Journal of Privacy and Confidentiality
Vol. 8 (1) 2018
Issue in honor of S. Fienberg
Submitted Dec. 20, 2018
Published December 2018

STATISTICAL APPROXIMATING DISTRIBUTIONS UNDER DIFFERENTIAL PRIVACY

YUE WANG, DANIEL KIFER, JAEWOO LEE, AND VISHESH KARWA

Department of Computer Science and Engineering; The Pennsylvania State University
e-mail address: yuw140@cse.psu.edu

Department of Computer Science and Engineering; The Pennsylvania State University
e-mail address: dkifer@cse.psu.edu

Department of Computer Science; University of Georgia
e-mail address: jwlee@cs.uga.edu

Department of Statistical Science; Fox School of Business; Temple University
e-mail address: vishesh@temple.edu

Abstract. Statistics computed from data are viewed as random variables. When they are used for tasks like hypothesis testing and confidence intervals, their true sampling distributions are often replaced by approximating distributions that are easier to work with (for example, the Gaussian, which results from using approximations justified by the Central Limit Theorem). When data are perturbed by differential privacy, the approximating distributions also need to be modified. Prior work provided various competing methods for creating such approximating distributions with little formal justification beyond the fact that they worked well empirically.

In this paper, we study the question of how to generate statistical approximating distributions for differentially private statistics, provide finite sample guarantees for the quality of the approximations. We also provide illustrative examples, mostly focusing on chi-squared testing.

1. Introduction

The increasing use of formal privacy methods, such as differential privacy [Dwork et al., 2006, Bun and Steinke, 2016], to create and release *sanitized data* (instead of the original sensitive data) has raised the question of how much to trust analyses derived from the sanitized data. These concerns have led to a line of research [Gaboardi et al., 2016, Karwa, 2016, D'Orazio et al., 2015, Sheffet, 2015, Rogers and Kifer, 2017, Uhler et al., 2013, Yu et al., 2014, Smith, 2011, Vu and Slavkovic, 2009, Dwork et al., 2015, Solea, 2014, Karwa and Slavkovic, 2016, Johnson and Shmatikov, 2013] on how to properly incorporate differential

Key words and phrases: Differential Privacy, Approximating Distributions, Finite Sample Guarantees.
Supported by NSF grants 1702760 and 1228669.

privacy in statistics to account for errors due to (1) sampling of the original data from a data-generating distribution, and (2) additional noise due to privacy protection schemes.

The typical process we consider is to start with a sensitive dataset D_n of size n, compute a noisy privacy-preserving statistic $\tilde{S}(D_n)$ and then approximate the sampling distribution of $\tilde{S}(D_n)$. This approximating distribution can then be used for various purposes, such as creating confidence intervals, testing hypotheses, etc.[1] Previous approaches in this direction used ad-hoc techniques that were specifically customized to how $\tilde{S}(D_n)$ was computed or used simplifying assumptions, such as truly Gaussian distributed data.

In the non-private case, a simple example of the use of approximating distributions is in the generation of confidence intervals around the mean of n i.i.d. random variables.

Example 1.1. *Let $D_n = \{x_1, x_2, \ldots, x_n\}$ be a sequence of numbers that are bounded between 0 and 1 and are generated independently from some distribution F with known variance σ^2. The mean μ of F can be estimated as $\hat{\mu}_n \equiv \frac{1}{n}\sum x_i$. To get a confidence interval around $\hat{\mu}_n$ that is likely to contain the true mean, the standard approach is to use the Central Limit Theorem to conclude that $\sqrt{n}\frac{\hat{\mu}_n - \mu}{\sigma}$ converges in distribution to a standard Gaussian $N(0,1)$ as $n \to \infty$. Next approximate $\sqrt{n}\frac{\hat{\mu}_n - \mu}{\sigma}$ as a Gaussian and find an interval $[-\alpha, \alpha]$ that contains 95% of the probability mass of a standard Gaussian. Thus $P(\sqrt{n}(\hat{\mu}_n - \mu)/\sigma \in [-\alpha, \alpha]) \approx 0.95$. Simple algebra then shows that $P(\mu \in [\hat{\mu}_n - \sigma\alpha/\sqrt{n}, \hat{\mu}_n + \sigma\alpha/\sqrt{n}]) \approx 0.95$ and this gives an (approximate) 95% confidence interval for μ.*

In Example 1.1, the Gaussian was used as an approximation to the true distribution of the average of n observations in the data. In practice, this approximation is often fairly accurate (e.g., see Figure 1).

In the privacy-preserving case, we would not have direct access to $x_1, x_2, \ldots, x_n$ or even to $\sum_i x_i$ but we could have access to a privacy-preserving noisy sum, such as $Y_\epsilon + \sum_i x_i$, where Y_ϵ is a Laplace random variable with scale $1/\epsilon$ and variance $2/\epsilon^2$. We can estimate the true mean μ of the distribution F as $\tilde{\mu}_n = \frac{1}{n}(Y_\epsilon + \sum_i x_i)$ – this is our privacy-preserving statistic $\tilde{S}(D_n)$. If we can find an approximating distribution for $\sqrt{n}(\tilde{\mu}_n - \mu)/\sigma$, then we could proceed as in Example 1.1 to generate a confidence interval for the unknown true

I was fortunate enough to meet Steve at a variety of privacy workshops over the years. More than anyone I have met, he encouraged early-career researchers and showed interest in their work. Part of the reason was Steve's enthusiasm for statistical disclosure control, which comes across in every conversation. Once, he (half-jokingly) told me that everyone should be limited to lifetime total of 20 published pages, so that people focus their energies on deep progress in the problems they care about. Late in his career, Steve also mentioned that he started telling prospective students that he doesn't need to publish anymore. Of course, he continued his research in privacy — not because he needed to, but because he wanted to. One of our shared common interests was in developing techniques for proper statistical analysis of differentially private data, and overcoming the associated computational difficulties. I am grateful for his support and encouragement — to a first-year faculty, it means a lot when a legend tells you he had read all of your papers, and I will miss his enthusiasm.

Dan Kifer

DOI: 10.29012/jpc.705

[1] We note that other approaches are also possible, such as using MCMC (e.g., [Charest, 2011]) or, in certain cases, computing the exact distribution (e.g., [Solea, 2014]).

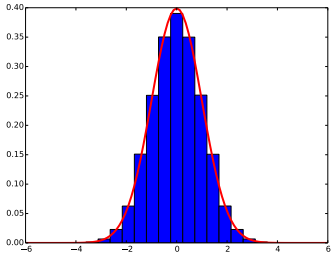

Figure 1: Nonprivate Case: Distribution of $\sqrt{n}(\hat{\mu}-\mu)/\sigma$ (as a histogram) for data generated by Bernoulli($n = 4,000$, $\mu = 0.5$) vs. $N(0,1)$.

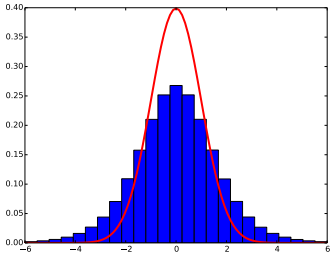

Figure 2: Private Case: Distribution of $\sqrt{n}(\tilde{\mu}-\mu)/\sigma \equiv \frac{Y_\epsilon+\sum_i(x_i-\mu)}{\sqrt{n\sigma^2}}$ (with $\epsilon = 0.033$) vs. approximating distribution from Option 1 in Section 1.

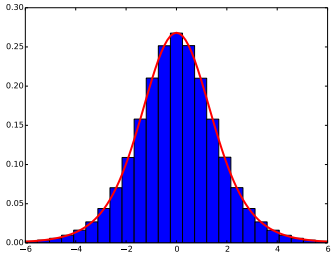

Figure 3: Private Case: Distribution of $\sqrt{n}(\tilde{\mu}-\mu)/\sigma$ (with $\epsilon = 0.033$) vs. the approximating distribution from Options 2 and 3.

mean μ. The challenge is finding such an approximating distribution. There are several natural possibilities:

(Option 1) Approximating $\sqrt{n}(\tilde{\mu}_n-\mu)/\sigma$ by taking the limit as $n \to \infty$ (e.g., Johnson and Shmatikov [2013]). This results in a Gaussian approximation and is often inaccurate (see Figure 2).

(Option 2) Rewriting $\sqrt{n}(\tilde{\mu}_n-\mu)/\sigma$ as $\frac{Y_\epsilon}{\sqrt{n\sigma^2}} + \frac{\sum(x_i-\mu)}{\sqrt{n\sigma^2}}$ and approximating the second term as a Gaussian. The overall approximation is then the sum of a $N(0,1)$ random variable and an independent Laplace random variable with scale $1/(\epsilon\sqrt{n\sigma^2})$.

(Option 3) Recalling that ϵ is the privacy parameter of the noise in $\tilde{\mu}_n$, take the limit of $\tilde{\mu}_n$ as $n \to \infty, \epsilon \to 0$ in such a way that $\epsilon\sqrt{n}$ remains constant (e.g., Rogers and Kifer [2017], Gaboardi et al. [2016]). This approach results in the same approximating distribution to the mean as in Option 2 [Rogers and Kifer, 2017, Gaboardi et al., 2016].

The quality of the approximations in Options 2 and 3 are shown in Figure 3. The main justification for Options 2 and 3 is that they empirically work well in the specific problem they were applied to (chi-squared testing with input perturbation [Gaboardi et al., 2016, Rogers and Kifer, 2017]); however there were no formal justifications for why these would be good approximations to use.

The goal of this paper is to provide a methodology for creating approximating distributions in more general settings and to study finite sample guarantees for the quality of the approximations.

One approach is to use techniques like the Berry-Esseen theorem [Berry, 1941] or Stein's method [Stein, 1972] to upper bound the rate of convergence of the sampling distribution of a statistic to its approximating distribution. However, we take a different approach to provide *relative* guarantees that compare the convergence of a privacy-preserving statistic to the rate of convergence of the non-private statistic.

Specifically, let $S(D_n)$ be a non-private statistic whose sampling distribution F is approximated by some distribution G. Let $\tilde{S}(D_n)$ be a privacy preserving version of this statistic and let $\tilde{G}$ be an approximation of its sampling distribution $\tilde{F}$. We are looking to provide guarantees of the form $d(\tilde{F}, \tilde{G}) \leq \alpha d(F, G)$, where d is a distance measure between distributions (e.g., the Kolmogorov-Smirnov distance). These relative guarantees allow us to directly take advantage of non-private convergence results (e.g., the Berry-Esseen theorem) to bound the distance between $\tilde{G}$ and $\tilde{F}$. Furthermore, in cases where the non-private statistic converges faster than the theoretical results imply (i.e. the constants in the theoretical results may be too loose), relative guarantees would state that the privacy-preserving statistics also converge quickly to their approximating distributions.

As a simple example, our techniques show that the approximation used in Options 2 and 3 (Figure 3) is just as good (in terms of Kolmogorov-Smirnov or Wasserstein distance) as the Gaussian approximation used by the Central Limit Theorem in the non-private case (Figure 1). In summary, our contributions are as follows:

- We provide a generic recipe for creating approximations to the sampling distributions of privacy-preserving statistics. While conceptually simple, we note that existing works either use poor approximations (e.g., [Johnson and Shmatikov, 2013]) or propose methods specific to a given application (e.g., [Uhler et al., 2013, Yu et al., 2014, Gaboardi et al., 2016, Rogers and Kifer, 2017]).
- We provide accuracy guarantees for these approximating distributions. The accuracy guarantees are relative to the guarantees that would hold in the non-private case and depend on the mechanism M that outputs privacy-preserving statistics.
- In many cases, the noisy statistic $M(S(D))$ undergoes further processing (for example, when using the value of the noisy statistic and its approximating distribution to create confidence intervals or to do a chi-squared test). We provide positive and negative results on how such post-processing can affect approximation accuracy. For some statistical analyses, we cannot provide an accuracy guarantee, so one of the purposes of this paper is also to point out possible research directions in non-private rates of convergence that can help address the negative results.

