), it follows that the algorithm is (ϵ, δ) -differentially private. We now prove that when the algorithm runs beyond Step 1, the

interval produced in Step 6 is an $(1 - \alpha)$ -level confidence interval, of fixed width 2w i.e.,

$$\mathbb{P}_{\substack{\underline{X} \sim N(\mu, \sigma^2) \\ M}} (|\mu - \tilde{\mu}| < w) \ge 1 - \alpha,$$

where w is defined in Step 5 of the algorithm. We have,

$$|\tilde{\mu} - \mu| = \left| \hat{\mu} + Z_1 + \frac{1}{n} \sum_i (X_i - Y_i) - \mu \right|$$

$$\leq |\hat{\mu} - \mu| + \left| \frac{1}{n} \sum_i (X_i - Y_i) \right| + |Z_1|.$$

We will now analyze these three terms. Note that

$$\mathbb{P}\left(|\hat{\mu}-\mu| > \frac{\sigma}{\sqrt{n}} z_{1-\alpha_0/2}\right) \le \alpha_0,$$

since $\hat{\mu} - \mu$ has a normal distribution with mean 0 and variance σ^2/n and $z_{1-\alpha_0/2} = \phi^{-1}(1-\alpha_0/2)$. Next,

$$\mathbb{P}\left(\frac{1}{n}\left|\sum_{i}(X_{i}-Y_{i})\right|>0\right)\leq 1-\mathbb{P}\left(\forall iX_{\min}\leq X_{i}\leq X_{\max}\right)\leq\alpha_{2},$$

from Theorem 3.1, since $[X_{\min}, X_{\max}]$ is a $(1 - \alpha_2)$ confident range of X_1, \ldots, X_n . Finally,

$$\mathbb{P}\left(|Z_1| > b_1 \log\left(\frac{1}{\alpha_1}\right)\right) \le \alpha_1$$

due to the tails of a Laplace distribution, Proposition 7.2. Thus, with probability at least $1 - (\alpha_0 + \alpha_1 + \alpha_2) = 1 - \alpha$, we have

$$|\tilde{\mu} - \mu| \le \frac{\sigma}{\sqrt{n}} z_{1-\alpha_0/2} + 0 + b_1 \log\left(\frac{1}{\alpha_1}\right) = w$$

as desired. To finish the proof, we need to upper bound the width of the confidence interval. Let $\alpha_0 = \alpha_1 = \alpha_2 = \alpha/3$. From Theorem 3.1, we have

$$w_0 = c\sigma \sqrt{\log\left(\frac{n}{\alpha_1}\right)}$$

Hence, the width of the confidence interval is 2w where,

$$w = \frac{\sigma}{\sqrt{n}} z_{1-\alpha_0/2} + \frac{c\sigma}{\epsilon n} \sqrt{\log\left(\frac{n}{\alpha_1}\right)} \cdot \log\left(\frac{1}{\alpha_2}\right)$$
$$= \frac{\sigma}{\sqrt{n}} z_{1-\alpha/6} + \frac{c\sigma}{\epsilon n} \sqrt{\log\left(\frac{3n}{\alpha}\right)} \cdot \log\left(\frac{3}{\alpha}\right)$$

By using the tail bound on the normal distribution in Proposition 7.3 and the relation between the tail bound and a quantile in Proposition 7.6, we obtain the fact that $z_{1-\alpha} \leq \sqrt{2\log(1/\alpha)}$. Hence we get,

$$w \leq \frac{\sigma}{\sqrt{n}} \mathcal{O}\left(\sqrt{\log\left(\frac{1}{\alpha}\right)}\right) + \frac{c_2 \sigma}{\epsilon n} \sqrt{\log\left(\frac{n}{\alpha}\right)} \cdot \log\left(\frac{3}{\alpha}\right)$$
$$= \frac{\sigma}{\sqrt{n}} \mathcal{O}\left(\sqrt{\log\left(\frac{1}{\alpha}\right)}\right) + \frac{\sigma}{\epsilon n} \cdot \operatorname{polylog}\left(\frac{n}{\alpha}\right)$$

Nearly optimal width for finite sample sizes The differentially private confidence intervals obtained in Theorem 4.1 have a multiplicative increase in their length when compared to a classical non-private confidence interval due to the hidden constant in the term $\sigma/\sqrt{n} \cdot \left(\sqrt{\log(1/\alpha)}\right)$. We show that one can avoid this behavior and obtain a differentially private confidence interval with an additive increase, i.e. the differentially private length is the sum of two terms: the first term is the non-private length and the second term that vanishes faster than the non-private length, with a mildly worse dependence on other parameters.

Theorem 4.2. Let $\alpha_1 = \alpha_2 = \alpha/(2\sqrt{n})$. If,

$$n \ge c \min\left\{\frac{1}{\epsilon}\log\left(\frac{R}{\epsilon\sigma\alpha}\right), \frac{1}{\epsilon}\log\left(\frac{1}{\delta\alpha\epsilon}\right)\right\},\$$

then

$$w = \frac{\sigma}{\sqrt{n}} z_{1-\alpha/2} + \frac{\sigma}{\epsilon n} \text{polylog}\left(\frac{n}{\alpha}\right)$$

The key point is that there is no hidden constant in the first term, which matches the length of the non-private confidence interval.

We will first start with an asymptotic expansion of the quantile function of a standard normal variable, which is given in Proposition 4.1 below.

Proposition 4.1. Let $z_{1-p} = \phi^{-1}(1-p)$ denote the quantile function of the standard normal distribution. For any $\alpha \in (0,1)$ and $0 < \Delta < \alpha$, we have

$$z_{1-(\alpha-\Delta)/2} \le z_{1-\alpha/2} + \frac{c_1\Delta}{\alpha-\Delta}$$

where c_1 is a fixed constant.

Proof. Let $q(p) = z_{1-p/2} = \phi^{-1}(1-p/2)$. By the mean value theorem, we have

$$q\left(\alpha - \Delta\right) = q\left(\alpha\right) - \Delta q'\left(c\right)$$

for some $c \in [\alpha - \Delta, \alpha]$. One can show that

$$q'(p) = -\sqrt{\frac{\pi}{2}} \cdot e^{q(p)^2/2}$$

and Also, $q(p) = \phi^{-1}(1-p/2) \le \sqrt{2\log(2/p)}$, which can be obtained by applying Proposition 7.6 to a Gaussian tail bound. Hence we have,

$$\begin{aligned} q(\alpha - \Delta) &= q(\alpha) + \sqrt{\frac{\pi}{2}} \cdot e^{q(c)^2/2} \cdot \Delta \\ &\leq q(\alpha) + \sqrt{\frac{\pi}{2}} \cdot e^{q(\alpha - \Delta)^2/2} \cdot \Delta, \\ &(\text{Since } q(p) \text{ is decreasing in } [\alpha - \Delta, \alpha]) \\ &\leq q(\alpha) + \sqrt{\frac{\pi}{2}} \cdot \frac{2}{\alpha - \Delta} \cdot \Delta \text{ (Since } q^2(p) \leq 2\log(2/p) \text{)} \end{aligned}$$

Proof of Theorem 4.2. Let $\alpha_1 = \alpha_2 = \alpha/(2\sqrt{n})$. Hence $\alpha_0 = \alpha - \alpha/\sqrt{n}$. Applying Proposition 4.1 with $\Delta = \alpha/\sqrt{n}$, we get,

$$z_{1-\alpha_0/2} \le z_{1-\alpha/2} + \frac{c_1}{\sqrt{n}}$$

The width of the confidence interval is

$$w = \frac{\sigma}{\sqrt{n}} z_{1-\alpha_0/2} + \frac{c\sigma}{\epsilon n} \sqrt{\log\left(\frac{n}{\alpha_1}\right)} \cdot \log\left(\frac{1}{\alpha_2}\right)$$
$$\leq \frac{\sigma}{\sqrt{n}} \cdot \left(z_{1-\alpha/2} + \frac{c_1}{\sqrt{n}}\right) + \frac{c\sigma}{\epsilon n} \sqrt{\log\left(\frac{n\sqrt{n}}{\alpha}\right)} \cdot \log\left(\frac{\sqrt{n}}{\alpha}\right)$$
$$\leq \frac{\sigma}{\sqrt{n}} z_{1-\alpha/2} + \frac{\sigma}{\epsilon n} \text{polylog}\left(\frac{n}{\alpha}\right)$$

Finally, we need

$$n \ge c_1 \min\left\{\frac{1}{\epsilon}\log\left(\frac{R\sqrt{n}}{\sigma\alpha}\right), \frac{1}{\epsilon}\log\left(\frac{\sqrt{n}}{\delta\alpha}\right)\right\}.$$

For this, it suffices to have

$$n \ge c_2 \min\left\{\frac{1}{\epsilon}\log\left(\frac{R}{\epsilon\sigma\alpha}\right), \frac{1}{\epsilon}\log\left(\frac{1}{\delta\alpha\epsilon}\right)\right\}.$$

Г		

5 Estimating a confidence interval of a mean with unknown variance

In this section, we present a differentially private algorithm for estimating the confidence interval of the mean of a normal population when the variance is unknown. Let us first recall the interval in the non-private case. The standard confidence interval for μ in the unknown variance case is given by $\bar{X} \pm s/\sqrt{n} \cdot t_{n-1,\alpha/2}$, where

$$s^{2} = \frac{1}{n-1} \sum_{i=1}^{n} (X_{i} - \bar{X})^{2}$$

is the sample variance, $t_{n-1,\alpha/2}$ is the $\alpha/2^{th}$ quantile of a t-distribution with n-1 degrees of freedom (see Section 7.1 for definitions.)

Note that unlike the known variance case, we need to use the sample variance s^2 as an estimate of σ^2 , and to account for this, we need to replace the z-quantile with a t-quantile. Moreover, when the variance is unknown, the optimal confidence interval is no longer of fixed length, in contrast with the known variance case. The expected length of the interval is given by

$$\frac{2\sigma}{\sqrt{n}} \cdot k_n \cdot t_{n-1,1-\alpha/2} = \Theta\left(\sigma\sqrt{\log(1/\alpha)/n}\right)$$

where

$$k_n = \sqrt{\frac{2}{n-1}} \cdot \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)} = \mathcal{O}\left(1 - \frac{1}{n}\right).$$

(see for example Lehmann and Romano (2006) and Bolch (1968).)

Let us now consider a differentially private algorithm for estimating the confidence interval when the variance is unknown. Recall that in the known variance case given in Algorithm 4, we first estimate the range of the data using σ , and then estimate the mean of the data truncated to the estimated range, using a Laplace mechanism. A conservative confidence interval is estimated by considering all sources of randomness, including the noise due to privacy, and the range estimation and truncation step. When σ is unknown, we proceed by using Algorithm 2 to get a crude estimate of σ . Once we have an estimate of σ , we proceed as in the known variance case to estimate the range, followed by using a Laplace mechanism to compute the mean of the data truncated to the estimated range. For construction of a differentially private confidence interval in the unknown variance case, we also need a conservative estimate of the sample variance of the truncated data using Laplace Mechanism. Finally, the analysis of the algorithm takes into account all sources of randomness to ensure the required coverage is obtained. Algorithm 5 describes the estimator of confidence interval with unknown variance.

Theorem 5.1. Let \mathbb{I} be the set of all possible intervals in \mathbb{R} . For every $n \in \mathbb{N}$, $\sigma_{\min} < \sigma_{\max} \in [0, \infty], \epsilon, \delta > 0$, $\alpha \in (0, 1/2), R \in [0, \infty)$, there is an (ϵ, δ) -differentially private algorithm $M : \mathbb{R}^n \to \mathbb{I}$ such that if X_1, \ldots, X_n are iid random variables from $N(\mu, \sigma^2)$, where $\mu \in (-R, R)$ and $\sigma \in (\sigma_{\min}, \sigma_{\max})$, and $I \leftarrow M(X_1, \ldots, X_n)$, then

$$\mathbb{P}_{\substack{\underline{X} \sim N(\mu, \sigma^2) \\ M}} \left(I(X_1, \dots, X_n) \ni \mu \right) \ge 1 - \alpha.$$

Moreover, if

$$n \ge \frac{c_1}{\epsilon} \min\left\{ \max\left\{ \log\left(\frac{R}{\sigma_{\min}}\right), \log\left(\frac{\sigma_{\max}}{\sigma_{\min}}\right) \right\}, \log\left(\frac{1}{\delta}\right) \right\} + \frac{c_2}{\epsilon} \log\left(\frac{\log\left(\frac{1}{\epsilon}\right)}{\alpha}\right) \right\}$$

(where c_1 and c_2 are universal constants), then,

$$\beta := \mathop{\mathbb{E}}_{\substack{\underline{X} \sim N(\mu, \sigma^2) \\ M}} \left[|I(X_1, \dots, X_n)| \right] \le \max\left\{ \frac{\sigma}{\sqrt{n}} \mathcal{O}\left(\sqrt{\log\left(\frac{1}{\alpha}\right)}\right), \frac{\sigma}{\epsilon n} \text{polylog}\left(\frac{1}{\alpha}\right) \right\}$$

As in the case of known variance, Theorem 5.1 asserts that there exists an (ϵ, δ) -differentially private algorithm that produces a $(1 - \alpha)$ -confidence interval, no matter what the sample size is. Moreover, if n is large enough, the width of the confidence interval is non-trivial. It is the maximum of two terms. The first term being the width of interval obtained without privacy which is $\Theta\left((\sigma/\sqrt{n}) \cdot \sqrt{\log(1/\alpha)}\right)$, and hence is necessary. Using the fact that a lower bound of the known variance applies to the unknown variance, it will be shown that the second term of $\Omega\left(\sigma/(\epsilon n) \cdot \log(1/\alpha)\right)$ is also necessary for privacy. (See Theorem 6.2 in Section 6.) Thus we can match the lower bound upto polylog factors.

Proof. We first start with the algorithm:

Algorithm 5 Differentially private confidence interval with unknown variance.

Input: $X_1, \ldots, X_n, \alpha_0, \alpha_1, \alpha_2, \alpha_3, \epsilon_1, \epsilon_2, \epsilon_3, \delta, \sigma_{\min}, \sigma_{\max}, R$

Output: An $(\epsilon = \epsilon_1 + \epsilon_2 + \epsilon_3, \delta)$ -differentially private $(1 - \alpha)$ -level confidence interval of μ , where $\alpha = \alpha_0 + \alpha_1 + \alpha_2 + \alpha_3$.

1: If

$$n < c \min\left\{ \max\left(\frac{1}{\epsilon_3} \log\left(\frac{R}{\sigma_{\min}\alpha_3}\right), \frac{1}{\epsilon_3} \log\left(\frac{\log_2 \frac{\sigma_{\max}}{\sigma_{\min}}}{\alpha_3}\right)\right), \frac{1}{\epsilon_3} \log\left(\frac{1}{\delta\alpha_3}\right) \right\},$$

output (-R, R).

Run an (ε₃, δ) differentially private range estimation algorithm for unknown variance from Theorem 3.3 on (X₁,..., X_n) to get a (1 - α₃) confident estimate of the range [X_{min}, X_{max}]. Let w₀ = X_{max} - X_{min}.
 Let

$$Y_i = \begin{cases} X_i & \text{if } X_i \in [X_{\min}, X_{\max}] \\ X_{\max} & \text{if } X_i > X_{\max} \\ X_{\min} & \text{if } X_i < X_{\min} \end{cases}$$

4: Let

$$\tilde{\mu} = \frac{\sum_i Y_i}{n} + Z_1$$

where Z_1 is a Laplace random variable with mean 0 and scale parameter

$$b_1 = \frac{w_0}{\epsilon_1 n}.$$

5: Truncate $\tilde{\mu}$ to lie in the interval $[X_{\min}, X_{\max}]$.

6: Let

$$s_1^2 = \frac{\sum_i (Y_i - \tilde{\mu})^2}{n - 1}$$

7: Let $\tilde{s}^2 = s_1^2 + Z_2 + b_2 \log\left(\frac{1}{\alpha_2}\right)$ where $Z_2 \sim Lap(0, b_2)$ and

$$b_2 = \frac{w_0^2}{\epsilon_2 \cdot (n-1)}$$

8: If $\tilde{s}^2 < 0$ or $\tilde{s}^2 > \sigma_{\max}^2$, set $\tilde{s}^2 = \sigma_{\max}^2$. 9:

$$w = \frac{\tilde{s}}{\sqrt{n}} t_{n-1,\alpha_0/2} + b_1 \log\left(\frac{1}{\alpha_1}\right)$$

where $\alpha_0 + \alpha_1 + \alpha_2 + \alpha_3 = \alpha$, and $t_{n-1,\alpha_0/2}$ is the $(1 - \alpha_0/2)$ -quantile of a t-distribution with n-1 degrees of freedom. (See Section 7.1 for Definitions.)

10: Output the interval:

$$\left[\tilde{\mu}-w,\tilde{\mu}+w\right].$$

In the algorithm, we need to partition $\epsilon = \epsilon_1 + \epsilon_2 + \epsilon_3$ and $\alpha = \alpha_0 + \alpha_1 + \alpha_2 + \alpha_3$. We will defer this choice to the step when we bound the expected length of the confidence interval.

If the condition in Step 1 is satisfied, the output is trivially an (ϵ, δ) -differentially private $(1 - \alpha)$ -level confidence interval. So we focus on the case when the algorithm runs beyond Step 1. The proof has three parts, proof of privacy, coverage guarantee, and the expected length. We begin with the proof of privacy.

Privacy: The proof of privacy is the same as in Theorem 4.1, except we also need to account for the computation of the estimate \tilde{s} of the standard deviation. Note that the global sensitivity of s_1^2 in Step 6 is $w_0^2/(n-1)$, so

 \tilde{s}^2 is $(\epsilon_2, 0)$ -differentially private. By composition of differential privacy (Theorem 2.1) and the properties of Laplace mechanism (Theorem 2.2), it follows that the algorithm is (ϵ, δ) -DP.

Coverage: Now we will show that the algorithm outputs an interval with the desired coverage. Let E be the following event:

$$\left\{ \left| \tilde{\mu} - \mu \right| \le \frac{\tilde{s}}{\sqrt{n}} t_{n-1,\alpha_0/2} + b_1 \log\left(\frac{1}{\alpha_1}\right) \right\}$$

We need to show $\mathbb{P}(E) \ge 1 - \alpha$. Let $\hat{\mu} = (\sum_i X_i)/n$, and

$$s^2 = \frac{1}{(n-1)} \sum_{i} (X_i - \hat{\mu})^2$$

Consider the following events:

$$E_{0} = \left\{ |\hat{\mu} - \mu| \leq \frac{s}{\sqrt{n}} t_{n-1,\alpha_{0}/2} \right\}$$

$$E_{1} = \left\{ |Z_{1}| \leq b_{1} \log \left(\frac{1}{\alpha_{1}}\right) \right\}$$

$$E_{2} = \left\{ |Z_{2}| \leq b_{2} \log \left(\frac{1}{\alpha_{2}}\right) \right\}$$

$$E_{3} = \left\{ \forall i, X_{i} = Y_{i} \text{ and } 1 \leq \frac{\hat{\sigma}}{\sigma} \leq 8 \right\}$$

Claim 1 $E_0 \cap E_1 \cap E_2 \cap E_3 \implies E$

Proof of Claim 1. Assuming that E_0, E_1, E_2, E_3 all hold. We will first show that \tilde{s}^2 is a conservative estimate of s^2 . From E_3 , we have $X_i = Y_i$. Hence $\tilde{\mu} = \hat{\mu} + Z_1$ and

$$s_1^2 = \frac{\sum_i (Y_i - \hat{\mu} + Z_1)^2}{n - 1}$$

= $\frac{\sum_i (X_i - \hat{\mu})^2}{n - 1} + \frac{n}{n - 1} \cdot Z_1^2 + 2 \frac{\sum_i (X_i - \hat{\mu})}{n - 1}$
= $\frac{\sum_i (X_i - \hat{\mu})^2}{n - 1} + \frac{n}{n - 1} \cdot Z_1^2 + 2 \frac{\sum_i X_i - n\hat{\mu}}{n - 1}$
= $s^2 + \frac{n}{n - 1} \cdot Z_1^2$.

We have

$$\tilde{s}^{2} = s_{1}^{2} + Z_{2} + b_{2} \log\left(\frac{1}{\alpha_{2}}\right)$$

$$= s^{2} + \frac{n}{n-1}Z_{1}^{2} + Z_{2} + b_{2} \log\left(\frac{1}{\alpha_{2}}\right)$$

$$> s^{2}, \text{ since by } E_{2}, |Z_{2}| \le b_{2} \log\left(\frac{1}{\alpha_{2}}\right)$$
(7)

Hence, since $s^2 \leq \tilde{s}^2$, from E_0 we have,

$$|\hat{\mu} - \mu| \le \frac{s}{\sqrt{n}} t_{n-1,\frac{\alpha_0}{2}} \le \frac{\tilde{s}}{\sqrt{n}} t_{n-1,\frac{\alpha_0}{2}}$$

Finally, using E_1, E_3 and the fact that

$$|\tilde{\mu} - \mu| \le |\hat{\mu} - \mu| + \left|\frac{\sum_{i}(X_i - Y_i)}{n}\right| + |Z_1|,$$

we get,

$$|\tilde{\mu} - \mu| \le \frac{\tilde{s}}{\sqrt{n}} t_{n-1,\frac{\alpha_0}{2}} + b_1 \log\left(\frac{1}{\alpha_1}\right)$$

Thus, by union bound, to prove that $\mathbb{P}(E) > 1 - \alpha$, it is enough to show the following claim:

Claim 2 $\mathbb{P}(E_0) \ge 1 - \alpha_0$, $\mathbb{P}(E_1) \ge 1 - \alpha_1$, $\mathbb{P}(E_2) \ge 1 - \alpha_2$ and $\mathbb{P}(E_3) \ge 1 - \alpha_3$.

Proof of Claim 2. By the properties of a non-private confidence interval, (see for e.g. (Lehmann and Romano, 2006)), we have, $\mathbb{P}(E_0) \ge 1 - \alpha_0$. By tail properties of a Laplace distribution given in Proposition 7.2, we have $\mathbb{P}(E_1) \ge 1 - \alpha_1$, $\mathbb{P}(E_2) \ge 1 - \alpha_2$. Finally, by Theorem 3.3, if

$$n \ge c \min\left\{ \max\left(\frac{1}{\epsilon_3} \log\left(\frac{R}{\sigma_{\min}\alpha_3}\right), \frac{1}{\epsilon_3} \log\left(\frac{\log(\sigma_{\max}/\sigma_{\min})}{\alpha_3}\right)\right), \frac{1}{\epsilon_3} \log\left(\frac{1}{\delta\alpha_3}\right) \right\},\$$

then,

$$\mathbb{P}(\forall i, X_i = Y_i) = \mathbb{P}(\forall i, X_{\min} \le X_i \le X_{\max}) \ge 1 - \alpha_3$$

Length of the Interval: We now need to upper bound the expected length of the interval. Recall that the length of the interval is 2w, where

$$w = \frac{\tilde{s}}{\sqrt{n}} t_{n-1,\alpha_0/2} + b_1 \log\left(\frac{1}{\alpha_1}\right).$$

Here, \tilde{s}^2 and b_1 are the only random variables; The following proposition upper bounds the expectation of both these terms for a specific choice of $\alpha_0, \alpha_1, \alpha_2$, and α_3 .