In Section 2, we review related work. In Section 3, we review background material in statistics and privacy. We discuss desiderata for approaches for creating approximating distributions in Section 4. We present a general recipe for approximating distributions in Section 5, along with relative accuracy guarantees (Section 5.1) and degradation results under post-processing (Section 5.2). In Section 6, we show how the proposed recipe can be applied to a variety of applications, what guarantees can be provided, and what kinds of additional non-private results are needed.

2. Related Work

The ability to use differential privacy to protect data has raised the question of how useful would the data be in statistical applications [Wasserman and Zhou, 2010, Fienberg et al., 2010, Dwork and Lei, 2009, Vu and Slavkovic, 2009, Smith, 2011, Chaudhuri and Hsu, 2012]. Early experimental evaluations were not encouraging [Fienberg et al., 2010] but since then there have been numerous efforts at designing statistical tests and confidence intervals with improved performance [Johnson and Shmatikov, 2013, Uhler et al., 2013, Solea, 2014, Yu

et al., 2014, Sheffet, 2015, Gaboardi et al., 2016, Rogers and Kifer, 2017, Karwa, 2016, Karwa and Slavkovic, 2016, D'Orazio et al., 2015, Cai et al., 2017].

The most common applications of statistics involve *estimation* (e.g., estimating the mean of a population from a sample), *confidence intervals*, and *hypothesis testing*. The latter two tasks involve computing a statistic S_n over a dataset of size n and often (but not always) approximating the sampling distribution of S_n and then using the approximating distribution to create confidence intervals or to provide evidence against statistical hypotheses. In the case of differential privacy, such decisions must be made from a noisy statistic $\tilde{S}_n$ rather than the true statistic S_n. Then most, but not all (e.g., Cai et al. [2017], Acharya et al. [2017], Karwa and Vadhan [2018], Awan and Slavković [2018], Charest [2011]) approaches base their confidence intervals or hypothesis test on an approximation to the true distribution of $\tilde{S}_n$.

There are various approaches to generating such approximating distributions. The most direct approach is to compute the limiting distribution of $\tilde{S}_n$ as $n \to \infty$ and use that as the approximation [Johnson and Shmatikov, 2013, Smith, 2011, Karwa and Slavkovic, 2016]. The resulting approximations are often inaccurate unless n is so large that the variance of the privacy noise is insignificant compared to the variance in the data. Other approaches make specific assumptions on the data (such as normality) to avoid using approximation techniques like the CLT [Karwa, 2016, Solea, 2014, Sheffet, 2015, Chen et al., 2016].

When dealing with non-normal data, one common approach is to perform direct replacements, such as finding terms like $\sum_i (x_i - \mu)/\sqrt{n\sigma^2}$ and replacing them with Gaussian random variables [D'Orazio et al., 2015, Gaboardi et al., 2016]. Other approaches involve setting up a limiting process [Rogers and Kifer, 2017, Uhler et al., 2013, Yu et al., 2014] and using the limiting distribution as the finite sample approximation. An advantage of this approach is that it can be used with other limit theorems, like Slutsky's theorem [Rogers and Kifer, 2017, Ferguson, 1996]. In all cases, the methods are narrow – they are designed for the specific privacy mechanisms that generate the privacy-preserving statistics and do not provide approximation guarantees.

Recent work by Rinott et al. [2018] considered several cases where the likelihood function of noisy tables is tractable and parameter estimates can be obtained by optimization procedures. This likelihood function was, for example, used in testing the null hypothesis of independence between rows and columns of a table. They noted that the sampling distribution empirically matched the asymptotic distribution that one gets as the sample size $n \to \infty$. However, in general it is not always possible to optimize the likelihood function efficiently (Rinott et al. [2018] noted two special cases where it is: Poisson data with truncated noise and Binomial data with truncated noise – it is worth noting that truncated noise is used to ensure the computational complexity does not scale with the sample size, but forces the use of approximate differential privacy). Furthermore, when the privacy noise does not depend on the sample size, letting $n \to \infty$ is generally not a reliable way of finding a good approximation to the sampling distribution [Uhler et al., 2013, Gaboardi et al., 2016] (since the added noise always drops out of the resulting distribution). We conjecture that when our recipe is applied to the techniques of Rinott et al. [2018], the approximating distribution will coincide with the distribution they obtained by letting $n \to \infty$.

Charest [2011] proposed some of the earliest Bayesian techniques to analyze differentially private data by taking advantage of conjugate priors in the differential privacy mechanism and using MCMC. In general, Bayesian techniques for the analysis of differentially private data are computationally expensive and in many cases would need to resort to approximations for efficiency. One exception, which we discuss later in this section, is the case where each

record is individually perturbed. However, such approaches lead to more perturbation and a greater loss of power since the privacy noise scales with the sample size.

Smith [Smith, 2011] and Wasserman and Zhou [Wasserman and Zhou, 2010] focus on finite sample guarantees about the distribution of data protected by differential privacy, while Chaudhuri and Hsu [2012] study the convergence of point estimators derived from empirical distribution functions. In particular, Smith [Smith, 2011] studies the convergence of differentially private statistics to the Gaussian distribution; however, valid statistical tests can be performed even if the noisy data are still far from Gaussian [Gaboardi et al., 2016, Rogers and Kifer, 2017] (as long as a different approximating distribution is used). The results of Wasserman and Zhou [Wasserman and Zhou, 2010] can be viewed as the rate at which sample size overpowers fixed privacy noise when estimating the true data-generating distribution. Again, valid statistical inference can be performed over differentially private data even when the data do not overpower the privacy noise (for example when the variance of the privacy noise is a constant fraction of the variance due to data sampling) [Gaboardi et al., 2016, Rogers and Kifer, 2017, Karwa, 2016].

An important line of work focuses on record-level perturbations. That is, rather than adding noise to sufficient statistics, each record is individually perturbed. This approach generally leads to a higher loss of power since the amount of perturbation scales with the size of the data. However, it can be used to provide stronger notions of local privacy where even the data collector is not trusted [Duchi et al., 2018]. Duchi et al. [2018] study mechanisms with optimal minimax rates under this setting. Since the noise scales with the sample size, there is no issue with using asymptotics where $n \to \infty$ (since it doesn't cause the noise to drop out of the equation). However, in this setting it is common to use Bayesian techniques (e.g., MCMC) to fit models from the perturbed data [Goldstein and Shlomo, 2018, Polettini and Arima, 2015] or EM for point estimates and rely on standard asymptotics for their properties (such as unbiasedness) [Woo and Slavkovic, 2015].

3. Notation and Background

In this paper, we use the notation $\text{Lap}(b)$ for the Laplace distribution with density function $f(x) = \frac{1}{2b}e^{-|x|/b}$; it has mean 0 and standard deviation $b\sqrt{2}$. We use the notation $\text{Lap}_k(b)$ for a vector of k independent $\text{Lap}(b)$ random variables.

3.1. Concepts in Privacy.

The main privacy definitions we will be considering are ϵ-differential privacy and ρ-zcdp. This allows us provide examples of our framework using different kinds of privacy noise.

Definition 3.1 (Dwork et al. [2006])**.** *A randomized algorithm M satisfies ϵ-differential privacy if for all possible outputs ω and pairs of datasets D_1, D_2 differing on the value of one record, $P(M(D_1) = \omega) \leq e^{\epsilon} P(M(D_2) = \omega)$, where the randomness comes from the algorithm M.*

Differential privacy controls how much influence one person's data could have on the output of a randomized algorithm [Dwork et al., 2006, Kifer and Machanavajjhala, 2014, Machanavajjhala and Kifer, 2015]. A relaxation of differential privacy, known as ρ-zero mean concentrated differential privacy (ρ-zcdp) [Bun and Steinke, 2016], is defined as:

Definition 3.2 (ρ-zcdp [Bun and Steinke, 2016])**.** *A randomized algorithm M satisfies ρ-zero-mean concentrated differential privacy (ρ-zcdp) if for each pair of datasets D_1, D_2 that differ on the value of one record and all $\alpha > 1$,*

$$\sum_{\omega \in \text{range}(M)} \left(\frac{P(M(D_1) = \omega)}{P(M(D_2) = \omega)}\right)^{\alpha-1} P(M(D_1) = \omega) \leq e^{(\alpha-1)\rho\alpha}.$$

Informally, it takes the *privacy loss random variable*, defined as $\frac{P(M(D_1)=\omega)}{P(M(D_2)=\omega)}$, and instead of bounding it absolutely by e^ϵ (as in differential privacy), it allows the privacy loss to be a random variable with constraints on its moments.

Some commonly used algorithms for both privacy definitions are based on the concept of sensitivity. The L_p *sensitivity* of a function f, denoted by $\text{Sen}_p(f)$, is the largest change in f resulting from a change in one of the input records. More precisely,

$$\text{Sen}_p(f) = \max_{D_1, D_2} ||f(D_1) - f(D_2)||_p$$

where the max is over all pairs of datasets that differ on one record.

A very simple algorithm for enforcing ϵ-differential privacy is known as the *Laplace Mechanism* [Dwork et al., 2006].

Lemma 3.3 (Laplace Mechanism [Dwork et al., 2006])**.** *Given a function f, and a database D, and a differential privacy parameter ϵ, the Laplace Mechanism, which releases the noisy answer $f(D) + Lap(Sen_1(f)/\epsilon)$, satisfies ϵ-differential privacy.*[2]

On the other hand, for ρ-concentrated differential privacy, we can use the Gaussian mechanism [Bun and Steinke, 2016].

Lemma 3.4 (Gaussian Mechanism [Bun and Steinke, 2016])**.** *Given a function f, a database D, and a zcdp parameter ρ, the Gaussian Mechanism, which releases the noisy answer $f(D) + N(\vec{0}, \Sigma = \frac{Sen_2^2(f)}{2\rho}\mathbf{I})$, satisfies ρ-zcdp.*

Even though ρ-concentrated differential privacy is a relaxation of ϵ-differential privacy, the two definitions are often compared by setting $\rho = \frac{\epsilon^2}{2}$ as this setting guarantees that an ϵ-differential privacy algorithm satisfies ρ-zcdp [Bun and Steinke, 2016].

3.2. **Other Notation.** For a random variable X, we let F_X denote its cumulative distribution function (cdf), so that $F_X(t) = P(X \leq t)$. We denote multidimensional random variables in bold (e.g., $\mathbf{X}$) and multidimensional scalars like $\vec{t}$. The multidimensional cdf is $F_{\mathbf{X}}(\vec{t}) = P(\mathbf{X} \preceq \vec{t})$, where $\preceq$ is component-wise inequality.

One way to measure distance between distributions is a variant of the Kolmogorov-Smirnov distance:

Definition 3.5 ($KS(\mathcal{L})$ Distance)**.** *Let $\mathbf{X}$ and $\mathbf{Y}$ be d-dimensional random variables with cumulative distribution functions $F_{\mathbf{X}}$ and $F_{\mathbf{Y}}$. The $KS(\mathcal{L})$ distance $d_{KS(\mathcal{L})}$ is defined as: $d_{KS(\mathcal{L})}(\mathbf{X}, \mathbf{Y}) = \sup_{\vec{t} \in \mathcal{L}} |F_{\mathbf{X}}(\vec{t}) - F_{\mathbf{Y}}(\vec{t})|$. When $\mathcal{L} = \mathbb{R}^d$, we simply write it as d_{KS}.*

[2] If f is vector valued, the noise is a vector of independent Laplace random variables.

The set $\mathcal{L}$ is used to exclude certain points, such as points of discontinuity, which are frequently omitted in statistics when measuring convergence to distributions [Ferguson, 1996]. We also use the total variation distance d_V:

$$d_V(F_{\mathbf{X}}, F_{\mathbf{Y}}) = \sup |P(F_{\mathbf{X}} \in S) - P(F_{\mathbf{Y}} \in S)|.$$

When $\mathbf{X}$ and $\mathbf{Y}$ are both discrete or both have densities, d_V is half the L_1 distance between $\mathbf{X}$ and $\mathbf{Y}$.

We say $\mathbf{X}_1, \mathbf{X}_2, \ldots$ converge in distribution to $\mathbf{Y}$ if $F_{\mathbf{X}_n}(\vec{t}) \to F_{\mathbf{Y}}(\vec{t})$ at all points $\vec{t}$ at which $F_{\mathbf{Y}}$ is continuous.

Another way to measure distance between distributions is the Wasserstein metric. We use its dual representation [Edwards, 2011]:

Definition 3.6 (Wasserstein distance)**.** *Let μ_1 and μ_2 be two Borel probability measures over $\mathbb{R}^d$ such that $E_{\mathbf{X}\sim\mu_1}||\mathbf{X}||_2 < \infty$ and $E_{\mathbf{Y}\sim\mu_2}||\mathbf{Y}||_2 < \infty$. Let Ω be the set of all real-valued 1-Lipschitz continuous functions on $\mathbb{R}^d$. Then $d_W(\mu_1, \mu_2) = \sup_{f\in\Omega} E_{\mathbf{X}\sim\mu_1} f(\mathbf{X}) - E_{\mathbf{Y}\sim\mu_2} f(\mathbf{Y})$. When $F_{\mathbf{X}}$ and $F_{\mathbf{Y}}$ are the corresponding cdfs of μ_1 and μ_2, we also write $d_W(F_{\mathbf{X}}, F_{\mathbf{Y}})$ as $d_W(\mathbf{X}, \mathbf{Y})$.*

4. Desiderata for Approximating Distributions

In this section, we propose desiderata for methods that generate approximating distributions of differentially private statistics: they should be general and compatible with existing limit theorems.

4.1. **Generality.** There have been several approaches to generating asymptotic approximations of differentially private statistics. To compare them, consider the following example of a chi-squared goodness-of-fit test, which tests if data did not come from a Multinomial$(n_0, \vec{\theta})$ distribution. The null hypothesis is that the data *did* come from this distribution. The data points are k-dimensional vectors $\vec{x}_1, \vec{x}_2, \cdots, \vec{x}_{n_0}$ that are one-hot coded (in each vector $\vec{x}_i$, one component is 1 and the rest are 0 – this corresponds to the item that is sampled from a Multinomial). Let $\mathbf{S}_{n_0} = \sum_{i=1}^{n_0} \vec{x}_i$, which is a Multinomial random variable. The null hypothesis is that $\mathbf{S}_{n_0}$ was sampled from a Multinomial$(n_0, \vec{\theta})$ distribution with a specific value of the parameter $\vec{\theta}$. In the non-private case, one would commonly use the test statistic $\phi_{n_0}(\mathbf{S}_{n_0}) = \sum_{j=1}^{k} \frac{(\mathbf{S}_{n_0}[j] - n_0\vec{\theta}[j])^2}{n_0\vec{\theta}[j]}$ and approximate its sampling distribution under the null hypothesis (i.e. the assumption that $\vec{\theta}$ is the true parameter) by calculating the distribution of $\lim_{n\to\infty} \phi_n(S_n)$. Let F be the cumulative probability function of this limit. The quantity $1 - F(\phi_{n_0}(S_{n_0}))$ is therefore an estimate, under the null hypothesis, of seeing a value of the test statistic as extreme as what was observed on the data at hand. This quantity is called the *p-value* and if it is small, then this data is unlikely to be a sample from the Multinomial$(n_0, \vec{\theta})$ distribution.

For the sake of illustration, consider three different ways of making this statistic satisfy differential privacy or ρ-zcdp:[3]

[3]Please note that we are not proposing any of those to be used as a test, we are merely proposing three different situations to evaluate the suitability of previously proposed methods for generating approximating distributions.

Case 1: **[Input Perturbation]**. Set $\tilde{\mathbf{S}}_{n_0} = \mathbf{S}_{n_0} + Lap_k(b_1/\epsilon)$ (that is, add k independent, appropriately scaled Laplace random variables to $\mathbf{S}_{n_0}$ – for our purposes here, the specific value of b_1 is not relevant) and use the test statistic $\sum_{j=1}^{k} \frac{(\tilde{\mathbf{S}}_{n_0}[j]-n_0\vec{\theta}[j])^2}{n_0\vec{\theta}[j]}$. For ρ-zcdp with $\rho = \frac{\epsilon^2}{2}$ one can use $N(0, b_1^2/\epsilon^2\mathbf{I}_k)$ noise instead of Laplace.

Case 2: **[Output Perturbation]**. Add noise directly to the test statistic: $\sum_{j=1}^{k} \frac{(\mathbf{S}_{n_0}[j]-n_0\vec{\theta}[j])^2}{n_0\vec{\theta}[j]} + Lap(b_2/\epsilon)$. For ρ-zcdp ($\rho = \frac{1}{2}\epsilon^2$), use $N(0, b_2^2/\epsilon^2)$ noise instead.

Case 3: **[Hybrid].** Set $\tilde{\mathbf{S}}_{n_0} = \mathbf{S}_{n_0} + Lap_k(b_3/\epsilon)$ and use the test statistic $\sum_{j=1}^{k} \frac{(\tilde{\mathbf{S}}_{n_0}[j]-n_0\vec{\theta}[j])^2}{n_0\vec{\theta}[j]} + Lap(b_4/\epsilon)$. Again, for ρ-zcdp with $\rho = \frac{1}{2}\epsilon^2$, replace Lap$(b_3/\epsilon)$ and Lap(b_4/ϵ) with $N(0, b_3^2/\epsilon^2)$ and $N(0, b_4^2/\epsilon^2)$, respectively.

If we want to generate an approximating asymptotic distribution, there are two prior proposals: compute the distribution of the test statistic as $n \to \infty$ [Johnson and Shmatikov, 2013, Uhler et al., 2013] (Approach 1) or let $n \to \infty$ while keeping $\epsilon\sqrt{n}$ constant [Gaboardi et al., 2016, Rogers and Kifer, 2017] (Approach 2). Although the resulting distributions are complicated, their key properties can be determined:

(1) In Case 1, under the null hypothesis, Approach 1 yields a chi-squared distribution with $k-1$ degrees of freedom regardless of the variance of the added Gaussian or Laplace noise. Prior work [Gaboardi et al., 2016] showed that Approach 1 produces an approximating distribution with a much smaller first moment than the true sampling distribution and noted that Approach 2 is much more accurate empirically.
(2) In Case 2, under the null hypothesis, Approach 1 results in the sum of a chi-squared random variable (with $k-1$ degrees of freedom) and a Laplace(b_2/ϵ) (or Gaussian, depending on the added noise) random variable. This approximation is intuitive and has the same first moment as the sampling distribution.[4] Meanwhile, Approach 2 diverges since the noise scale approaches infinity ($n \to \infty$ and $n\sqrt{\epsilon} \to$constant imply that $\epsilon \to 0$).
(3) In Case 3, under the null hypothesis, the same arguments as in Gaboardi et al. [2016] show that Approach 1 again results in a distribution that underestimates the first moment while Approach 2 again diverges.

These examples show that no single asymptotic regime will provide good approximation to the sampling distributions of the statistics – even the first moment of the approximation can be significantly different from that of the sampling distribution.

In particular, the "right" limit to use depends on the privacy enforcement mechanism M (e.g., input perturbation or output perturbation). Thus, the choice of limit must depend on the mechanism M.

Please note that in this specific example, the sampling distribution of the test statistic under the null hypothesis can be exactly sampled from (e.g., sample a dataset from Multinomial$(n_0, \vec{\theta})$ and use it to compute the test statistic) and so the p-value $1-F(\phi_{n_0}(S_{n_0}))$ can be estimated using Monte Carlo techniques. In general this will not always be the case, hence the need for approximating distributions.

[4]To see the result about equality of the first moment, replace $(S_{n_0} - n_0\vec{\theta})/\sqrt{n}$ by a Gaussian having the same first and second moments. Standard results in statistics [Ferguson, 1996] show that the result has a chi-squared distribution with $k-1$ degrees of freedom. Meanwhile the first moment is unchanged because it only depends on the first and second moments of $(S_{n_0} - n_0\vec{\theta})/\sqrt{n}$.

4.2. **Compatibility with Limit Theorems.** Setting up limits and using well-defined asymptotic regimes is not the only way of creating approximating distributions. For example, in all of the cases in Section 4.1, another method is to rewrite the test statistic as a joint function of (1) the quantity $(\mathbf{S}_{n_0} - n_0\vec{\theta})/\sqrt{n_0}$ and (2) a Laplace (or Gaussian) random variable. Then one could substitute $(\mathbf{S}_{n_0} - n_0\vec{\theta})/\sqrt{n_0}$ with its central limit approximation $N(\vec{0}, \Sigma)$.

For example, in Case 1 of Section 4.1, one could rewrite the test statistic as follows. Let $\vec{u} = [u_1, u_2, \ldots, u_k]$ be a vector where $u_i = (\mathbf{S}_{n_0}[i] - n\vec{\theta}[i])/\sqrt{n_0}$ for all i and let $\vec{V} = [v_1, \ldots, v_k]$ be a vector of independent Laplace(b_1/ϵ) random variables. The test statistic can then be written as $\sum_{i=1}^{k}(u_i + v_i/\sqrt{n_0})^2/\vec{\theta}[i]$. Since the Central Limit Theorem approximates $[u_1, \ldots, u_k]$ by a multivariate Gaussian, the ad-hoc substitution approach simply replaces $\vec{u} = [u_1, \ldots, u_k]$ by a vector of Gaussians. The result is a generalized chi-squared distribution [Gaboardi et al., 2016].

However, ad-hoc substitutions may not always be applicable, especially in cases where multiple limit theorems are used to derive approximating distributions in the non-private setting (in general, one must be careful about combining mathematical theorems and ad-hoc substitutions in the same result).

One example is the chi-squared test of independence. Consider an $r \times c$ table of counts T_{n_0} with n_0 records over two variables R with r possible values and C with c possible values, so that $T_{n_0}[i, j]$ is the number of records in which $R = i$ and $C = j$. Such a table is modeled as a sample from a Multinomial(n_0, θ) distribution where $\theta[i, j]$ is the probability of a record with $R = i$ and $C = j$. Thus $\sum_i \sum_j T[i, j] = n_0$. Let us use the notation $T[\bullet, j] = \sum_i T[i, j]$ and $T[i, \bullet] = \sum_j T[i, j]$ to represent column and row marginals, respectively. A typical question to ask is whether the rows and columns of T are not independent (independence test) [Gaboardi et al., 2016, Rogers and Kifer, 2017, Uhler et al., 2013, Yu et al., 2014, Wang et al., 2015]. The non-private chi-squared statistic for this independence test is computed as $\phi_{n_0}(T_{n_0}) = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(T[i,j]-E[i,j])^2}{E[i,j]}$ where $E[i, j] = T[i, \bullet]T[\bullet, j]/n_0$.