Proposition 5.1. Let

$$\alpha_3 = \min\left\{\frac{\alpha}{4}, \frac{1}{\log^2 n}\right\},\,$$

 $\alpha_0 = \alpha_1 = \alpha_2 = (\alpha - \alpha_3)/3$, and $\epsilon_1 = \epsilon_2 = \epsilon_3 = \epsilon/3$, then we have

$$\mathbb{E}[\tilde{s}] \le k_n \cdot \sigma + \sigma \cdot \mathcal{O}\left(\sqrt{\frac{\log\left(\frac{n}{\alpha}\right)}{\epsilon n}} \cdot \sqrt{\log\left(\frac{1}{\alpha}\right)}\right)$$
(8)

where

$$k_n = \sqrt{\frac{2}{n-1}} \cdot \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)}$$

and

$$\mathbb{E}[b_1] \le \frac{\sigma}{\epsilon n} \cdot \mathcal{O}\left(\sqrt{\log(n/\alpha)}\right).$$
(9)

Deferring the proof of Proposition 5.1 to the end, let us upper bound w. Using the settings of α_i 's specified by Proposition 5.1, the expected value of w is:

$$\mathbb{E}[w] = \frac{\mathbb{E}[\tilde{s}]}{\sqrt{n}} \cdot t_{n-1,\alpha_0/2} + \mathbb{E}[b_1] \cdot \log\left(\frac{1}{\alpha_1}\right)$$

$$\leq \frac{\sigma}{\sqrt{n}} \cdot t_{n-1,\alpha_0/2} \left[k_n + \mathcal{O}\left(\sqrt{\frac{\log(n/\alpha)}{\epsilon n}} \cdot \sqrt{\log\left(\frac{1}{\alpha}\right)}\right) \right]$$

$$+ \mathcal{O}\left(\frac{\sigma}{\epsilon n}\sqrt{\log\left(\frac{n}{\alpha}\right)} \cdot \log\left(\frac{1}{\alpha}\right)\right)$$

$$\leq k_n \cdot \frac{\sigma}{\sqrt{n}} t_{n-1,\alpha_0/2} + \frac{\sigma}{\epsilon n} \cdot t_{n-1,\alpha_0/2} \cdot \operatorname{polylog}\left(\frac{n}{\alpha}\right)$$
(10)

From Proposition 7.7 Part (2), if $n \geq \mathcal{O}(\log(1/\alpha_0)) = \mathcal{O}(\log(1/\alpha))$, we have,

$$t_{n-1,\alpha_0/2} \leq \mathcal{O}\left(\sqrt{\log(1/\alpha_0)}\right) = \mathcal{O}\left(\sqrt{\log(1/\alpha)}\right),$$

since $\alpha_0 \in [\alpha/4, \alpha/3]$. Substituting this in equation 10, we get

$$\mathbb{E}\left[w\right] \leq \frac{k_n \sigma}{\sqrt{n}} \mathcal{O}\left(\sqrt{\log(1/\alpha)}\right) + \frac{\sigma}{\epsilon n} \mathcal{O}\left(\sqrt{\log(1/\alpha)}\right) \operatorname{polylog}\left(\frac{n}{\alpha}\right)$$
$$= \frac{k_n \sigma}{\sqrt{n}} \mathcal{O}\left(\sqrt{\log(1/\alpha)}\right) + \frac{\sigma}{\epsilon n} \operatorname{polylog}\left(\frac{n}{\alpha}\right)$$

Finally, let us verify the conditions on the sample size n. We have used the following two conditions on n: We need

$$n \ge \frac{c}{\epsilon_3} \min\left\{ \max\left(\log\left(\frac{R}{\sigma_{\min}\alpha_3}\right), \log\left(\frac{\log(\sigma_{\max}/\sigma_{\min})}{\alpha_3}\right) \right), \log\left(\frac{1}{\delta\alpha_3}\right) \right\},\$$

and from, Theorem 3.3 to obtain the upper bound on the expectation of the variance, we need

$$n \ge \frac{c}{\epsilon_3} \min\left\{\log\left(\frac{\sigma_{\max}}{\sigma_{\min}}\right), \log\left(\frac{1}{\alpha_3\delta}\right)\right\}.$$

Since $\alpha_3 = \min \{ \alpha/4, 1/\log^2 n \}$ and $\epsilon_3 = \epsilon/3$, it suffices to require

$$n \ge c \min\left\{ \max\left(\frac{1}{\epsilon} \log\left(\frac{R \log(1/\epsilon)}{\sigma_{\min}\alpha}\right), \frac{1}{\epsilon} \log\left(\frac{\sigma_{\max} \log(1/\epsilon)}{\sigma_{\min}\alpha}\right)\right), \frac{1}{\epsilon} \log\left(\frac{\log(1/\epsilon)}{\delta\alpha}\right) \right\}.$$

Taking the common terms out, we get,

$$n \ge c \min\left\{ \max\left\{ \log\left(\frac{R}{\sigma_{\min}}\right), \log\left(\frac{\sigma_{\max}}{\sigma_{\min}}\right) \right\}, \log\left(\frac{1}{\delta}\right) \right\} + \frac{c}{\epsilon} \log\left(\frac{\log(1/\epsilon)}{\alpha}\right).$$

All that remains is a proof of Proposition 5.1, which is given below.

Proof. Let $A = \{ \forall i, X_i = Y_i \}$. Under the event A, from equation 7, we have,

$$\tilde{s}^{2} = s^{2} + \frac{n}{n-1}Z_{1}^{2} + Z_{2} + b_{2}\log\left(\frac{1}{\alpha_{2}}\right)$$
$$:= s^{2} + Y_{1}.$$

Under the event A^c , we can use the following upper bound of \tilde{s}^2 :

$$\tilde{s}^2 \le \frac{w_0^2}{\epsilon_2 \cdot (n-1)} := Y_2.$$

Let \mathbb{I}_A be the indicator function of event A. We have,

$$\begin{split} \tilde{s}^2 &= \tilde{s}^2 \cdot \mathbb{I}_A + \tilde{s}^2 \cdot \mathbb{I}_{A^c} \\ &\leq (s^2 + Y_1) \cdot \mathbb{I}_A + Y_2 \cdot \mathbb{I}_{A^c} \\ &\leq s^2 + Y_1 + Y_2. \end{split}$$

Using the inequality $\sqrt{a+b+c} \leq \sqrt{a} + \sqrt{b} + \sqrt{c}$ for non-negative reals and Jensen's inequality for $f(x) = \sqrt{x}$, we have,

$$\mathbb{E}[\tilde{s}] \le \mathbb{E}[s] + \sqrt{\mathbb{E}[Y_1]} + \sqrt{\mathbb{E}[Y_2]}.$$
(11)

Let us compute the expectation of the three terms in equation 11.

(1) First note that $\mathbb{E}[s] = k_n \sigma$, where

$$k_n = \sqrt{\frac{2}{n-1}} \cdot \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)}.$$

(See, for example, Bolch (1968).)

(2) Next, we have

$$\mathbb{E}[Y_1] = \mathbb{E}\left[\frac{n}{n-1} \cdot Z_1^2 + Z_2 + b_2 \log\left(\frac{1}{\alpha_2}\right)\right]$$
$$= \frac{2n}{n-1} \cdot \mathbb{E}[b_1^2] + 0 + \mathbb{E}[b_2] \cdot \log\left(\frac{1}{\alpha_2}\right)$$
$$= \frac{2n}{n-1} \cdot \mathbb{E}\left[\frac{w_0^2}{\epsilon_1^2 n^2}\right] + \mathbb{E}\left[\frac{w_0^2}{\epsilon_2 \cdot (n-1)}\right] \cdot \log\left(\frac{1}{\alpha_2}\right)$$

Note that b_1 and b_2 , the scales of Z_1 and Z_2 respectively, are also random. We have used the fact that if Z is a Laplace random variable with mean 0 and scale b, then $\mathbb{E}[Z] = 0$ and $\mathbb{E}[Z^2] = 2b^2$, see Proposition 7.2. Now let us upper bound the expectation of w_0^2 . Recall that

$$w_0 = X_{\max} - X_{\min} = \hat{\sigma} \sqrt{\log(n/\alpha_3)}.$$

Given that

$$\alpha_3 = \min\left\{\frac{\alpha}{4}, \frac{1}{\log^2 n}\right\},\,$$

and $\alpha_0 = \alpha_1 = \alpha_2 = (\alpha - \alpha_3)/3 \in [\alpha/4, \alpha/3]$. From Theorem 3.3, if $n \ge (c/\epsilon) \log(\sigma_{\max}/\sigma_{\min})$, we have,

$$\mathbb{E}[\hat{\sigma}^2] \le \sigma^2 \left(c_1 + c_2 \alpha_3 \log^2(n) \right) \le \sigma^2 \left(c_1 + c_2 \right) \le K \sigma^2$$

Hence,

$$\mathbb{E}[w_0^2] \le K\sigma^2 \cdot \log\left(\frac{n}{\min\left\{\alpha/4, 1/(\log^2 n)\right\}}\right) = \sigma^2 \cdot \mathcal{O}\left(\log(n/\alpha)\right).$$

Hence we have,

$$\mathbb{E}[Y_1] \le \frac{2n}{n-1} \cdot \mathbb{E}\left[\frac{w_0^2}{\epsilon_1^2 n^2}\right] + \mathbb{E}\left[\frac{w_0^2}{\epsilon_2 \cdot (n-1)}\right] \log\left(\frac{1}{\alpha_2}\right)$$
$$\le \sigma^2 \left(\mathcal{O}\left(\frac{\log(n/\alpha)}{\epsilon^2 n^2}\right) + \mathcal{O}\left(\frac{\log(n/\alpha)}{\epsilon n} \cdot \log\left(\frac{1}{\alpha}\right)\right)\right)$$
(12)

since $\epsilon_1 = \epsilon_2 = \epsilon_3 = \epsilon/3$ and $\alpha_2 \in [\alpha/4, \alpha/3]$.

(3) Finally, we have

$$\mathbb{E}\left[Y_2\right] \le \frac{\mathbb{E}[w_0^2]}{\epsilon_2 \cdot (n-1)} \le \mathcal{O}\left(\frac{\sigma^2}{\epsilon n} \cdot \log\left(\frac{n}{\alpha}\right)\right)$$
(13)

Combining inequalities 12 and 13, and plugging in equation 11, we have

$$\mathbb{E}[\tilde{s}] \le k_n \sigma + \sigma \cdot \left[\mathcal{O}\left(\sqrt{\frac{\log(n/\alpha)}{\epsilon n}} \cdot \sqrt{\log(1/\alpha)}\right) + \mathcal{O}\left(\frac{\sqrt{\log(n/\alpha)}}{\epsilon n}\right) \right]$$

Similarly, we have,

$$\mathbb{E}[b_1] = \mathbb{E}\left[\frac{w_0}{\epsilon_1 n}\right] \le \frac{\sqrt{\mathbb{E}\left[w_0^2\right]}}{\epsilon_1 n} \le \frac{\sigma}{\epsilon n} \cdot \mathcal{O}\left(\sqrt{\log(n/\alpha)}\right),$$

since $\epsilon_1 = \epsilon/3$.

Nearly optimal width for finite sample sizes. As in the case of known variance, we can obtain a differentially private confidence interval whose increase in length is additive as opposed to multiplicative with a minor increase in the sample complexity.

Theorem 5.2. Let $\alpha_1 = \alpha_2 = \alpha_3 = \alpha/(3\sqrt{n})$. If,

$$n \ge c \min\left\{ \max\left(\frac{1}{\epsilon} \log\left(\frac{R}{\epsilon \sigma_{\min} \alpha}\right), \frac{1}{\epsilon} \log\left(\frac{\sigma_{\max}}{\epsilon \sigma_{\min} \alpha}\right)\right), \frac{1}{\epsilon} \log\left(\frac{1}{\epsilon \delta \alpha}\right) \right\},$$

then

$$w = k_n \cdot \frac{\sigma}{\sqrt{n}} t_{n-1,1-\alpha/2} + \frac{\sigma}{\epsilon n} \cdot \operatorname{polylog}\left(\frac{n}{\alpha}\right) + \mathcal{O}\left(\frac{\sigma}{n}\right) t_{n-1,1-\alpha/2}$$

where

$$k_n = \sqrt{\frac{2}{n-1}} \cdot \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)}$$

We will first start with an asymptotic expansion of the quantile function of a t-distribution, which is given in Proposition 5.2 below.