The null hypothesis is that the table is generated by *any* Multinomial distribution in which rows and columns are independent. Thus, unlike in Section 4.1, even under the null hypothesis, we do not know the true distribution over the data and hence cannot write down or sample from the true sampling distribution of the test statistic.

Nevertheless under the null hypothesis of independence between rows and columns, this test statistic converges in distribution to a chi-squared random variable with $(r - 1)(c - 1)$ degrees of freedom [Ferguson, 1996]. This result is obtained by using a combination of two limits: the Central Limit Theorem (which approximates $(T[i, j] - n\theta[i, j])/\sqrt{n\theta[i, j]}$ by a Gaussian) and the fact that $E[i, j]/n \to \theta[i, j]$ in probability. Slutsky's theorem provides a justification for combining the two, to show that $(T[i, j] - n\theta[i, j])/\sqrt{nE[i, j]/n}$ converges to the same Gaussian distribution (additional algebra then yields the asymptotic distribution of the test statistic [Ferguson, 1996]).

In the privacy-preserving case, differential privacy can be achieved by adding noise directly to the table T [Gaboardi et al., 2016, Rogers and Kifer, 2017, Johnson and Shmatikov, 2013] or to the test statistic itself [Uhler et al., 2013, Yu et al., 2014].

However, ad-hoc substitution cannot be used to create an approximating distribution for the differentially private test statistic because θ is unknown. In fact, existing techniques must first estimate θ [Gaboardi et al., 2016, Rogers and Kifer, 2017], argue that the estimate converges in probability to θ, and then use Slutsky's theorem to merge this estimate with

Table 1: List of Symbols in the Asymptotic Regime

n	The variable to take to infinity; hypothetical sample size
n_0	The true sample size
$\mathbf{S}_n$	A statistic computed from data with size n
ϕ_n	The statistic we are interested in: $\phi_{n_0}(\mathbf{S}_{n_0})$
h_n	Transformation for which $h_n(\mathbf{S}_n)$ converges in distribution as $n \to \infty$
$\mathbf{Z}$	$\lim_{n\to\infty} h_n(\mathbf{S}_n) \to \mathbf{Z}$ in distribution
$\mathbf{W}$	An approximation to the distribution of $\mathbf{S}_n$
M	A privacy mechanism that can be applied to $\mathbf{S}_n$

the Central Limit Theorem. Ad-hoc substitutions generally cannot be combined with limit theorems like Slutsky's theorem because they will not always guarantee that the necessary conditions of the theorems are satisfied.

Thus, any method of constructing approximating distributions should be compatible with other mathematical limit theorems. In particular, this suggests that the method of constructing approximating distributions should be setting up a limiting process.

5. Recipe for Approximating Distributions

In this section, we present a recipe for generating approximating distributions for differentially private statistics and present results about the relative accuracy that can be expected. Then in Section 6 we present applications of these results.

To describe our setup, we consider two alternate worlds, the privacy-preserving world and the non-private world. The data scientist has a planned analysis for the hypothetical non-private world, but works in the privacy-preserving world and would like to modify the planned analysis appropriately.

The privacy-preserving world. There is a dataset $D = \{\vec{x}_1, \ldots, \vec{x}_{n_0}\}$ of size n_0 and a statistic $\mathbf{S}_{n_0}$ is computed from it (for example $\mathbf{S}_{n_0} = \sum_{i=1}^{n_0} \vec{x}_i$). A differentially private or ρ-zcdp mechanism M is applied to $\mathbf{S}_{n_0}$ – for example, M can add a Laplace random variable $\mathbf{Y}$ to $\mathbf{S}_{n_0}$.

The non-private world. In the non-private world, the analyst would have access to $\mathbf{S}_{n_0}$ directly and would be interested in some function $\phi_{n_0}(\mathbf{S}_{n_0})$ (for example, the chi-squared statistic). To approximate the distribution of $\phi_{n_0}(\mathbf{S}_{n_0})$, statisticians often hypothesize "what if the sample size were not a constant n_0, but a variable n that can be manipulated." Such an approach provides approximation results, like the central limit theorem, which states that there is a transformation $h_n(\mathbf{S}_n)$ that converges in distribution to some random variable $\mathbf{Z}$ as $n \to \infty$ (for example $h_n(\mathbf{S}_n) = \sqrt{n}(n^{-1}\mathbf{S}_n - \vec{\mu})$ where $\vec{\mu}$ is the mean for the i.i.d. samples $\vec{x}_i$'s). In our case, we assume h_n invertible and piecewise continuous (with a finite number of pieces).[5]

We summarize this notation in Table 1. In the non-private case, the statistician would compute the limit: $\lim_{n\to\infty} \phi_n(\mathbf{S}_n)$ and use the convergence in distribution of $h_n(\mathbf{S}_n)$ to $\mathbf{Z}$ as part of the computations. This limit would then be used as an approximation of $\phi_{n_0}(\mathbf{S}_{n_0})$.

[5]If h_n is not one-to-one, one can often augment it so that it becomes one-to-one. For example, if $h_n(\mathbf{S}_n) = |\mathbf{S}_n - n\vec{\mu}|/\sqrt{n}$, then h_n is not one-to-one, but it can be extended to a function $g_n(\mathbf{S}_n) = (|\mathbf{S}_n - n\vec{\mu}|/\sqrt{n},\ \text{sign}(\mathbf{S}_n - n\vec{\mu}))$ that returns the sign as well.

In the privacy-preserving case, we propose instead to compute the limit:

$$\lim_{n\to\infty} \phi_{n_0}(M(h_{n_0}^{-1}(h_n(\mathbf{S}_n))))$$

and use it as an approximation to $\phi_{n_0}(M(\mathbf{S}_{n_0}))$, since the analyst only has access to $M(\mathbf{S}_{n_0})$ instead of $\mathbf{S}_{n_0}$. Another way to think of it is that if $h_n(\mathbf{S}_n)$ converges in distribution to $\mathbf{Z}$ and we would like to approximate $h_{n_0}(\mathbf{S}_{n_0})$ with $\mathbf{Z}$, then we could approximate the distribution of $\mathbf{S}_{n_0}$ as $h_{n_0}^{-1}(\mathbf{Z})$ and thus approximate $\phi_{n_0}(M(\mathbf{S}_{n_0}))$ with

$$\phi_{n_0}(M(h_{n_0}^{-1}(\mathbf{Z}))).$$

As a simple toy example, let us revisit Example 1.1 where we have a dataset $\{x_1, \ldots, x_{n_0}\}$ of scalars. Here S_{n_0} is the non-private sum, $\phi_{n_0}(S_{n_0}) = \sqrt{n_0}(n_0^{-1}S_{n_0} - \mu)/\sigma$ is the statistic we are interested in when creating non-private confidence intervals, $h_n(S_n) = \sqrt{n}(n^{-1}S_n - \mu)/\sigma$ is the transformation of S_n whose sampling distribution can be approximated in the non-private setting. M is the mechanism that adds Laplace noise Y_ϵ with scale $1/\epsilon$ to S_n. The recipe tells us that we should derive the approximation for the sampling distribution of $\phi_{n_0}(M(S_{n_0})) = \sqrt{n_0}(n_0^{-1}(Y_\epsilon + S_{n_0}) - \mu)/\sigma$ from the limit (note that equality here means equality in distribution):

$$\begin{aligned}
\lim_{n\to\infty} \phi_{n_0}(M(h_{n_0}^{-1}(h_n(S_n)))) &= \lim_{n\to\infty} \phi_{n_0}(M(h_{n_0}^{-1}\left(\frac{\sum_{i=1}^n (x_i - \mu)}{\sigma\sqrt{n}}\right))) \\
&= \lim_{n\to\infty} \phi_{n_0}(M\left(n_0\mu + \sqrt{n_0}\frac{\sum_{i=1}^n (x_i - \mu)}{\sqrt{n}}\right)) \\
&= \lim_{n\to\infty} \phi_{n_0}\left(Laplace(1/\epsilon) + n_0\mu + \sqrt{n_0}\frac{\sum_{i=1}^n (x_i - \mu)}{\sqrt{n}}\right) \\
&= \lim_{n\to\infty} \frac{Laplace(1/\epsilon)}{\sigma\sqrt{n_0}} + \frac{\sum_{i=1}^n (x_i - \mu)}{\sigma\sqrt{n}} = \frac{Laplace(1/\epsilon)}{\sigma\sqrt{n_0}} + N(0,1).
\end{aligned}$$

This yields the intuitively expected approximation for the sampling distribution of $\sqrt{n_0}(\frac{Y_\epsilon + S_{n_0}}{n_0} - \mu)/\sigma$. Thus, if one wanted to form confidence intervals for the mean, one would find an α such that the interval $[-\alpha, \alpha]$ contains 95% of the probability mass of the convolution of a standard Gaussian with a Laplace random variable with scale $1/(\epsilon\sqrt{n_0}\sigma)$. The confidence interval for μ is then $[\frac{Y_\epsilon + S_{n_0}}{n_0} - \frac{\alpha\sigma}{\sqrt{n_0}}, \frac{Y_\epsilon + S_{n_0}}{n_0} + \frac{\alpha\sigma}{\sqrt{n_0}}]$. We defer more examples of this computation for various applications to Section 6.

The next question to ask is how accurate is the approximation to the sampling distribution – if $\sqrt{n_0}(n_0^{-1}S_{n_0} - \mu)/\sigma$ is well approximated by the standard Gaussian, then is the proposed approximation to the sampling distribution of $\sqrt{n_0}(n_0^{-1}(Y_\epsilon + S_{n_0}) - \mu)/\sigma$ also accurate? Such questions are addressed next.

5.1. **Relative Guarantees.** We first study how the mechanism M affects distance between random variables. For example, if the Kolmogorov-Smirnov distance $d_{KS}(\mathbf{X}, \mathbf{Y}) \leq \delta$ (or total variation distance $d_V(\mathbf{X}, \mathbf{Y}) \leq \delta$) then what can we say about $d_{KS}(M(\mathbf{X}), M(\mathbf{Y}))$? We will apply this result to answer the following question: if (in the non-private case) the distribution of $\mathbf{S}_{n_0}$ is close to the distribution of some random variable Φ, then will the distribution of $M(\mathbf{S}_{n_0})$ remain close to the distribution of $M(\Phi)$ (i.e., in the private case, will M preserve distances relative to the non-private case)?

The first result is that if $\mathbf{X}$ and $\mathbf{Y}$ are random variables whose distributions are very similar – their total variation distance $d_V(\mathbf{X}, \mathbf{Y}) \leq \delta$, then even if the total variation distance between $M(\mathbf{X})$ and $M(\mathbf{Y})$ is large, the Kolmogorov-Smirnov (KS) distance will be small.

Theorem 5.1. *Let $\mathbf{X}$ and $\mathbf{Y}$ be two random variables using a common measure μ, so that $F_{\mathbf{X}}(\vec{t}) = \int_{\vec{x} \preceq \vec{t}} f_{\mathbf{X}}(\vec{x})\ d\mu(\vec{x})$ and $F_{\mathbf{Y}}(\vec{y}) = \int_{\vec{y} \preceq \vec{t}} f_{\mathbf{Y}}(\vec{y})\ d\mu(\vec{y})$. Let $M : \mathbb{R}^k \to \mathbb{R}^\ell$ be a randomized function associated with conditional probability $g(\cdot \mid \cdot)$ (so that $P(M(\vec{x}) \in S) = \int_{\vec{z} \in S} g(\vec{z} \mid \vec{x})\ d\nu(\vec{z})$). Then*

$$d_{KS}(M(\mathbf{X}), M(\mathbf{Y})) \leq 2 d_V(\mathbf{X}, \mathbf{Y})$$

where the probability is over the randomness in the function M and the random variables $\mathbf{X}$ and $\mathbf{Y}$.