Proposition 5.2. Let $t_{n,1-p} = \phi^{-1}(1-p)$ denote the $1-p^{th}$ quantile of a t-distribution with n degrees of freedom. For any $\alpha \in (0,1/2)$ and $0 < \Delta < \alpha$, we have

$$t_{n,1-(\alpha-\Delta)/2} \le t_{n,1-\alpha/2} + \frac{c_1\Delta}{(\alpha-\Delta)}$$

where c_1 is a fixed constant.

Proof. Let $q(p) = t_{n,1-p/2} = G_n^{-1}(1-p/2)$ where $G_n(.)$ is the cdf of a t-distribution with n degrees of freedom. By the mean value theorem, we have

$$q\left(\alpha - \Delta\right) = q\left(\alpha\right) - \Delta q'\left(c\right)$$

for some $c \in [\alpha - \Delta, \alpha]$. By taking the derivative, using Proposition 7.7 for a bound on the quantile of a *t*-distribution, one can show that

$$-q'(p) = \frac{\sqrt{n\pi}}{2} \frac{\Gamma(n/2)}{\Gamma((n+1)/2)} \left(1 + \frac{q^2(p)}{n}\right)^{(n+1)/2}$$
$$\leq \sqrt{\frac{\pi}{2}} \cdot \sqrt{\frac{n}{n-1}} \left(1 + \frac{q^2(p)}{n}\right)^{(n+1)/2}$$
$$\leq c \cdot \exp\left(q^2(p) \cdot \frac{n+1}{2n}\right)$$
$$\leq \frac{c}{p} \text{ (Proposition 7.5 Part(2))}$$

Hence we have,

$$q(\alpha - \Delta) \le q(\alpha) + \Delta \frac{c_1}{(\alpha - \Delta)}$$

Proof of Theorem 5.2. Let $\alpha_1 = \alpha_2 = \alpha_3 = \alpha/(3\sqrt{n})$. Hence $\alpha_0 = \alpha - \alpha/\sqrt{n}$. Applying Proposition 5.2 with $\Delta = \alpha/\sqrt{n}$, we get,

$$t_{n-1,1-\alpha_0/2} \le t_{n-1,1-\alpha/2} + \frac{c_1}{\sqrt{n}}$$

From equation 10 in the proof of Theorem 5.1, we have,

$$\mathbb{E}[w] \le k_n \cdot \frac{\sigma}{\sqrt{n}} \cdot t_{n-1,1-\alpha_0/2} + \frac{\sigma}{\epsilon n} \cdot \operatorname{polylog}\left(\frac{n}{\alpha}\right)$$
$$\le k_n \cdot \frac{\sigma}{\sqrt{n}} \left(t_{n-1,1-\alpha/2} + \frac{c_1}{\sqrt{n}}\right) + \frac{\sigma}{\epsilon n} \operatorname{polylog}\left(\frac{n}{\alpha}\right)$$
$$\le k_n \cdot \frac{\sigma}{\sqrt{n}} t_{n-1,1-\alpha/2} + \frac{\sigma}{\epsilon n} \operatorname{polylog}\left(\frac{n}{\alpha}\right)$$

For the sample complexity, we need

$$n \ge c \min\left\{ \max\left(\frac{1}{\epsilon} \log\left(\frac{R\sqrt{n}}{\sigma_{\min}\alpha}\right), \frac{1}{\epsilon} \log\left(\frac{\sigma_{\max}\sqrt{n}}{\sigma_{\min}\alpha}\right)\right), \frac{1}{\epsilon} \log\left(\frac{\sqrt{n}}{\delta\alpha}\right) \right\},\$$

It suffices to have

$$n \ge c \min\left\{ \max\left(\frac{1}{\epsilon} \log\left(\frac{R}{\epsilon \sigma_{\min}\alpha}\right), \frac{1}{\epsilon} \log\left(\frac{\sigma_{\max}}{\epsilon \sigma_{\min}\alpha}\right)\right), \frac{1}{\epsilon} \log\left(\frac{1}{\epsilon \delta \alpha}\right) \right\}.$$

6 Lower Bounds for confidence intervals with known variance

In this section, we prove lower bounds on the expected size of $(1 - \alpha)$ -level confidence sets obtained with and without differential privacy. We begin with deriving lower bounds on the expected size of confidence sets without the privacy requirement, followed by the case when the confidence sets are required to be (ϵ, δ) differentially private. The non-private lower bounds are useful for comparisons and to understand the price that we have to pay to ensure differential privacy.

6.1 Lower bounds on Confidence Intervals without privacy

In this subsection, we derive lower bounds on confidence sets for the normal mean where there are no privacy constraints. The non-private confidence set estimation problem that we consider here differs from the standard problem of a estimating confidence set for the mean of the normal population. For ϵ -differential privacy, we require $\mu \in (-R, R)$. The corresponding problem without privacy is called the *bounded normal mean problem* and has a long history, see for example Pratt et al. (1963); Evans et al. (2005); Schafer and Stark (2009) for the problem of estimating a minimax confidence interval for the bounded normal mean, and Casella and Strawderman (1981); Bickel et al. (1981); Marchand and Strawderman (2004) for the problem of point estimation of the bounded normal mean.

It is well known (see for example Lehmann and Romano (2006)), that the classic confidence interval $X \pm \sigma/\sqrt{n}$ is of minimax expected size among all $1 - \alpha$ level confidence sets for the normal mean with known variance, when $\mu \in \mathbb{R}$. On the other hand, when we assume that $\mu \in (-R, R)$, the classic confidence interval is no longer minimax. When $R \leq 2\sigma z_{1-\alpha/2}$, Evans et al. (2005) show that the so called *truncated Pratt interval* (Pratt et al., 1963) is of expected minimax length. In general, the minimax confidence set is not known.

However, we are interested in estimating a confidence set in a regime where R is on the order of, or much larger than the standard deviation, i.e. $R = \Theta(\sigma)$ or $R \gg \sigma$. Under this regime, we derive a lower bound on the expected size of any $(1 - \alpha)$ confidence set of the bounded mean. The main result of this section is stated below.

Theorem 6.1. Let $X_1, \ldots, X_n \stackrel{iid}{\sim} N(\mu, \sigma^2)$ where $\mu \in (-R, R)$, σ^2 is known and $R > c\sigma^2$ for a fixed (but arbitrarily small) constant c > 0. Let $I = I(X_1, \ldots, X_n)$ be any measurable set where $I \subseteq (-R, R)$ such that for every $\alpha \in (0, 1/2)$, and $\mu \in (-R, R)$, if,

$$\mathbb{P}_{\underline{X} \sim N(\mu, \sigma^2)} (I(X_1, \dots, X_n) \ni \mu) \ge 1 - \alpha.$$
$$\mathbb{E}_{\underline{X} \sim N(\mu, \sigma^2)} [|I(X_1, \dots, X_n)|] \ge \frac{2\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}} - \tilde{\mathcal{O}}\left(\frac{1}{n}\right).$$

Theorem 6.1 states that when $R = \Omega(\sigma)$, the lower bound on the expected size of a confidence set without privacy is equal to the length of the classic interval minus a term that goes to 0 at the rate 1/n, upto logarithmic factors. Before we present a proof of Theorem 6.1, we need some intermediate results that are stated below:

Fact 6.1 (See for example Hogg and Craig (1995)). If $\mu \sim N(0, \gamma^2)$, and $X_1, \ldots, X_n | \mu \stackrel{iid}{\sim} N(\mu, \sigma^2)$, then the conditional distribution of $\mu | X_1, \ldots, X_n$ is a normal with mean μ_p and variance σ_p^2 where

$$\mu_p = \frac{n/\sigma^2}{n/\sigma^2 + 1/\gamma^2} \cdot \bar{X}$$

and

$$\sigma_p^2 = \frac{1}{n/\sigma^2 + 1/\gamma^2}$$

Proposition 6.1. Let $X \sim N(\mu, \sigma^2)$ and I(X) be any measurable set, then $\mathbb{P}(X \in I) \leq 2\Phi(|I|/(2\sigma)) - 1$, where $\Phi(\cdot)$ is the cdf of a standard normal distribution and |I| is the Lebesgue measure of the set I.

Proof sketch. Let I be any measurable set. Since the normal distribution is symmetric and unimodal at μ , the density is highest at μ and decreasing away from μ . Thus, probability mass in I will be maximized when I is an interval centered at μ . Hence an interval I with the highest mass is $[\mu - |I|/2, \mu + |I|/2]$. The result follows by computing the probability of this interval.

Proposition 6.2. Let $\alpha \in (0, 1/2)$ and $0 < \Delta < \alpha$, then

$$z_{1-(\alpha+\Delta)/2} \ge z_{1-\alpha/2} - \frac{c\Delta}{\alpha}$$

Proof. Apply Proposition 4.1 with $\alpha' = \alpha + \Delta$.

We are now ready to prove Theorem 6.1. The key idea in proving this result is to use a normal prior distribution on μ and obtain a posterior distribution of μ given the data. For a normal prior, the posterior has a simple form given by Fact 6.1. Then using Proposition 6.1, we obtain an upper bound on how much posterior probability mass can be packed into any set I. On the other hand, since the set must have $1 - \alpha$ coverage, we have a lower bound on the posterior probability mass contained inside I whenever $\mu \in [-R, R]$. We have to eliminate the case when μ does not lie [-R, R]. Combining this with Jensen's inequality gives the result.

Proof. Let $\mu \sim N(0, R^2\beta^2)$ where β will be specified later. Let $\underline{X}|\mu = (X_1, \ldots, X_n)|\mu \sim N(\mu, \sigma^2)$ and $I(\underline{X})$ be any subset of (-R, R). With a minor abuse of notation, let $\mathbb{P}_{\underline{X}}(\cdot)$ denote the probability with respect to \underline{X} and $\mathbb{E}_{\mu}[\cdot]$ denote the expectation taken with respect to the distribution of μ . We have the following lower bound:

$$\mathbb{E}_{\mu \sim N(0, R^{2} \beta^{2})} \left[\mathbb{P}_{\underline{X} \sim N(\mu, \sigma^{2})} \left(\mu \in I \right) \right] \geq \mathbb{E}_{\mu} \left[\mathbb{P}_{\underline{X}} \left(\mu \in I \right) \cdot \mathbb{I}(\mu \in (-R, R)) \right] \\
\geq (1 - \alpha) \cdot \mathbb{P}_{\mu}(\mu \in (-R, R)) \\
\geq (1 - \alpha) \cdot \gamma,$$
(14)

where $\gamma = \mathbb{P}_{\mu}(\mu \in (-R, R))$. Next, using the Fact 6.1, $\mu | \underline{X} \sim N(\mu_p, \sigma_p^2)$ where

$$\sigma_p = \frac{\sigma}{\sqrt{n}} \frac{1}{\sqrt{1 + \sigma^2/(nR^2\beta^2)}},$$

we have the following upper bound:

$$\mathbb{E}_{\mu} \left[\mathbb{P}_{\underline{X}}(\mu \in I) \right] = \mathbb{E}_{\underline{X}} \left[\mathbb{P}_{\mu | \underline{X}}(\mu \in I) \right] \\
\leq \mathbb{E}_{\underline{X}} \left[2\Phi \left(\frac{|I|}{2\sigma_p} - 1 \right) \right] \quad (By \text{ Proposition 6.1}) \\
\leq 2\Phi \left(\frac{E[|I|]}{2\sigma_p} \right) - 1 \quad (By \text{ Jensen's inequality applied to } \Phi(x) \text{ for } x > 0 \quad (15)$$

Combining the upper bound and the lower bound from equations 15 and 14, we get,

$$(1-\alpha)\gamma \le 2\Phi\left(\frac{E[|I|]}{2\sigma_p}\right) - 1 \implies \mathbb{E}[|I|] \ge \Phi^{-1}\left(\frac{\gamma(1-\alpha)+1}{2}\right)2\sigma_p \tag{16}$$

Let us consider the term $\Phi^{-1}(\cdot)$ in the lower bound. Let $\beta = 1/\log(\sqrt{n})$. By the Gaussian tail bound in Proposition 7.3, we have

$$\gamma \ge 1 - 2 \exp\left(-1/(2\beta^2)\right) = 1 - \Omega(1/\sqrt{n}).$$

Hence $\gamma(1-\alpha) = 1 - \alpha - \Omega(1/\sqrt{n})$ and from Proposition 6.2, we have

$$\Phi^{-1}\left(\frac{\gamma(1-\alpha)+1}{2}\right) \ge \Phi^{-1}\left(1-\alpha/2 - \Omega(1/\sqrt{n})\right)$$
$$\ge \Phi^{-1}\left(1-\alpha/2\right) - \frac{c/\sqrt{n}}{\alpha}$$
$$\ge \Phi^{-1}\left(1-\alpha/2\right) - \frac{c}{\alpha\sqrt{n}}$$

Next, let us now consider the σ_p term. We have,

$$\sigma_p = \frac{\sigma}{\sqrt{n}} \cdot \frac{1}{\sqrt{1 + \sigma^2/(nR^2\beta^2)}}$$
$$\geq \frac{\sigma}{\sqrt{n}} \left(1 - \Omega\left(\frac{\sigma^2}{nR^2\beta^2}\right) \right)$$
$$\geq \frac{\sigma}{\sqrt{n}} \left(1 - \Omega\left(\frac{\sigma^2\log n}{nR^2}\right) \right)$$
$$\geq \frac{\sigma}{\sqrt{n}} \left(1 - \tilde{\Omega}\left(\frac{1}{n}\right) \right)$$

when $R = \mathcal{O}(\sigma)$. Hence combining the two terms and applying them to equation 16, we have

$$\mathbb{E}[|I|] \ge \frac{2\sigma}{\sqrt{n}} \cdot \left(1 - \tilde{\Omega}\left(\frac{1}{n}\right)\right) \cdot \left(\Phi^{-1}\left(1 - \frac{\alpha}{2} - \Omega\left(\frac{1}{n}\right)\right)\right)$$
$$\ge \frac{2\sigma}{\sqrt{n}}\Phi^{-1}\left(1 - \frac{\alpha}{2}\right) - \tilde{\Omega}\left(\frac{1}{n}\right)$$
(17)

6.2 Lower bounds on confidence sets with privacy

In this subsection, we derive lower bounds on the expected size of any $(1 - \alpha)$ -level (ϵ, δ) -differentially private confidence sets. The following theorem is the main result of this section.

Theorem 6.2. Let $M(X_1, \ldots, X_n)$ be any (ϵ, δ) -DP algorithm that outputs a random measurable set $S(X_1, \ldots, X_n) \subset (-R, R)$ such that, for every $\alpha \in (0, 1)$, and $\mu \in (-R, R)$, whenever $X_1, \ldots, X_n \stackrel{iid}{\sim} N(\mu, \sigma^2)$ where $\mu \in (-R, R)$ and σ^2 is known, if

1.
$$\mathbb{P}_{\substack{X \sim N(\mu, \sigma^2) \\ M}} (S(X_1, \dots, X_n) \ni \mu) \ge 1 - \alpha$$

2.
$$\mathbb{E}_{\substack{X \sim N(\mu, \sigma^2) \\ M}} [|S(X_1, \dots, X_n)|] \le \beta$$

then,

1. For all $\alpha \in (0, \frac{1}{2})$, and $\delta < \alpha/2n$,

$$\beta \ge c \cdot \min\left(\frac{\sigma}{\epsilon n} \log\left(\frac{1}{\alpha}\right), R\right).$$

2. If $\beta < \sigma < R$, then,

$$n \ge c_1 \cdot \min\left(\frac{1}{\epsilon}\log\left(\frac{1}{\alpha}\right), \frac{1}{\epsilon}\log\left(\frac{1}{\delta}\right)\right)$$

and

$$n \ge c_2 \cdot \min\left(\frac{1}{\epsilon} \log\left(\frac{R}{\sigma}\right), \frac{1}{\epsilon} \log\left(\frac{1}{\delta}\right)\right)$$

Theorem 6.2 gives a lower bound on the expected size of any $(1 - \alpha)$ -level confidence set produced by a differentially private algorithm. Note that the lower bounds apply to any set and not just intervals.

We will now compare the lower bounds with the upper bounds obtained in the previous sections. Let us focus on the case of known variance. Theorem 4.1 asserts the existence of a differentially private algorithm that outputs a $(1 - \alpha)$ -level confidence interval of length β such that if

$$n \ge \frac{c}{\epsilon} \min\left\{\log\left(\frac{R}{\sigma}\right), \log\left(\frac{1}{\delta}\right)\right\} + \frac{1}{\epsilon}\log\left(\frac{1}{\alpha}\right),$$

then,

$$\beta \le \max\left\{\frac{\sigma}{\sqrt{n}}\mathcal{O}\left(\sqrt{\log\left(\frac{1}{\alpha}\right)}\right), \frac{\sigma}{\epsilon n} \text{polylog}\left(\frac{n}{\alpha}\right)\right\}.$$

Note that the upper bound on the width of β is a maximum two terms. The first term $\mathcal{O}\left((\sigma/\sqrt{n}) \cdot \sqrt{\log(1/\alpha)}\right)$ is needed even without privacy, since it is the optimal width of the interval without privacy, as shown in Theorem 6.1. The second term is $\sigma/(\epsilon n) \cdot \log(1/\alpha)$ up to $\operatorname{polylog}(n, 1/\alpha)$ factors. The lower bound in Part (1) of Theorem 6.2 shows that this term is unavoidable, unless we produce a very large interval of length proportional to the original range (-R, R) of μ .

Part (2) of Theorem 6.2 gives a lower bound on n if we require β to be smaller than σ and R. Note that the lower bound on the sample size matches the sample size requirement of Theorem 4.1 to obtain a non-trivial confidence interval. In the unknown variance case, the lower bounds of Theorem 6.2 apply. We are interested in the regime when $R \ge \Omega(\sigma_{\max})$. In this case the known-variance lower bound on sample complexity for $\sigma = \sigma_{\min}$ matches our unknown-variance upper bound given in Theorem 5.1.

Before we present a proof of Theorem 6.2, we will introduce some additional notation for the proof. Let X_1, \ldots, X_n be iid samples from a distribution $\mathbb{D}_{(\theta,\gamma)}$. As before, we abuse the notation and use \mathbb{D}_{θ} to denote the distribution when γ is a nuisance parameter. Let $M(x_1, \ldots, x_n)$ be an (ϵ, δ) -differentially private mechanism that runs on the realization x_1, \ldots, x_n and let $Q(E|X_1 = x_1, \ldots, X_n = x_n)$ denote the conditional distribution of M given x_1, \ldots, x_n , where E is any event. For any event E, let

$$\mathbb{M}_{\theta}(E) = \int Q(E|X_1, \dots, X_n) d\mathbb{P}_{\theta}(X_1, \dots, X_n)$$
(18)

be the marginal distribution on the outputs induced by the DP mechanism when the data are generated from the distribution \mathbb{D}_{θ} ; here \mathbb{P}_{θ} is short hand to denote the probability computed under the distribution \mathbb{D}_{θ} . Similarly \mathbb{M}_{θ} is a shorthand to denote the probability computed uner the distribution \mathbb{D}_{θ} and the mechanism M. Let \mathbb{D}_{θ_0} and \mathbb{D}_{θ_1} be two distributions, and denote the total variation distance between \mathbb{D}_{θ_0} and \mathbb{D}_{θ_1} by $||\mathbb{D}_{\theta_0} - \mathbb{D}_{\theta_1}||_{tv}$.

To prove Theorem 6.2 we need a lemma that bounds the multiplicative distances between two differentially private marginal distributions induced by two different distributions \mathbb{D}_{θ_0} and \mathbb{D}_{θ_1} on the data. This is done in Lemma 6.1 below. We built on Lemma 4.12 of Bun et al. (2015) where they use a "secrecy of sampling argument" that appeared in Kasiviswanathan et al. (2011) and Smith (2009) to amplify privacy when a differentially private algorithm is run on a re-sampled version of a dataset \underline{X} . We use a similar technique to compute the privacy amplification for datasets generated by two different distributions. Lemma 4.12 of Bun et al. (2015) can be recovered from our result by letting \mathbb{D}_{θ_i} be an empirical cdf of \underline{X} and \underline{X}' of two neighboring datasets.

Lemma 6.1. For every pair of distributions \mathbb{D}_{θ_0} and \mathbb{D}_{θ_1} , every (ϵ, δ) -differentially private algorithm $M(x_1, \ldots, x_n)$, if \mathbb{M}_{θ_0} and \mathbb{M}_{θ_1} are two induced marginal distributions on the output of M evaluated on input dataset X_1, \ldots, X_n sampled iid from \mathbb{D}_{θ_0} and \mathbb{D}_{θ_1} respectively, $\epsilon' = 6\epsilon n ||\mathbb{D}_{\theta_0} - \mathbb{D}_{\theta_1}||_{tv}$ and $\delta' = 4e^{\epsilon'} n\delta ||\mathbb{D}_{\theta_0} - \mathbb{D}_{\theta_1}||_{tv}$, then, for every event E.

$$\mathbb{M}_{\theta_0}(E) \le e^{\epsilon'} \mathbb{M}_{\theta_1}(E) + \delta'.$$

Proof. We will first construct a coupling between \mathbb{D}_{θ_0} and \mathbb{D}_{θ_1} that allows us to control the hamming distance between iid samples generated from \mathbb{D}_{θ_0} and \mathbb{D}_{θ_1} . Let us start with some notation. Let $p = ||\mathbb{D}_{\theta_0} - \mathbb{D}_{\theta_1}||_{tv}$, $F = \max(\mathbb{D}_{\theta_0} - \mathbb{D}_{\theta_1}, 0), \ G = \max(\mathbb{D}_{\theta_1} - \mathbb{D}_{\theta_0}, 0), \ C = \min(\mathbb{D}_{\theta_0}, \mathbb{D}_{\theta_1})$. It is easy to see that $\mathbb{D}_{\theta_0} = F + C$ and $\mathbb{D}_{\theta_1} = G + C$. Consider the following algorithm to generate 2n samples:

- 1. Generate H_1, \ldots, H_n iid Bernoulli(p) and let $H = \sum_{i=1}^n H_n$.
- 2. For i = 1 to n,
 - (a) If $H_i = 1$, sample $X_i \propto F$ and $X'_i \propto G$
 - (b) If $H_i = 0$, sample $X_i \propto C$ and set $X'_i = X_i$.

Here $X \propto F$ means that X is generated from a distribution defined by normalizing F. Under this construction, one can verify the following:

- 1. $\underline{X} := (X_1, \dots, X_n) \stackrel{iid}{\sim} \mathbb{D}_{\theta_0}$ 2. $\underline{X}' := (X'_1, \dots, X'_n) \stackrel{iid}{\sim} \mathbb{D}_{\theta_1}$
- 3. $d(\underline{X}, \underline{X'}) = H$.