The proof of Theorem 5.1 is in Appendix A.1.

In some cases the KS distance between $\mathbf{X}$ (the true sampling distribution) and $\mathbf{Y}$ (its approximation) can be small even though the total variation distance between $\mathbf{X}$ and $\mathbf{Y}$ is large. This often happens when one of the distributions is discrete while the other is continuous. In this case, the Kolmogorov-Smirnov distance (restricted to some set $\mathcal{L}$) between $M(\mathbf{X})$ and $M(\mathbf{Y})$ depends strongly on the properties of M.

Theorem 5.2. *Let $\mathcal{L} \subseteq \mathbb{R}^k$. Let $\mathbf{X}$ and $\mathbf{Y}$ be random variables over $\mathbb{R}^k$ whose cumulative distribution functions can be written as $\int_{\vec{x} \preceq \vec{t}} f_{\mathbf{X}}(\vec{x})\ d\mu_1(\vec{x})$ and $\int_{\vec{y} \preceq \vec{t}} f_{\mathbf{Y}}(\vec{y})\ d\mu_2(\vec{y})$, respectively.[6] Let $M : \mathbb{R}^k \to \mathbb{R}^\ell$ be a randomized function associated with conditional probability $g(\cdot \mid \cdot)$ satisfying the following conditions:*

- *Denote $H(\vec{t} \mid \vec{x}) = \int_{\vec{z} \preceq \vec{t}} g(\vec{z} \mid \vec{x})\ d\nu(\vec{z})$, where ν is the measure used to integrate over the output space of M, and suppose there is a function $\bar{g}^\dagger$ such that $H(\vec{t} \mid \vec{x}) = \int_{\vec{s} \succeq \vec{x}} \bar{g}^\dagger(\vec{t} \mid \vec{s})\ d\mu_3(\vec{s})$ for some measure μ_3.*
- *$\int_{\vec{s} \in \mathbb{R}^k} \mathbb{1}_{\{\vec{s} \notin \mathcal{L}\}}\ d\mu_3(\vec{s}) = 0$.*

Then for all $\vec{t} \in \mathbb{R}^\ell$:

$$d_{KS}(M(\mathbf{X}), M(\mathbf{Y})) \leq d_{KS(\mathcal{L})}(\mathbf{X}, \mathbf{Y}) \sup_{\vec{t} \in \mathcal{L}} ||\bar{g}^\dagger(\vec{t} \mid \cdot)||_1,$$

where $||\bar{g}^\dagger(\vec{t} \mid \cdot)||_1 = \int_{\vec{s} \in \mathbb{R}^k} |\bar{g}^\dagger(\vec{t} \mid \vec{s})|\ d\mu_3(\vec{s})$.

The proof of Theorem 5.2 is in Appendix A.2.

The use of possibly different base measures μ_1, μ_2, μ_3 in Theorem 5.2 are necessary because the sampling distribution $\mathbf{X}$ and its approximating distribution can be of various types (e.g., discrete, continuous, mixture). Intuitively, the function $\bar{g}^\dagger$ is the derivative of the conditional CDF of M with respect to the second argument, and the blowup in distance is controlled by the L_1 norm of $\bar{g}^\dagger$. A simple corollary is that if M has a translation invariant conditional CDF (e.g., if M adds an independent random noisy variable), then the blowup in distance is at most 1:

Corollary 5.3. *Using the notation of Theorem 5.2, if the conditional CDF $g(\cdot \mid \cdot)$ of M is translation invariant (i.e. $g(\vec{z} + \vec{t} \mid \vec{x} + \vec{t}) = g(\vec{z} \mid \vec{x})$) and uses the same base measure as $\mathbf{Y}$ (i.e. $P(M(\vec{x}) \preceq \vec{t}) = \int_{\vec{z} \preceq \vec{t}} g(\vec{z} \mid \vec{x})\ d\mu_2(\vec{z})$). Then*

$$d_{KS(\mathcal{L})}(M(\mathbf{X}), M(\mathbf{Y})) \leq d_{KS(\mathcal{L})}(\mathbf{X}, \mathbf{Y}).$$

[6] Note that μ_1 and μ_2 can be different measures, so, for example $\mathbf{X}$ can be discrete and $\mathbf{Y}$ can be continuous.

The proof of Corollary 5.3 is in Appendix A.3.

This result allows us to re-use existing convergence results, such as the Berry-Esseen inequality:

Theorem 5.4 Feller [1971]**.** *Let $X_1, X_2, \ldots,$ be i.i.d. random variables with $E[X_i] = \mu$, $E[|X_i - \mu|^2] = \sigma^2 > 0$, and $E[|X_i - \mu|^3] = \rho < \infty$. Let Y be a standard Gaussian random variable. Then $d_{KS}(\frac{\sum_{i=1}^n (X_i - \mu)}{\sigma\sqrt{n}}, Y) \leq \frac{3\rho}{\sigma^3\sqrt{n}}$.* □

Combing Corollary 5.3 with Theorem 5.4 immediately yields:

Corollary 5.5. *Let $X_1, X_2, \ldots,$ be i.i.d. random variables with $E[X_i] = \mu$, $E[|X_i - \mu|^2] = \sigma^2 > 0$, and $E[|X_i - \mu|^3] = \rho < \infty$. Let Y be a standard Gaussian random variable. Let Z_1 and Z_2 be independent Laplace$(1/\epsilon)$ random variables. Then $d_{KS}(\frac{Z_1 + \sum_{i=1}^n (X_i - \mu)}{\sigma\sqrt{n}}, Y + \frac{Z_2}{\sigma\sqrt{n}}) \leq \frac{3\rho}{\sigma^3\sqrt{n}}$.* □

Continuing our toy example from Example 1.1 whose sampling distribution was approximated in the beginning of Section 5, Corollary 5.3 tells us that this approximation is as close to the sampling distribution of the private statistic as the central limit theorem is close to the sampling distribution of the non-private statistic. Meanwhile, Corollary 5.5 provides the rate of convergence.

Similar results can be obtained for the Wasserstein metric. For these results, we view a differentially private algorithm as a function of two variables: $f(\mathbf{X}, \mathbf{B})$. The first is a k-dimensional vector corresponding to the input (for example, it could contain the mean and estimated standard deviation of the data, or it could contain the value of a non-private chi-squared statistic). The second is a finite-dimensional random variable (e.g., a sequence of random bits that guide the operation of the algorithm).

Lemma 5.6. *Let M be a mechanism that satisfies ϵ-differential privacy and that works as follows. On input $\mathbf{X} \in \mathbb{R}^k$, M samples a finite-dimensional random vector $\mathbf{B}$ from a distribution μ_b and then returns $f(\mathbf{X}, \mathbf{B})$ for some function $f : \mathbb{R}^k \times \mathbb{R}^b \to \mathbb{R}^\ell$. Let $\mathbf{X}$ and $\mathbf{Y}$ be random vectors in $\mathbb{R}^k$ having distributions μ_1 and μ_2, respectively. If all of the following conditions are satisfied:*

- *f is L-Lipschitz continuous in its first argument (i.e. $\mathbf{X}$)*
- *$E[||f(\vec{0}, \mathbf{B})||_2] < \infty$*
- *$E[||\mathbf{X}||_2] < \infty$*
- *$E[||\mathbf{Y}||_2] < \infty$*

then the following holds: $d_W(\mathbf{X}, \mathbf{Y}) \leq \delta \Rightarrow d_W(M(\mathbf{X}), M(\mathbf{Y})) \leq L\delta$.

For proof see Appendix A.4.

Clearly, if M adds independent random noise (i.e. $f(\mathbf{X}, \mathbf{B}) = \mathbf{X} + \mathbf{B}$) then f is 1-Lipschitz continuous. This allows us to re-use convergence results based on Stein's method, such as the following.

Theorem 5.7 Ross [2011]**.** *Let $X_1, X_2, \ldots,$ be i.i.d. random variables with $E[X_i] = \mu$, $E[|X_i - \mu|^2] = \sigma^2 > 0$, $E[|X_i - \mu|^3] = \rho_3 < \infty$, and $E[|X_i - \mu|^4] = \rho_4 < \infty$. Let Y be a standard Gaussian random variable. Then $d_W(\frac{\sum_{i=1}^n (X_i - \mu)}{\sigma\sqrt{n}}, Y) \leq \frac{1\rho_3}{\sigma^3\sqrt{n}} + \frac{\sqrt{2\rho_4}}{\sqrt{\pi n \sigma^2}}$.* □

Combining Lemma 5.6 and Theorem 5.7, we get the following guarantee for the approximating distribution of a noisy sum that arises from the proposed recipe:

Corollary 5.8. *Let $X_1, X_2, \ldots,$ be i.i.d. random variables with $E[X_i] = \mu$, $E[|X_i - \mu|^2] = \sigma^2 > 0$, and $E[|X_i - \mu|^3] = \rho < \infty$. Let Y be a standard Gaussian random variable. Let Z_1 and Z_2 be independent Laplace$(1/\epsilon)$ random variables. Then $d_W(\frac{Z_1 + \sum_{i=1}^{n}(X_i - \mu)}{\sigma\sqrt{n}}, Y + \frac{Z_2}{\sigma\sqrt{n}}) \leq \frac{1\rho_3}{\sigma^3\sqrt{n}} + \frac{\sqrt{2\rho_4}}{\sqrt{\pi n}\sigma^2}$.* □

5.2. **Degradation due to Postprocessing.** The results of Section 5.1 bound the degradation in approximation quality when M is applied to a random variable (compared to when M is applied to its approximation).

However, a data scientist is most often interested in some function $\phi_{n_0}(M(\mathbf{S}_{n_0}))$ and would like to know if applying ϕ_{n_0} will preserve the quality of the approximation. For some simple operations, the approximation is preserved but for others it can become arbitrarily bad. Such considerations can inform the design of privacy-preserving analysis. In the following theorem we consider common operations used to create test statistics (for example, in chi-squared testing [Gaboardi et al., 2016, Rogers and Kifer, 2017], the postprocessing would include pointwise squaring followed by summation). In this theorem we now interpret $\mathbf{X}$ to be $M(\mathbf{S}_{n_0})$, $\mathbf{W}$ to be the approximation of $\mathbf{S}_{n_0}$, and $\mathbf{Y}$ to be $M(\mathbf{W})$.

Theorem 5.9. *Let $\mathbf{X} = (X_1, \ldots, X_k)$ and $\mathbf{Y} = (Y_1, \ldots, Y_k)$ be k-dimensional random variables and suppose $d_{KS(\mathcal{L})}(\mathbf{X}, \mathbf{Y}) \leq \delta$ then:*

1. *If ϕ is the coordinate projection operator that selects a fixed subset of the components of a vector (e.g., $\phi(\vec{t}) = (t_2, t_4, t_5)$) then $d_{KS(\mathcal{L})}(\phi(\mathbf{X}), \phi(\mathbf{Y})) \leq \delta$.*
2. *If ϕ is the sum of coordinates (i.e. $\phi(\vec{t}) = t_1 + t_2 + \cdots + t_k$) then there exist $\mathbf{X}$ and $\mathbf{Y}$ such that $d_{KS(\mathcal{L})}(\mathbf{X}, \mathbf{Y}) \leq \delta$ but $d_{KS(\mathcal{L})}(\phi(\mathbf{X}), \phi(\mathbf{Y})) = 1$.*
3. *If $\phi(\vec{t}) \equiv c\vec{t} + \vec{a}$, where c is a scalar, then $d_{KS(\mathcal{L})}(\phi(\mathbf{X}), \phi(\mathbf{Y})) \leq \delta$.*
4. *Even if ϕ is continuous, is one-to-one, and preserves partial orders (i.e. $\vec{t} \preceq \vec{s} \Rightarrow \phi(\vec{t}) \preceq \phi(\vec{s})$) then there still exist $\mathbf{X}$ and $\mathbf{Y}$ such that $d_{KS(\mathcal{L})}(\mathbf{X}, \mathbf{Y}) \leq \delta$ but $d_{KS(\mathcal{L})}(\phi(\mathbf{X}), \phi(\mathbf{Y})) = 1$.*

The proof is in Appendix A.5.