Fix any event E. Let

$$q_{\theta_j}(h) = \mathbb{P}_{\theta_j}(E|H=h) := \int_{\underline{x}} \mathbb{Q}(E|\underline{X}=\underline{x}) d\mathbb{P}_{\theta_j}(\underline{X}|H=h)$$

for $j \in \{0, 1\}$ and

$$p(h) = \mathbb{P}(H=h) = \binom{n}{h} p^h (1-p)^{n-h},$$

since $H \sim Binomial(n, p)$. For $j \in \{0, 1\}$, we have, by definition,

$$\mathbb{M}_{\theta_j}(E) = \sum_{h=0}^n q_{\theta_j}(h) p(h)$$

Claim 1: For $j \in \{0,1\}$, $q_{\theta_j}(h) \leq e^{\epsilon} q_{\theta_j}(h-1) + \delta$ for $h = 1, \ldots, n$, and $q_{\theta_1}(0) = q_{\theta_0}(0)$. We will defer the proof of Claim 1 towards the end. By Claim 1, for $\theta_j \in \{0,1\}$, we have

$$q_{\theta_j}(h) \le e^{h\epsilon} q_{\theta_j}(0) + \frac{e^{h\epsilon} - 1}{e^{\epsilon} - 1} \delta$$
(19)

In the following, we will use the standard fact of the moment generating function of a Binomial distribution: Fact 6.2. If $H \sim Binomial(n, p)$ then, for anyt > 0,

$$\mathbb{E}\left[e^{tH}\right] = (1 - p + p \cdot e^t)^n$$

Consider,

$$\mathbb{M}_{\theta_j}(E) = \sum_{h=0}^n p(h)q_{\theta_j}(h) = \mathbb{E}\left[q_{\theta_j}(H)\right]
\leq \mathbb{E}\left[e^{H\epsilon}q_{\theta_j}(0) + \frac{e^{H\epsilon} - 1}{e^{\epsilon} - 1}\delta\right] (\text{From equation 19})
= q_{\theta_j}(0) \cdot \mathbb{E}\left[e^{H\epsilon}\right] + \frac{\delta}{e^{\epsilon} - 1} \cdot \left(\mathbb{E}\left[e^{H\epsilon}\right] - 1\right)
= q_{\theta_j}(0) \cdot (1 - p + p \cdot e^{\epsilon})^n + \frac{\delta}{e^{\epsilon} - 1} \cdot \left((1 - p + p \cdot e^{\epsilon})^n - 1\right)$$
(20)

Similarly, we can show that,

$$\mathbb{M}_{\theta_j}(E) \ge q_{\theta_j}(0) \left(1 - p + p \cdot e^{-\epsilon}\right)^n + \frac{\delta}{e^{-\epsilon} - 1} \cdot \left((1 - p + p e^{-\epsilon})^n - 1\right)$$
(21)

Combining inequalities 20 and 21, we get:

$$\mathbb{M}_{\theta_0}(E) \le \left(\frac{1-p+p \cdot e^{\epsilon}}{1-p+p \cdot e^{-\epsilon}}\right)^n \cdot \left(\mathbb{M}_{\theta_1}(E) + \frac{1-(1-p+p \cdot e^{-\epsilon})^n}{1-e^{-\epsilon}} \cdot \delta\right) + \frac{(1-p+p \cdot e^{\epsilon})^n - 1}{e^{\epsilon} - 1} \cdot \delta \tag{22}$$

Finally, we have,

$$n\log\left(\frac{1-p+p\cdot e^{\epsilon}}{1-p+p\cdot e^{-\epsilon}}\right) \le n\cdot 6\epsilon p$$

and

$$\begin{split} e^{6\epsilon n \cdot p} \cdot \frac{1 - (1 + p \cdot (e^{\epsilon} - 1))^n}{1 - e^{-\epsilon}} \cdot \delta + \frac{(1 + p \cdot (e^{\epsilon} - 1))^n - 1}{e^{\epsilon} - 1} \cdot \delta \\ \leq e^{6\epsilon n \cdot p} \cdot \frac{1 - \exp(2np \cdot (e^{-\epsilon} - 1))}{1 - e^{-\epsilon}} \cdot \delta + \frac{\exp(2np \cdot (e^{\epsilon} - 1)) - 1}{e^{\epsilon} - 1} \cdot \delta \\ \leq e^{6\epsilon n \cdot p} \cdot 2np \cdot \delta + 2np \cdot \delta \\ \leq e^{6\epsilon n \cdot p} \cdot 4np \cdot \delta, \end{split}$$

which shows that

$$\mathbb{M}_{\theta_0}(E) \le e^{\epsilon'} \mathbb{M}_{\theta_1}(E) + \delta'$$

for $\epsilon' = 6np \cdot \epsilon$ and $\delta' = 4e^{\epsilon'}np \cdot \delta$. We will now prove Claim 1.

Proof of Claim 1. We will prove the claim for j = 0, the other case is similar. Let us introduce some notation. Fix a $(h_1, \ldots, h_n) \in \{0, 1\}^n$. Let $I' = \{i : h_i = 1\}$, $J = \{i : h_i = 0\}$ and let r be any index in I'. Let $I = I'/\{r\}$ and consider the following partition of \underline{X} into three parts:

$$\underline{X} = (\underline{X}_I, X_r, \underline{X}_J)$$

where \underline{X}_I is the vector \underline{X} subsetted by the indices in I. By definition of the coupling, $\underline{X}_I \stackrel{iid}{\sim} F, X_r \sim F,$ $\underline{X}_J \stackrel{iid}{\sim} C$. Now let $X'_r \sim C$ and let

$$\underline{X}' = (\underline{X}_I, X'_r, \underline{X}_J).$$

Also, let h'_1, \ldots, h'_n be the binary indicators corresponding to <u>X</u>. By construction, we have the following:

- 1. $h_i = h'_i$ for all $i \neq r$ 2. $h_r = 1$ and $h_r = 0$ 3. $\sum_{i=1}^n h_i = h$ and $\sum_{i=1}^n h'_i = h - 1$ 4. $\mathbb{P}_{\theta_i}(\underline{X}|H_1 = h_1, \dots, H_n = h_n) = \mathbb{P}_F(\underline{X}_I)\mathbb{P}_F(X_r)\mathbb{P}_C(\underline{X}_I)$
- 5. $\mathbb{P}_{\theta_i}(\underline{X}'|H_1 = h'_1, \dots, H_n = h'_n) = \mathbb{P}_F(\underline{X}_I)\mathbb{P}_F(X'_r)\mathbb{P}_C(\underline{X}_I)$

Now consider the following:

$$\begin{split} \mathbb{P}_{\theta_{j}}(E|H_{1} = h_{1}, \dots, H_{n} = h_{n}) \\ &= \int_{\underline{x}} \mathbb{Q}(E|\underline{X} = \underline{x}) d\mathbb{P}_{\theta_{j}}(\underline{X}|H_{1} = h_{1}, \dots, H_{n} = h_{n}) \\ &= \int_{\underline{x}_{I}} \int_{x_{r}} \int_{\underline{x}_{J}} \mathbb{Q}(E|\underline{x}_{I}, x_{r}, \underline{x}_{J}) d\mathbb{P}_{F}(\underline{X}_{I}) d\mathbb{P}_{F}(X_{r}) \mathbb{P}_{C}(\underline{X}_{J}) \\ &\leq \int_{\underline{x}_{I}} \int_{x_{r}} \int_{\underline{x}_{J}} \left(e^{\epsilon} \mathbb{Q}(E|\underline{x}_{I}, x_{r}', \underline{x}_{J}) + \delta \right) d\mathbb{P}_{F}(\underline{X}_{I}) d\mathbb{P}_{F}(X_{r}) \mathbb{P}_{C}(\underline{X}_{J}) \\ &\leq \int_{\underline{x}_{I}} \int_{\underline{x}_{J}} \left(e^{\epsilon} \mathbb{Q}(E|\underline{x}_{I}, x_{r}', \underline{x}_{J}) + \delta \right) d\mathbb{P}_{F}(\underline{X}_{I}) d\mathbb{P}_{C}(\underline{X}_{J}) \\ &\leq \int_{\underline{x}_{I}} \int_{x_{r}'} \int_{\underline{x}_{J}} \left(e^{\epsilon} \mathbb{Q}(E|\underline{x}_{I}, x_{r}', \underline{x}_{J}) + \delta \right) d\mathbb{P}_{F}(\underline{X}_{I}) d\mathbb{P}_{C}(X_{r}') \mathbb{P}_{C}(\underline{X}_{J}) \\ &\leq \int_{\underline{x}} \left(e^{\epsilon} \mathbb{Q}(E|\underline{x}') + \delta \right) d\mathbb{P}_{\theta_{j}}(\underline{X}|H_{1} = h_{1}', \dots, H_{n} = h_{n}') \\ &\leq e^{\epsilon} \mathbb{P}_{\theta_{j}}(E|H_{1} = h_{1}', \dots, H_{n} = h_{n}') + \delta. \end{split}$$

Hence we have shown that

$$\mathbb{P}_{\theta_j}(E|H_1 = h_1, \dots, H_n = h_n, H = h) \le e^{\epsilon} \mathbb{P}_{\theta_j}(E|H_1 = h'_1, \dots, H_n = h'_n, H = h - 1) + \delta$$

Taking expectations on both sides with respect to (H_1, \ldots, H_n) , we get

$$\mathbb{P}_{\theta_j}(E|H=h) \le e^{\epsilon} \mathbb{P}_{\theta_j}(E|H=h-1) + \delta$$

which proves the claim.

Let us consider a corollary of Lemma 6.1 when \mathbb{P}_{θ_0} and \mathbb{P}_{θ_1} are two normal distributions with $\theta_0 = (\mu_0, \sigma^2)$ and $\theta_1 = (\mu_1, \sigma^2)$. Since the total variation distance between $N(\mu_0, \sigma^2)$ and $N(\mu_1, \sigma^2)$ is upper bounded by $|\mu_0 - \mu_1|/\sigma$ (see for example DasGupta (2008)), we get the following Corollary:

Corollary 1. Let \mathbb{P}_{θ_0} and \mathbb{P}_{θ_1} be two normal distributions where $\theta_0 = (\mu_0, \sigma^2)$ and $\theta_1 = (\mu_1, \sigma^2)$, where μ_1 and μ_0 are the means μ_0 and σ^2 is the variance. Let $M(x_1, \ldots, x_n)$ be any (ϵ, δ) -differentially private algorithm that induces the marginal distribution \mathbb{M}_{θ_0} and \mathbb{M}_{θ_1} as defined in equation 18. For any event E, we have

$$\mathbb{M}_{\theta_0}(E) \le e^{6\epsilon n \cdot k} \left(\mathbb{M}_{\theta_1}(E) + 4n\delta \cdot k \right),$$

where

$$k = \min\left\{\frac{|\mu_0 - \mu_1|}{\sigma}, 1\right\}.$$

We are now ready to prove Theorem 6.2.