The consequence of the theorem is the following. If the distribution of a random vector $\mathbf{X}$ is close (in the KS distance) to the distribution of a random vector $\mathbf{Y}$ then any marginal of the variables in $\mathbf{X}$ is equally close to the corresponding marginal of $\mathbf{Y}$. However, the sum of the random variables $\sum X_i$ might have a significantly different distribution (in the KS distance) than $\sum Y_i$.

Thus, in general, when applied to statistics, the distribution of a random variable $\mathbf{W}$ could be a good approximation of the sampling distribution of $\mathbf{S}_{n_0}$ (in the sense of the KS-distance), however this does not guarantee that $\phi_{n_0}(\mathbf{W})$ is a good approximation to $\phi_{n_0}(\mathbf{S}_{n_0})$ (unless $\mathbf{W}$ is close to $\mathbf{S}_{n_0}$ under a stronger metric). Despite this negative result, simulations (e.g., in Section 6.1) suggest that the worst-case is pathological. However, from a purely statistical perspective, output perturbation (designing a mechanism that directly releases a noisy version of $\phi_{n_0}(\mathbf{S}_{n_0})$) appears preferable to input perturbation (i.e. releasing a noisy version of $\mathbf{S}_{n_0}$ and then applying ϕ_{n_0} to the result) as there is no post-processing to worry about.

For the Wasserstein distance d_W, the following postprocessing result is almost trivial:

Lemma 5.10. *If ϕ is L-Lipschitz continuous then $d_W(\mathbf{X}, \mathbf{Y}) \leq \delta \Rightarrow d_W(\phi(\mathbf{X}), \phi(\mathbf{Y})) \leq L\delta$.* □

The coordinate projection operation is 1-Lipschitz continuous and so preserves the Wasserstein distance. The summation of coordinates is $\sqrt{d}$-Lipschitz continuous (where d is the dimensionality) and so blows up the Wasserstein distance by at most $\sqrt{d}$, unlike the KS distance which has no such guarantee. However, an operation such as squaring (which is used in chi-squared testing) is not Lipschitz continuous and so could distort two distributions arbitrarily. In such cases even stronger notions of distances are needed. We will examine such cases in Section 6.

6. Applications

We present a variety of applications and show how to apply the recipe for approximating distributions in this section. We also study relative guarantees and discuss research directions in non-private convergence rates that are needed.

6.1. Chi-squared Goodness-of-fit Test with Input Perturbation.

We use the same setup and notation as in Section 4.1.

In the non-private world, a statistician would typically compute the test statistic $\phi_{n_0}(\mathbf{S}_{n_0}) = \sum_{j=1}^{k} \frac{(\mathbf{S}_{n_0}[j]-n_0\vec{\theta}[j])^2}{n_0\vec{\theta}[j]}$. If $\mathbf{S}_{n_0}$ really is generated by Multinomial$(n_0, \vec{\theta})$, the Central Limit Theorem states that $\lim_{n\to\infty} \frac{\mathbf{S}_n - n\vec{\theta}}{\sqrt{n}}$ converges in distribution to the multivariate Gaussian $N(\vec{0}, \text{diag}(\vec{\theta}) - \vec{\theta}\vec{\theta}^t)$ and it follows that $\lim_{n\to\infty} \phi_n(\mathbf{S}_n)$ converges in distribution to a chi-squared random variable with $k-1$ degrees of freedom [Ferguson, 1996]. Thus the analyst computes the probability that a chi-squared random variable exceeds the actual value of $\sum_{j=1}^{k} \frac{(\mathbf{S}_{n_0}[j]-n_0\vec{\theta}[j])^2}{n_0\vec{\theta}[j]}$. If this probability (i.e. the p-value) is small enough (e.g., ≤ 0.01), then the analyst could conclude that the data were probably not generated by Multinomial$(n_0, \vec{\theta})$.

In this example, we consider the use of Gaussian noise to perturb the table to protect privacy (as in Gaboardi et al. [2016]). Specifically, to satisfy ρ-zcdp, a mechanism M could add $N(\vec{0}, (1/\rho)\mathbf{I})$ noise to $\mathbf{S}_{n_0}$ (on the other hand, if one wanted to use ϵ-differential privacy, one could use Laplace noise instead). The data analyst would only be given access to $M(\mathbf{S}_{n_0}) = \mathbf{Y} + \mathbf{S}_{n_0} \equiv \tilde{\mathbf{S}}_{n_0}$ with $\mathbf{Y} \sim N(\vec{0}, (1/\rho)\mathbf{I})$. The analyst can then compute $\phi_{n_0}(\tilde{\mathbf{S}}_{n_0}) \equiv \sum_{j=1}^{k} \frac{(\tilde{\mathbf{S}}_{n_0}[j]-n_0\vec{\theta}[j])^2}{n_0\vec{\theta}[j]}$.

Lemma 6.1. *The above computation of $\phi_{n_0}(\tilde{\mathbf{S}}_{n_0})$ satisfies ρ-zcdp.*

Proof. Modifying the record of one person's data would lead to one cell from $\mathbf{S}_{n_0}$ increasing by one and another cell decreasing by one. So $\text{Sen}_2(\mathbf{S}_{n_0}) = \sqrt{2}$. Then, by the Gaussian Mechanism in Lemma 3.4, adding noise from $N(\vec{0}, (1/\rho)\mathbf{I})$ to $\mathbf{S}_{n_0}$ would satisfy ρ-zcdp. The computation of $\phi_{n_0}(\tilde{\mathbf{S}}_{n_0})$ is just post-processing on $\tilde{\mathbf{S}}_{n_0}$, therefore it also satisfies ρ-zcdp by the post-processing property of zero-concentrated differential privacy. □

We still need an approximation to $\phi_{n_0}(\tilde{\mathbf{S}}_{n_0})$ under the null hypothesis. In our proposed scheme, we take the limit of $\phi_{n_0}(M(h_{n_0}^{-1}(h_n(\mathbf{S}_n))))$. We break this computation into simple steps:

- $h_n(\mathbf{S}_n) = (\mathbf{S}_n - n\vec{\theta})/\sqrt{n}$.
- $h_{n_0}^{-1}(h_n(\mathbf{S}_n)) = n_0\vec{\theta} + \frac{\sqrt{n_0}}{\sqrt{n}}(\mathbf{S}_n - n\vec{\theta})$.
- $M(h_{n_0}^{-1}(h_n(\mathbf{S}_n))) = \mathbf{Y} + n_0\vec{\theta} + \frac{\sqrt{n_0}}{\sqrt{n}}(\mathbf{S}_n - n\vec{\theta})$.
- $\phi_{n_0}(M(h_{n_0}^{-1}(h_n(\mathbf{S}_n)))) = \sum_{i=1}^{k}\left(\frac{\mathbf{Y}[i]}{\sqrt{n_0}} + \frac{\mathbf{S}_n[i]-n\vec{\theta}[i]}{\sqrt{n}}\right)^2 \frac{1}{\vec{\theta}[i]}$.

Now, taking the limit as $n \to \infty$ and using Slutsky's theorem [Ferguson, 1996] we get

$$\lim_{n\to\infty} \phi_{n_0}(M(h_{n_0}^{-1}(h_n(\mathbf{S}_n)))) = \lim_{n\to\infty} \sum_{i=1}^{k}\left(\frac{\mathbf{Y}[i]}{\sqrt{n_0}} + \frac{\mathbf{S}_n[i] - n\vec{\theta}[i]}{\sqrt{n}}\right)^2 \frac{1}{\vec{\theta}[i]} = \sum_{i=1}^{k}\left(\frac{\mathbf{Y}[i]}{\sqrt{n_0}} + \mathbf{A}[i]\right)^2 \frac{1}{\vec{\theta}[i]}$$

in distribution, where $\mathbf{Y} \sim N(\vec{0}, (1/\rho)\mathbf{I})$ and $\mathbf{A} \sim N(\vec{0}, \mathrm{diag}(\vec{\theta}) - \vec{\theta}\vec{\theta}^t)$.

To summarize, a data analyst computes a noisy chi-square statistic $\sum_{i=1}^{k}\left(\frac{\mathbf{Y}[i]}{\sqrt{n_0}} + \frac{\mathbf{S}_n[i]-n\vec{\theta}[i]}{\sqrt{n}}\right)^2 \frac{1}{\vec{\theta}[i]}$ and can estimate its sampling distribution in two ways.

- The first is to use the approximation $\sum_{i=1}^{k}\left(\frac{\mathbf{Y}[i]}{\sqrt{n_0}} + \mathbf{A}[i]\right)^2 \frac{1}{\vec{\theta}[i]}$, whose tails can be numerically evaluated [Gaboardi et al., 2016] and ends up as equivalent to the intuitive process of replacing the true data $\frac{\mathbf{S}_{n_0}[i]-n_0\vec{\theta}[i]}{\sqrt{n_0}}$ with its Gaussian approximation.
- The second method is to directly sample from the sampling distribution: sample a new copy $\mathbf{S}_{n_0}$ from Multinomial$(n_0, \vec{\theta})$, fresh noise $\mathbf{Y}$, and compute $\frac{\mathbf{Y}[i]}{\sqrt{n_0}} + \frac{\mathbf{S}_{n_0}[i]-n_0\vec{\theta}[i]}{\sqrt{n_0}}$. Many samples are required to get an estimate of the tail probabilities under the null hypothesis.

The first method is much faster, but an important question is how accurate it is in practice. Corollary 5.3 and Lemma 5.6 imply that if the distribution of the true data $\frac{\mathbf{S}_{n_0}-n_0\vec{\theta}}{\sqrt{n_0}}$ is well-approximated by the Gaussian $\mathbf{A}$ then the distribution of the noisy data $(\frac{\mathbf{Y}}{\sqrt{n_0}} + \frac{\mathbf{S}_{n_0}-n_0\vec{\theta}}{\sqrt{n_0}})$ is approximated by the noisy Gaussian $(\frac{\mathbf{Y}}{\sqrt{n_0}} + \mathbf{A})$, in Wasserstein and Kolmogorov-Smirnov distances, without loss in approximation quality. However, squaring the terms and subsequent summation can decrease the approximation quality under the Kolmogorov-Smirnov and Wasserstein distances in the worst case (e.g., Theorem 5.9).

In practice [Wang et al., 2015, Gaboardi et al., 2016], the resulting approximating distribution for the noisy chi-square statistic appears accurate enough for hypothesis testing. This suggests that an even stronger notion of convergence is occurring. One possibility is to use several measures of the distance between two random variables $\mathbf{X}$ and $\mathbf{Y}$ as follows [Gaunt, 2015]. For example, let H_ℓ be a class of bounded functions that are ℓ-times continuously differentiable with bounded derivatives and let G_ℓ be a class of functions that are ℓ-times continuously differentiable with derivatives that are bounded by a specific polynomial. For instance, G_ℓ could consist of the single function $z \to z^2$ that performs squaring. Following Gaunt [2015], define:

$$d_{H_\ell}(\mathbf{X}, \mathbf{Y}) = \sup_{h\in H} E[h(\mathbf{X})] - E[h(\mathbf{Y})]$$

$$d_{H_\ell G_\ell}(\mathbf{X}, \mathbf{Y}) = \sup_{h \in H, g \in G} E[h(g(\mathbf{X}))] - E[h(g(\mathbf{Y}))].$$

Clearly, if $d_{H_\ell G_\ell}(\mathbf{X}, \mathbf{Y}) \leq \delta$ then $d_{H_\ell}(g(\mathbf{X}), g(\mathbf{Y})) \leq \delta$. Such a family of metrics would neatly capture worst case degradation due to postprocessing in the non-private case.

In the privacy preserving case, we are interested in a statistic ϕ of the privacy-preserving data $M(\mathbf{X})$ that is produced by a mechanism. We would like a similar type of guarantee: if $d_{H_\ell G_\ell}(\mathbf{X}, \mathbf{Y}) \leq \delta$ then $d_{H_\ell G_\ell}(M(\mathbf{X}), M(\mathbf{Y})) \leq \delta'$ (for some δ') and $d_{H_\ell}(\phi(M(\mathbf{X})), \phi(M(\mathbf{Y}))) \leq \delta''$ (for some δ''). It appears that additional restrictions on the class of functions G_ℓ would be required. For example,

Lemma 6.2. *Let M be a mechanism that adds some random variable $\mathbf{Z}$ to its input. Suppose that $\phi \in G_\ell$ and that G_ℓ is closed under translation of its input (i.e., if $g \in G$ then for every c, the function $x \to g(x + c)$ belongs to G_ℓ) then:*

$$d_{H_\ell G_\ell}(\mathbf{X}, \mathbf{Y}) \leq \delta \Rightarrow d_{H_\ell G_\ell}(M(\mathbf{X}), M(\mathbf{Y})) \leq \delta \Rightarrow d_{H_\ell}(\phi(M(\mathbf{X})), \phi(M(\mathbf{Y}))) \leq \delta.$$

The proof can be found in Appendix A.6

Nonprivate multivariate convergence results of the necessary form are an open area of research for many applications [Gaunt, 2015].

To facilitate the understanding for our recipe, we create several toy simulations and show the results in Figures 4 through 8. In each case, we set $\theta_0 = (0.1, 0.1, 0.3, 0.5)$. Figure 4 plots the sampling distribution of the non-private chi-squared statistic (with $n_0 = 10,000$) against the standard asymptotic approximation. Figure 7 shows the same plots but for $n = 2,000$. In both cases, the asymptotic approximation agrees well with the sampling distributions. Figures 5 and 6 show the corresponding results for the privacy-preserving case (ρ-zCDP) with $\rho = 0.001$ and 0.01, respectively. We see that our approximating distribution matches the true sampling distribution much better than the naive approximation (i.e. the asymptotic distribution when $n \to \infty$). Visually, the approximation is as good as in the non-private case (Figure 4). We note that the naive approximation gets better when the noise is smaller relative to the sample size (e.g., $\rho = 0.01$ when $n_0 = 10,000$). We can see similar results when comparing Figures 7 and 8. In this case the sample size is smaller but the proposed approximation is still accurate while the naive asymptotic distribution is inaccurate, even for $\rho = 0.01$.

We next look at how these approximations affect hypothesis testing. We use the same setting as Figures 7 and 8.

Example 6.3. *Consider the toy numerical example in Table 2a. Setting $\rho = 0.01$ we can achieve ρ-zCDP by adding $N(\vec{0}, (1/\rho)\mathbf{I})$ to the data. One such realization is shown in Table 2b. To test goodness of fit to the Multinomial$(n = 2,000, \vec{\theta}_0 = (0.1, 0.1, 0.3, 0.5))$ distribution, we compute the chi-squared statistics over both tables. The chi-squared statistic in Table 2a is equal to 3.557 with a corresponding p-value of 0.313 (when approximating the sampling distribution as $\chi^2(3)$). Note that the original table therefore does not provide much evidence against the null hypothesis. On the other hand, the chi-squared statistic in Table 2b is equal to 8.186. If we use the naive asymptotic distribution (i.e., $\chi^2(3)$) it leads to an estimated p-value of 0.042 because this approximation has much smaller tails than the true sampling distribution. But if we use the proposed approximating distribution, we get an estimated p-value of 0.12 for Table 2b. This illustrates that the privacy noise increases variability in the p-values. However, since the approximating distribution closely matches*

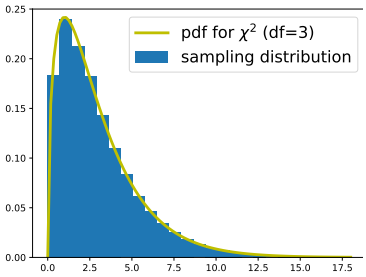

Figure 4: Sampling distribution of the non-private chi-squared statistic ($n_0 = 10,000$) compared to its standard approximation.

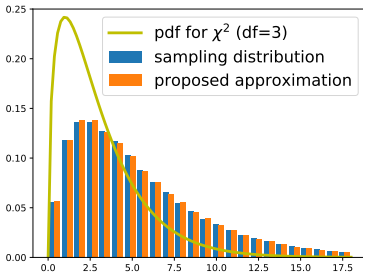

Figure 5: Sampling distribution of privacy preserving chi-squared statistic vs. naive asymptotic approximation (taking $n \to \infty$) vs. approximation provided by the recipe in Section 6.1 (with $\rho = 0.001$ in zCDP).

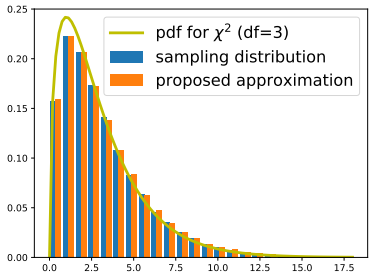

Figure 6: Same comparison as Figure 5 but with $\rho = 0.01$ in zCDP.

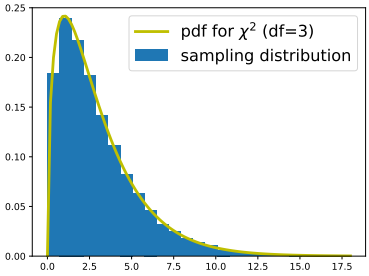

Figure 7: Sampling distribution of the non-private chi-squared statistic ($n_0 = 2,000$) compared to its standard approximation.

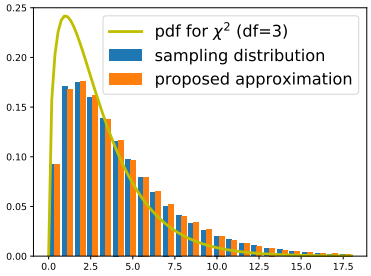

Figure 8: Sampling distribution of privacy preserving chi-squared statistic vs. naive asymptotic approximation (taking $n \to \infty$) vs. approximation provided by the recipe in Section 6.1 (with $\rho = 0.01$ in zCDP).

the sampling distribution under the null hypothesis (Figure 8), the p-value is correct (e.g., setting the rejection threshold at 0.05 *will result in a Type I error of approximately* 0.05*).*

197	201	637	965

(A) Original table

186.6	219.5	646.9	958.5

(B) 0.01-zCDP table

Table 2: A Toy Example on Goodness-of-fit Test

6.2. **Chi-squared Goodness-of-fit Test with Output Perturbation.** We now consider the same setup as in Sections 4.1 and 6.1, but using output perturbation instead of input perturbation.

To protect ϵ-differential privacy using output perturbation, a privacy mechanism M could add a random variable Y from the distribution $\text{Lap}(Sen_1(\phi_{n_0})/\epsilon)$ to $\phi_{n_0}(\mathbf{S}_{n_0}) \equiv \sum_{j=1}^{k} \frac{(\mathbf{S}_{n_0}[j]-n_0\vec{\theta}[j])^2}{n_0\vec{\theta}[j]}$ to obtain a noisy $\tilde{\phi}_{n_0}$. The sensitivity is:

Lemma 6.4. *$Sen_1(\phi_{n_0}) = \frac{1}{n_0}\max_{u,v:u\neq v}\left(\frac{2n_0-1}{\vec{\theta}[u]} + \frac{2n_0+1}{\vec{\theta}[v]}\right)$ for the chi-squared goodness of fit test statistic ϕ_{n_0}.*

The proof of Lemma 6.4 is in Appendix A.7.[7]

Thus, we are interested in an approximating distribution (under the null hypothesis) to $\tilde{\phi}_{n_0}(\mathbf{S}_{n_0}) = \phi_{n_0}(\mathbf{S}_{n_0}) + Y$ with $Y \sim \text{Lap}(Sen_1(\phi_{n_0})/\epsilon)$. In our proposed scheme, we take the limit of $M(\phi_{n_0}(h_{n_0}^{-1}(h_n(\mathbf{S}_n))))$. We break this computation into simple steps:

- $h_n(\mathbf{S}_n) = (\mathbf{S}_n - n\vec{\theta})/\sqrt{n}$.
- $h_{n_0}^{-1}(h_n(\mathbf{S}_n)) = n_0\vec{\theta} + \frac{\sqrt{n_0}}{\sqrt{n}}(\mathbf{S}_n - n\vec{\theta})$.
- $\phi_{n_0}(h_{n_0}^{-1}(h_n(\mathbf{S}_n))) = \sum_{i=1}^{k}\left(\frac{\mathbf{S}_n[i]-n\vec{\theta}[i]}{\sqrt{n}}\right)^2 \Big/ \vec{\theta}[i]$.
- $M(\phi_{n_0}(h_{n_0}^{-1}(h_n(\mathbf{S}_n)))) = \left[\sum_{i=1}^{k}\left(\frac{\mathbf{S}_n[i]-n\vec{\theta}[i]}{\sqrt{n}}\right)^2 \Big/ \vec{\theta}[i]\right] + Y$.

Now, taking the limit as $n \to \infty$, we get

$$\lim_{n\to\infty} M(\phi_{n_0}(h_{n_0}^{-1}(h_n(\mathbf{S}_n)))) = \lim_{n\to\infty}\left[\sum_{i=1}^{k}\left(\frac{\mathbf{S}_n[i]-n\vec{\theta}[i]}{\sqrt{n}}\right)^2 \Big/ \vec{\theta}[i]\right] + Y = \chi^2_{k-1} + Y$$

since the term in brackets is known to converge to χ^2_{k-1}, a chi-squared random variable with $k-1$ degrees of freedom [Ferguson, 1996]. Thus the approximating distribution is the convolution of a chi-square and a Laplace random variable (a similar approximation to what was proposed in Uhler et al. [2013] for the different problem of chi-squared independence testing). Corollary 5.3 and Lemma 5.6 then show that under the null hypothesis, this approximation is as close to the true sampling distribution of $\tilde{\phi}_{n_0}$ as the nonprivate approximation is to ϕ_{n_0}.

One advantage of using this approximating distribution instead of the true distribution under the null hypothesis is that the approximating distribution does not use $\vec{\theta}$ at all, so

[7]For independence testing, which uses a slightly different test statistic, the sensitivity can be found in Uhler et al. [2013].

one could develop a numerical algorithm for computing the tails of that distribution (a one-dimensional integral). In general, this could be faster than sampling from the true distribution (which would entail generating multiple datasets from $\vec{\theta}$ and computing the test statistic $\tilde{\phi}_{n_0}$ for each one – since a large number of samples would be needed to get an estimate for the tails of the distribution).

6.3. **Chi-squared Test of Independence with Input Perturbation.** We use the same setup for independence testing as Section 4.2.