Proof of Theorem 6.2. By definition of measure of a set S,

$$|S| = \int_{-R}^{R} \mathbb{I}(\mu \in S) d\mu,$$

where $\mathbb{I}(\cdot)$ is the indicator function. Hence, taking expectation on both sides with respect to $S \to M(X_1, \ldots, X_n)$ and $X_1, \ldots, X_n \sim N(\mu_0, \sigma^2)$, and changing the order of integration, (since $\mathbb{I}(\cdot)$ is non-negative, one can apply Tonelli's theorem), we get

$$\mathbb{E}_{\mu_0}[|S|] = \int_{-R}^{R} \mathbb{M}_{\mu_0} \left(\mu \in S\right) d\mu.$$

We will start by proving the first bound in Part (1). Let μ_0 and μ_1 be any two points in [-R, R]. Note that, $\forall \mu \in [\mu_0, \mu_1]$, we have,

$$\exp\left(6\epsilon n \frac{|\mu_0 - \mu|}{\sigma}\right) \le \exp\left(6\epsilon n \frac{|\mu_0 - \mu_1|}{\sigma}\right) \tag{23}$$

and by Corollary 1, we have,

$$\mathbb{M}_{\mu_0}(\mu \notin S) \leq e^{6\epsilon n \cdot k} \left(\mathbb{M}_{\mu}(\mu \notin S) + 4n\delta \cdot k \right) \\
\leq e^{6\epsilon n \cdot |\mu_0 - \mu| / \sigma} \left(\mathbb{M}_{\mu}(\mu \notin S) + 4n\delta \cdot 1 \right)$$
(24)

where $k = \min\{1, |\mu_0 - \mu|/\sigma\}$ and we have used an upper bound of 1 for k in the second term. Then we have,

$$\mathbb{E}_{\mu_{0}}[|S|] \geq \int_{\mu_{0}}^{\mu_{1}} \mathbb{M}_{\mu_{0}} (\mu \in S) d\mu \\
\geq \int_{\mu_{0}}^{\mu_{1}} (1 - \mathbb{M}_{\mu_{0}}(\mu \notin S)) d\mu \\
\geq |\mu_{1} - \mu_{0}| - \int_{\mu_{0}}^{\mu_{1}} (\mathbb{M}_{\mu}(\mu \notin S) + 4n\delta) \cdot e^{6\epsilon n \cdot |\mu_{0} - \mu| / \sigma} d\mu, \text{ (Equation 24)} \\
\geq |\mu_{1} - \mu_{0}| - (\alpha + 4n\delta) \cdot \int_{\mu_{0}}^{\mu_{1}} e^{6\epsilon n |\mu_{0} - \mu_{1}| / \sigma} d\mu \text{ (Equation 23)} \\
\geq |\mu_{1} - \mu_{0}| - (\alpha + 4n\delta) \cdot |\mu_{1} - \mu_{0}| \cdot e^{6\epsilon n |\mu_{0} - \mu_{1}| / \sigma} \\
= |\mu_{1} - \mu_{0}| \cdot \left(1 - (\alpha + 4n\delta) \cdot e^{6\epsilon n |\mu_{0} - \mu_{1}| / \sigma}\right)$$
(25)

We will consider two cases depending on how large R is relative to $\sigma/(\epsilon n)$:

Case 1: $2R \ge (\sigma/(6\epsilon n)) \cdot \log(1/(4\alpha))$. Let μ_0 and μ_1 be two points such that

$$|\mu_1 - \mu_0| = \frac{\sigma}{6\epsilon n} \cdot \log\left(\frac{1}{4\alpha}\right).$$

By the assumption on R two such points always exist. Substituting this in equation 25, we get,

$$\mathbb{E}_{\mu_0}[|S|] \ge |\mu_1 - \mu_0| \cdot \left(1 - (\alpha + 4n\delta) \cdot e^{6\epsilon n|\mu_0 - \mu_1|/\sigma}\right)$$
$$\ge \frac{\sigma}{6\epsilon n} \cdot \log\left(\frac{1}{4\alpha}\right) \left(1 - (\alpha + 4n\delta) \cdot \left(\frac{1}{4\alpha}\right)\right)$$
$$\ge \frac{\sigma}{6\epsilon n} \cdot \log\left(\frac{1}{4\alpha}\right) \left(\frac{3}{4} - \frac{n\delta}{\alpha}\right)$$
$$\ge \frac{\sigma}{24\epsilon n} \cdot \log\left(\frac{1}{4\alpha}\right) \text{ (Since } \delta < \alpha/(2n) \text{ which implies } 3/4 - (n\delta/\alpha) > 1/4)$$

Case 2: $2R < \sigma/(6\epsilon n) \cdot \log(1/(4\alpha)) < \infty$. Let $\mu_0 = -R$ and $\mu_1 = R$. Substituting this in equation 25, we get:

$$\mathbb{E}_{\mu_0}[|S|] \ge |\mu_1 - \mu_0| \cdot \left(1 - (\alpha + 4n\delta) \cdot e^{6\epsilon n |\mu_0 - \mu_1|/\sigma}\right)$$
$$= 2R \cdot \left(1 - (\alpha + 4n\delta) \cdot e^{12\epsilon n R/\sigma}\right)$$
$$\ge 2R \cdot \left(1 - (\alpha + 4n\delta) \left(\frac{1}{4\alpha}\right)\right)$$
$$\ge 2R \left(\frac{3}{4} - \frac{n\delta}{\alpha}\right)$$
$$\ge \frac{R}{2} \text{ (Since } \delta < \alpha/(2n) \text{ which implies } 3/4 - (n\delta/\alpha) > 1/4)$$

Combining the two cases , we have $\beta > R/2$ or $\beta > \sigma/(24\epsilon n) \cdot \log(1/(4\alpha))$ which proves the bound in Part (1). Let us now prove the bounds in Part (2). We will first show,

$$n \ge c \cdot \min\left\{\frac{1}{\epsilon} \log\left(\frac{R}{\sigma}\right), \frac{1}{\epsilon} \log\left(\frac{1}{\delta}\right)\right\}$$

By assumption, since $\beta < \sigma < R$, we have,

$$\sigma > \beta \ge \int_{-R}^{R} \mathbb{M}_{\mu_0}(\mu \in S) d\mu$$

$$\ge \int_{-R}^{R} \left(e^{-6\epsilon n} \cdot \mathbb{M}_{\mu}(\mu \in S) - 4n\delta \right) d\mu \text{ (By Corollary 1)}$$

$$\ge \int_{-R}^{R} \left(e^{-6\epsilon n} \cdot (1 - \alpha) - 4n\delta \right) d\mu$$

$$\ge 2R \cdot (1 - \alpha) \cdot e^{-6\epsilon n} - 4n\delta \cdot (2R)$$

$$\ge Re^{-6\epsilon n} - 4n\delta \cdot (2R)$$

Hence we have, either

$$\sigma > \frac{R}{2}e^{-6\epsilon n}$$
 or $4n\delta \cdot (2R) > \frac{R}{2} \cdot e^{-6\epsilon n}$

which implies that

$$n \ge \frac{c_1}{\epsilon} \log\left(\frac{R}{\sigma}\right) \text{ or } n \ge \frac{c_2}{\epsilon} \log\left(\frac{1}{\delta}\right).$$

Next we will show that

$$n \ge c \cdot \min\left\{\frac{1}{\epsilon} \log\left(\frac{1}{\alpha}\right), \frac{1}{\epsilon} \log\left(\frac{1}{\delta}\right)\right\}$$

Let μ_0 be a fixed point in (-R, R). The fact that if $\mathbb{E}_{\mu_0}(|S|) < R$, implies that there exists a $\mu_1 \in (-R, R)$ such that

$$\mathbb{M}_{\mu_0}\left(\mu_1 \notin S\right) > \frac{1}{2}$$

Indeed, if not, then

$$\mathbb{E}_{\mu_0}(|S|) = \int_{-R}^{R} \mathbb{M}_{\mu_0} \ (\mu \in S) \ d\mu \ge \frac{1}{2} 2R = R.$$

Thus from Corollary 1 we have

$$\frac{1}{2} < \mathbb{M}_{\mu_0} \left(\mu_1 \notin S \right) < e^{6\epsilon n \cdot |\mu_0 - \mu_1|/\sigma} \cdot \mathbb{M}_{\mu_1} \left(\mu_1 \notin S \right) + 4e^{6\epsilon n \cdot |\mu_0 - \mu_1|/\sigma} \cdot n\delta \cdot ||\mathbb{P}_{\mu_1} - \mathbb{P}_{\mu_0}||_{tv}$$
$$\leq (\alpha + 4n\delta) \cdot e^{6\epsilon n \frac{|\mu_0 - \mu_1|}{\sigma}} \leq (\alpha + 4n\delta) \cdot e^{6\epsilon n}.$$

This gives us

$$\alpha > \frac{1}{2}e^{-6\epsilon n} - 4\delta n$$

Hence either $4n\delta > e^{-6\epsilon n}/4$ or $\alpha > e^{-6\epsilon n}/4$, which gives us either $n \ge (c_1/\epsilon) \cdot \log(1/4\alpha)$ or $n \ge (c_2/\epsilon) \cdot \log(1/\delta)$.

7 Appendix: Basic distributions and tail bounds

We make use of several standard distributions and tail bounds that are defined here for completeness.

7.1 Basic Distributions

Definition 3 (Normal or Gaussian distribution). A random variable X has a normal distribution with mean μ and variance σ^2 if it's probability density function is given by

$$\frac{1}{\sqrt{2\pi\sigma^2}}e^{-(x-\mu)^2/(2\sigma^2)}$$

A standard Gaussian distribution is a Gaussian distribution with mean 0 and variance 1. The cumulative distribution function (cdf) of a standard normal distribution is denoted by $\Phi(\cdot)$ and the pdf is denoted by $\phi(\cdot)$.

Definition 4 (χ^2 -distribution). A random variable X has a χ^2 distribution with k degrees of freedom if it can be written as the sum of squares of k standard normal distributions, i.e. $X = \sum_{i=1}^{k} Z_i^2$ where each Z_i is a standard normal random variable.

Definition 5 (t-distribution). A random variable T_k has a t-distribution with k degrees of freedom, if it can be represented as follows: $T_k = Z/(\sqrt{Y/k})$ where Z is a standard normal distribution and Y is a χ^2 -distribution with k degrees of freedom. The density of a t-distribution is given by

$$\frac{\Gamma\left(\frac{k+1}{2}\right)}{\sqrt{k\pi}\Gamma\left(\frac{k}{2}\right)}\left(1+\frac{t^2}{k}\right)^{-\frac{k+1}{2}}$$

Definition 6 (Laplace distribution). A random variable X has a Laplace distribution with mean 0 and scale parameter b if it's density is given by

$$\frac{1}{2b}e^{-\frac{|x|}{b}}$$

7.2 Tail bounds

We make use of several tail bounds, some of which are well known and listed without proof and others are proved for completeness.

Proposition 7.1 (Chernoff Bounds). Let X_1, \ldots, X_n be independent random variables such that $X_i \in \{0, 1\}$, $\mu = \sum_i \mathbb{E}[X_i]$ and $X = \sum_i X_i$. Then for every $\delta \in (0, 1)$,

$$\mathbb{P}\left(X \le (1-\delta)\mu\right) \le \exp\left(-\frac{\delta^2}{2}\mu\right)$$
$$\mathbb{P}\left(X \ge (1+\delta)\mu\right) \le \exp\left(-\frac{\delta^2}{3}\mu\right).$$

Proposition 7.2 (Laplace tail bound). Let Z be a Laplace random variable with mean 0 and scale b then, for every t > 0,

$$\mathbb{P}(Z > t) = \frac{1}{2} \exp\left(-\frac{t}{b}\right) \text{ and}$$
$$\mathbb{P}(|Z| > t) = \exp\left(-\frac{t}{b}\right).$$

Proposition 7.3 (Gaussian tail bound). Let Z be a standard normal random variable with mean 0 and variance 1, then, for every t > 0, we have

$$\mathbb{P}\left(|Z| > t\right) \le 2\exp\left(-t^2/2\right).$$

Proposition 7.4 (χ^2 tail bound, Lemma 1 in Laurent and Massart (2000)). Let Y be a χ^2 random variable with n degrees of freedom, then for every x > 0 we have,

$$\mathbb{P}\left(Y \le n - 2\sqrt{nx}\right) \le \exp(-x).$$

Proposition 7.5 (A tail bound on t-distribution from Soms (1976)). Let T_n be a t-distribution with n degrees of freedom, then for every t > 0, we have,

$$\mathbb{P}\left(T_n \ge t\right) \le \frac{1}{t} \left(1 + \frac{t^2}{n}\right) f_{T_n}(t)$$

where $f_{T_n}(.)$ is the density of a t-distribution.

Proposition 7.6 (Relation between quantile and tail bound). Let X be a random variable with invertible cdf F(.) and let $q(\alpha) = F^{-1}(1-\alpha)$. If $\mathbb{P}(X \ge t) \le \alpha$, then $q(\alpha) \le t$.

Proof. By definition, $F(q(\alpha)) = 1 - \alpha$ and it is given that $F(t) \ge 1 - \alpha$. Since F(.) is a non-decreasing function, we have $t \ge q(\alpha)$.

Proposition 7.7 (Quantiles of a t-distribution). Let T_n be a t-distribution with n degrees of freedom and $t_{n,\alpha}$ be the $1 - \alpha$ quantile of T_n . Then,

1. If $t_{n,\alpha} > 1$, $t_{n,\alpha}^2 \le n \left(\sqrt{\frac{2\pi n}{n+1}} \left(\frac{1}{\alpha} \right)^{2/(n-1)} - 1 \right)$. 2. If $n \ge (64/9) \log(2/\alpha)$, then $t_{n,\alpha} \le \sqrt{8 \log(2/\alpha)}$.

Proof. In both the cases, the upper bound on the quantile follows from application of Proposition 7.6 and a tail bound on T_n . Thus, we only need to prove the tail bound on T_n in each case.

Part 1: We will prove the following tail bound:

$$\mathbb{P}\left(T_n \ge t\right) \le \frac{1}{\sqrt{2\pi}} \sqrt{\frac{n+1}{n}} \left(\frac{1}{t}\right) \left(1 + \frac{t^2}{n}\right)^{-(n-1)/2}$$

From Proposition 7.5, we have, $\mathbb{P}(T_n \ge t) \le \frac{1}{t} \left(1 + \frac{t^2}{n}\right) f_{T_n}(t)$ where $f_{T_n}(.)$ is the density of a t-distribution. Hence we have,

$$\mathbb{P}\left(T_n \ge t\right) \le \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \sqrt{\frac{1}{\pi n}} \cdot \frac{1}{t} \left(1 + \frac{t^2}{n}\right)^{-\frac{n-1}{2}}$$

Using the fact that for x > 0 and 0 < s < 1,

$$x^{1-s} < \frac{\Gamma(x+1)}{\Gamma(x+s)} \le (x+1)^{1-s}$$

and letting x = (n-1)/2 and s = 1/2, gives the desired result.

Part 2: We will prove the following tail bound: $\mathbb{P}\left(T_n \ge \sqrt{8\log(2/\alpha)}\right) \le \alpha$ Recall that $T_n = Z/(\sqrt{Y/n})$ where Z is a standard normal distribution and Y is a χ^2 -distribution with n degrees of freedom. Also, if $\{Z < t/2\}$ and $\{Y > n/4\}$, then $T_n < t$. Hence $T_n \ge t$ implies that either $Z \ge t/4$ or $Y \le n/4$. By a union bound, we have:

$$\mathbb{P}(T_n \ge t) \le \mathbb{P}(Z \ge t/2) + \mathbb{P}(Y \le n/4)$$
$$\le \exp(-t^2/8) + \exp(-9n/64),$$

where the last equation follows from the tail bounds on the Normal distribution in Proposition 7.3 and the χ^2 -distribution in Proposition 7.4. Finally, by setting $t = \sqrt{8 \log(2/\alpha)}$, and using the fact that $n \ge (64/9) \log(2/\alpha)$ gives the result.

Acknowledgments. We are grateful to the Harvard Privacy Tools differential privacy research group for illuminating and motivating discussions, particularly James Honaker, Gary King, Kobbi Nissim and Uri Stemmer. We would also like to thank Philip Leclerc, CDAR, Mathematical Statistician for carefully reading our paper and giving helpful feedback.

Bibliography

- Rina Foygel Barber and John C Duchi. Privacy and statistical risk: Formalisms and minimax bounds. arXiv preprint arXiv:1412.4451, 2014.
- Amos Beimel, Shiva Prasad Kasiviswanathan, and Kobbi Nissim. Bounds on the sample complexity for private learning and private data release. In *TCC*, volume 5978, pages 437–454. Springer, 2010.
- PJ Bickel et al. Minimax estimation of the mean of a normal distribution when the parameter space is restricted. The Annals of Statistics, 9(6):1301–1309, 1981.
- Avrim Blum, Cynthia Dwork, Frank McSherry, and Kobbi Nissim. Practical privacy: the sulq framework. In Proceedings of the twenty-fourth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems, pages 128–138. ACM, 2005.
- Ben W. Bolch. The teacher's corner: More on unbiased estimation of the standard deviation. *The American Statistician*, 22(3):27–27, 1968. doi: 10.1080/00031305.1968.10480476. URL http://dx.doi.org/10.1080/00031305.1968.10480476.
- Mark Bun, Kobbi Nissim, Uri Stemmer, and Salil Vadhan. Differentially private release and learning of threshold functions. In Proceedings of the 2015 IEEE 56th Annual Symposium on Foundations of Computer Science (FOCS), FOCS '15, pages 634–649, Washington, DC, USA, 2015. IEEE Computer Society. ISBN 978-1-4673-8191-8. doi: 10.1109/FOCS.2015.45. URL http://dx.doi.org/10.1109/FOCS.2015.45.
- Mark Bun, Kobbi Nissim, and Uri Stemmer. Simultaneous private learning of multiple concepts. In *Proceedings* of the 2016 ACM Conference on Innovations in Theoretical Computer Science, pages 369–380. ACM, 2016.
- Bryan Cai, Constantinos Daskalakis, and Gautam Kamath. Priv'it: Private and sample efficient identity testing. arXiv preprint arXiv:1703.10127, 2017.
- George Casella and William E Strawderman. Estimating a bounded normal mean. The Annals of Statistics, pages 870–878, 1981.
- Anirban DasGupta. Asymptotic theory of statistics and probability. Springer Science & Business Media, 2008.

- John Duchi, Martin J Wainwright, and Michael I Jordan. Local privacy and minimax bounds: Sharp rates for probability estimation. In Advances in Neural Information Processing Systems, pages 1529–1537, 2013a.
- John C Duchi, Michael I Jordan, and Martin J Wainwright. Local privacy and statistical minimax rates. In Foundations of Computer Science (FOCS), 2013 IEEE 54th Annual Symposium on, pages 429–438. IEEE, 2013b.
- Cynthia Dwork and Jing Lei. Differential privacy and robust statistics. In Proceedings of the forty-first annual ACM symposium on Theory of computing, pages 371–380. ACM, 2009.
- Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In *TCC*, pages 265–284. Springer, 2006.
- Steven N Evans, Ben B Hansen, and Philip B Stark. Minimax expected measure confidence sets for restricted location parameters. *Bernoulli*, pages 571–590, 2005.
- Stephen E. Fienberg, Alessandro Rinaldo, and Xiaolin Yang. Differential privacy and the risk-utility tradeoff for multi-dimensional contingency tables. In *Proceedings of the 2010 international conference on Privacy in statistical databases*, PSD'10, pages 187–199, Berlin, Heidelberg, 2010. Springer-Verlag. ISBN 3-642-15837-4, 978-3-642-15837-7. URL http://dl.acm.org/citation.cfm?id=1888848.1888869.
- Matthew Fredrikson, Eric Lantz, Somesh Jha, Simon Lin, David Page, and Thomas Ristenpart. Privacy in pharmacogenetics: An end-to-end case study of personalized warfarin dosing. In USENIX Security Symposium, pages 17–32, 2014.
- Marco Gaboardi, James Honaker, Gary King, Kobbi Nissim, Jonathan Ullman, and Salil Vadhan. Psi ({\Psi}): a private data sharing interface. arXiv preprint arXiv:1609.04340, 2016a.
- Marco Gaboardi, Ryan Rogers, and Salil Vadhan. Differentially private chi-squared hypothesis testing: Goodness of fit and independence testing. 2016b.
- Rob Hall, Alessandro Rinaldo, and Larry Wasserman. Random differential privacy. arXiv preprint arXiv:1112.2680, 2011.
- Rob Hall, Alessandro Rinaldo, and Larry Wasserman. Differential privacy for functions and functional data. Journal of Machine Learning Research, 14(Feb):703–727, 2013.
- Moritz Hardt and Kunal Talwar. On the geometry of differential privacy. In *Proceedings of the forty-second* ACM symposium on Theory of computing, pages 705–714. ACM, 2010.
- Michael Hay, Ashwin Machanavajjhala, Gerome Miklau, Yan Chen, and Dan Zhang. Principled evaluation of differentially private algorithms using dpbench. In Proceedings of the 2016 International Conference on Management of Data, pages 139–154. ACM, 2016.
- Robert V Hogg and Allen T Craig. Introduction to mathematical statistics. (5"" edition). Upper Saddle River, New Jersey: Prentice Hall, 1995.
- Vishesh Karwa and Aleksandra Slavković. Differentially private graphical degree sequences and synthetic graphs. In *Privacy in Statistical Databases*, pages 273–285. Springer, 2012.
- Vishesh Karwa and Aleksandra Slavković. Inference using noisy degrees: Differentially private β -model and synthetic graphs. The Annals of Statistics, 44(1):87–112, 2016.
- Vishesh Karwa, Aleksandra B Slavković, and Pavel Krivitsky. Differentially private exponential random graphs. In *International Conference on Privacy in Statistical Databases*, pages 143–155. Springer, 2014.

- Vishesh Karwa, Dan Kifer, and Aleksandra B Slavković. Private posterior distributions from variational approximations. arXiv preprint arXiv:1511.07896, 2015.
- Shiva Prasad Kasiviswanathan, Homin K Lee, Kobbi Nissim, Sofya Raskhodnikova, and Adam Smith. What can we learn privately? *SIAM Journal on Computing*, 40(3):793–826, 2011.
- Daniel Kifer and Ryan Rogers. A new class of private chi-square tests. arXiv preprint arXiv:1610.07662, 2016.
- Beatrice Laurent and Pascal Massart. Adaptive estimation of a quadratic functional by model selection. Annals of Statistics, pages 1302–1338, 2000.
- Erich L Lehmann and Joseph P Romano. *Testing statistical hypotheses*. Springer Science & Business Media, 2006.
- Eric Marchand and William E Strawderman. Estimation in restricted parameter spaces: A review. Lecture notes-monograph series, pages 21–44, 2004.
- Pascal Massart. The tight constant in the dvoretzky-kiefer-wolfowitz inequality. The Annals of Probability, pages 1269–1283, 1990.
- John W Pratt et al. Shorter confidence intervals for the mean of a normal distribution with known variance. The Annals of Mathematical Statistics, 34(2):574–586, 1963.
- Chad M Schafer and Philip B Stark. Constructing confidence regions of optimal expected size. Journal of the American Statistical Association, 104(487):1080–1089, 2009.
- Or Sheffet. Differentially private ordinary least squares. In International Conference on Machine Learning, pages 3105–3114, 2017.
- Adam Smith. Differential privacy and the secrecy of the sample. Blog: Oddly Shaped Pegs, 2009.
- Adam Smith. Privacy-preserving statistical estimation with optimal convergence rates. In Proceedings of the forty-third annual ACM symposium on Theory of computing, pages 813–822. ACM, 2011.
- Eftychia Solea. Differentially private hypothesis testing for normal random variables. 2014.
- Andrew P Soms. An asymptotic expansion for the tail area of the t-distribution. Journal of the American Statistical Association, 71(355):728–730, 1976.
- Thomas Steinke and Jonathan Ullman. Between pure and approximate differential privacy. arXiv preprint arXiv:1501.06095, 2015.
- Salil Vadhan. The complexity of differential privacy. Work. Pap., Cent. Res. Comput. Soc., Harvard Univ. http://privacytools. seas. harvard. edu/publications/complexity-differential-privacy, 2016.
- Salil Vadhan. The complexity of differential privacy. In *Tutorials on the Foundations of Cryptography*, pages 347–450. Springer, 2017.
- Duy Vu and Aleksandra Slavkovic. Differential privacy for clinical trial data: Preliminary evaluations. In *Data Mining Workshops, 2009. ICDMW'09. IEEE International Conference on*, pages 138–143. IEEE, 2009.
- Yue Wang, Jaewoo Lee, and Daniel Kifer. Differentially private hypothesis testing, revisited. arXiv preprint arXiv:1511.03376, 2015.
- Larry Wasserman. Minimaxity, statistical thinking and differential privacy. Journal of Privacy and Confidentiality, 4(1):3, 2012.

- Larry Wasserman and Shuheng Zhou. A statistical framework for differential privacy. J. Amer. Statist. Assoc., 105(489):375–389, 2010. ISSN 0162-1459. doi: 10.1198/jasa.2009.tm08651. URL http://dx.doi.org/10.1198/jasa.2009.tm08651.
- Oliver Williams and Frank McSherry. Probabilistic inference and differential privacy. In Advances in Neural Information Processing Systems, pages 2451–2459, 2010.