Following the same notation, the row marginals $T_{n_0}[1,\bullet], T_{n_0}[2,\bullet], \ldots, T_{n_0}[r,\bullet]$ are modeled as a Multinomial random variable $\mathbf{U} \sim \text{Multinomial}(n_0, \pi^{(1)})$, where $\pi^{(1)}$ is *unknown*. Similarly, the column marginals $T_{n_0}[\bullet,1]$, $T_{n_0}[\bullet,2], \ldots, T_{n_0}[\bullet,c]$ are modeled as a Multinomial random variable $\mathbf{V} \sim \text{Multinomial}(n_0, \pi^{(2)})$ where $\pi^{(2)}$ is unknown.

The null hypothesis of independence between $\mathbf{U}$ and $\mathbf{V}$ is that the entire T_{n_0} is generated from a Multinomial$(n_0, \vec{\theta})$ distribution in which $\vec{\theta}$ has the form $\vec{\theta} = \text{vec}\,(\pi^{(1)}(\pi^{(2)})^t)$ for some unknown $\pi^{(1)}$ and $\pi^{(2)}$ (so the alternate hypothesis is that R and C are correlated and so T_{n_0} is generated by some other Multinomial with row marginals Multinomial$(n_0, \phi^{(1)})$ and column marginals Multinomial$(n_0, \phi^{(2)})$).

In the non-private world, an analyst using the chi-square test of independence would first compute estimators for the parameters $\pi^{(1)}$ and $\pi^{(2)}$ as follows: $\pi^{(1)}_i = T[i,\bullet]/n_0$ and $\pi^{(2)}_j = T[\bullet,j]/n_0$. Then they would compute estimated cell counts under the null hypothesis: $E[i,j] = T[i,\bullet]T[\bullet,j]/n_0$. Finally, they would compute the chi-squared statistic: $\phi_{n_0}(T_{n_0}) = \sum_{i=1}^{r}\sum_{j=1}^{c} \frac{(T[i,j]-E[i,j])^2}{E[i,j]}$. Under the null hypothesis, the asymptotic distribution of the chi-squared statistic is a chi-squared random variable with $(r-1)(c-1)$ degrees of freedom [Ferguson, 1996]. The p-value can be approximated as the probability that the chi-squared random variable exceeds $\phi_{n_0}(T_{n_0})$. A low p-value (e.g., ≤ 0.01) indicates strong evidence against the null hypothesis.

In the non-private case, the convergence is proved in two steps. In the first step, one proves that $\phi_{n_0}(T_{n_0}) = \sum_{i=1}^{r}\sum_{j=1}^{c} \frac{(T[i,j]-E[i,j])^2}{n\vec{\theta}[i.j]}$ converges to the chi-square distribution with $(r-1)(c-1)$ degrees of freedom. The second step uses Slutsky's theorem and the fact that $E[i,j]/n$ converges to $\vec{\theta}[i,j]$ in probability to conclude that $\vec{\theta}[i,j]$ can be replaced in the denominator with $E[i,j]/n$ without affecting the nonprivate asymptotic distribution.

One way to perform independence testing with ρ-zcdp and input perturbation is the following: a mechanism M can add $N(\vec{0}, (1/\rho)\mathbf{I})$ to T_{n_0}, as justified by the Gaussian Mechanism in Lemma 3.4 and the fact that $Sen_2(T_{n_0}) = \sqrt{2}$. The data analyst can be given access to $M(T_{n_0}) = T_{n_0} + \mathbf{Y} \equiv \tilde{T}_{n_0}$ with $\mathbf{Y} \sim N(\vec{0}, (1/\rho)\mathbf{I})$. The analyst can then compute $\phi_{n_0}(\tilde{T}_{n_0}) \equiv \sum_{i=1}^{r}\sum_{j=1}^{c} \frac{(\tilde{T}[i,j]-\tilde{E}[i,j])^2}{\tilde{E}[i,j]}$ with $\tilde{E}[i,j] = \tilde{T}[i,\bullet]\tilde{T}[\bullet,j]/\tilde{T}[\bullet,\bullet]$. We also define $\widetilde{\theta}[i,j] = \tilde{E}[i,j]/n_0$.

We still need the approximating distribution for $\phi_{n_0}(\tilde{T}_{n_0})$ under the null hypothesis. In our proposed scheme, we take the limit of $\phi_{n_0}(M(h_{n_0}^{-1}(h_n(T_n))))$. We break this computation into simple steps:

- $h_n(T_n) = (T_n - n\vec{\theta})/\sqrt{n}$.
- $h_{n_0}^{-1}(h_n(T_n)) = n_0\vec{\theta} + \frac{\sqrt{n_0}}{\sqrt{n}}(T_n - n\vec{\theta})$.
- $M(h_{n_0}^{-1}(h_n(T_n))) = \mathbf{Y} + n_0\vec{\theta} + \frac{\sqrt{n_0}}{\sqrt{n}}(T_n - n\vec{\theta})$.

- $\phi_{n_0}(M(h_{n_0}^{-1}(h_n(T_n)))) = \sum_{i=1}^{r}\sum_{j=1}^{c} \frac{\left(\mathbf{Y}[i,j]+n_0\vec{\theta}[i,j]+\frac{\sqrt{n_0}}{\sqrt{n}}(T_n[i,j]-n\vec{\theta}[i,j])-n_0\widetilde{\theta}[i,j]\right)^2}{n_0\widetilde{\theta}[i,j]}$.

Now, taking the limit as $n \to \infty$ and using Slutsky's theorem [Ferguson, 1996] we get

$$\lim_{n\to\infty} \phi_{n_0}(M(h_{n_0}^{-1}(h_n(T_n)))) = \lim_{n\to\infty} \sum_{i=1}^{r}\sum_{j=1}^{c} \frac{\left(\mathbf{Y}[i,j] + n_0\vec{\theta}[i,j] + \frac{\sqrt{n_0}}{\sqrt{n}}(T_n[i,j] - n\vec{\theta}[i,j]) - n_0\widetilde{\theta}[i,j]\right)^2}{n_0\widetilde{\theta}[i,j]}$$
$$= \sum_{i=1}^{r}\sum_{j=1}^{c} \frac{(\mathbf{Y}[i,j] + \sqrt{n_0}\mathbf{A}[i,j])^2}{n_0\widetilde{\theta}[i,j]}$$

in distribution, where $\mathbf{Y} \sim N(\vec{0}, (1/\rho)\mathbf{I})$ and $\mathbf{A} \sim N(\vec{0}, \mathrm{diag}(\vec{\theta}) - \vec{\theta}\vec{\theta}^t)$. The last equality is achieved partially by $\lim_{n\to\infty} \widetilde{\theta} = \vec{\theta}$.

Note that this limit is equivalent to taking $\phi_{n_0}(\tilde{T}_{n_0}) = \phi_{n_0}(T_{n_0} + \mathbf{Y})$ (where $\mathbf{Y} \sim N(\vec{0}, (1/\rho)\mathbf{I})$ and taking the limit as $n_0 \to \infty$ and $n_0\rho \to$constant, as proposed by Rogers and Kifer [2017].

Thus there are two ways of approximating the tails of the resulting statistic:

- One can estimate the parameter $\vec{\theta}$ and plug it into the approximating distribution. Then one can numerically work out the tails of the above distribution (as it is a polynomial of Gaussians) or one can sample from it many times to estimate the tails.
- On the other hand, following Gaboardi et al. [2016], one could estimate $\tilde{\theta}[i,j] = \tilde{E}[i,j]/\tilde{T}[\bullet,\bullet]$ (as long as none of the noisy counts are negative), or one could use more complicated estimation methods [Gaboardi et al., 2016, Rogers and Kifer, 2017], then sample m datasets $T_{n_0}^{(1)}, \ldots, T_{n_0}^{(m)}$ from the Multinomial$(n_0, \tilde{\theta})$ distribution, run M on each one to get $\tilde{T}_{n_0}^{(1)}, \ldots, \tilde{T}_{n_0}^{(m)}$, and compute $\phi_{n_0}(\tilde{T}_{n_0}^{(1)}), \ldots, \phi_{n_0}(\tilde{T}_{n_0}^{(m)})$. The p-value (probability of incorrectly declaring that rows and columns are not independent) is $\frac{\left|\left\{i:\ \phi_{n_0}(\tilde{T}_{n_0}^{(i)}) \geq \phi_{n_0}(\tilde{T}_{n_0})\right\}\right|}{m}$.

In both of these cases, one can approximate the tails of the distribution of $\phi_{n_0}(\tilde{T}_{n_0})$ by sampling from a parametrized distribution whose parameters are estimated from the privacy-preserving data. This turns out to be problematic. For example, in the second method, we are approximating T_{n_0} ~Multinomial$(n_0, \vec{\theta})$ as a random variable T'_{n_0} ~Multinomial$(n_0, \tilde{\theta})$ distribution. For relative guarantees, we would be interested in statements such as if $d_W(T_{n_0}, T'_{n_0}) < \delta$, then how does it compare to $d_W(M(T_{n_0}), M(T'_{n_0}))$ and $d_W\left(\phi_{n_0}(M(T_{n_0})), \phi_{n_0}(M(T'_{n_0}))\right)$?

The difficulty is that the randomness in M is correlated with T'_{n_0} (because its distribution is, by definition, Multinomial$(n_0, \tilde{\theta})$ and $\tilde{\theta}$ was estimated from $M(T_{n_0})$), so that our prior results do not apply. That is, our results are of the form $d(X,Y) \leq \delta \Rightarrow d(M(X), M(Y)) \leq \delta$ when M and Y are independent. We leave open the problem of relative guarantees under these types of correlation. However, we do note that there is a workaround that we discuss in the next section.

6.4. **Alternative Chi-squared Test of Independence with Input Perturbation.** One alternative to the approach in the previous section is to divide the privacy budget into two pieces: $\rho = \rho_1 + \rho_2$. Using ρ_1 we estimate the Multinomial parameter as follows. First we obtain row marginals $R[i] = T_{n_0}[i,\bullet] + N(0, 2/\rho_1)$ and column marginals $C[j] = T_{n_0}[\bullet, j] + N(0, 2/\rho_1)$. We normalize the row and column marginals so that they add up to n_0

and are nonnegative. Then we set $\tilde{\theta}[i,j] = \frac{R[i]}{n_0}\frac{C[j]}{n_0}$. Then we obtain noisy table counts using ρ_2 of the privacy budget: $\tilde{T}_{n_0} = M(T_{n_0}) = T_{n_0} + N(\vec{0}, (1/\rho_2)\mathbf{I})$. The test statistic remains the same: $\phi_{n_0}(\tilde{T}_{n_0}) \equiv \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(\tilde{T}_{n_0}[i,j]-\tilde{E}[i,j])^2}{\tilde{E}[i,j]}$ with $\tilde{E}[i,j] = \tilde{T}_{n_0}[i,\bullet]\tilde{T}_{n_0}[\bullet,j]/\tilde{T}_{n_0}[\bullet,\bullet]$.

Let T'_{n_0} be a random variable following the Multinomial$(n_0, \tilde{\theta})$ distribution. Now $\tilde{\theta}$ is independent of M. To obtain relative guarantees, we note that the postprocessing (i.e. computation of ϕ_{n_0} from its input) is a rational function in which the numerator is a quartic polynomial and the denominator is a quadratic polynomial of the input to ϕ_{n_0}. Our strategy will be to define a secondary test statistic ϕ'_{n_0} as $\phi'_{n_0}(\tilde{T}_{n_0}) \equiv \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(\tilde{T}_{n_0}[i,j]-\tilde{E}[i,j])^2}{n_0\vec{\theta}[i,j]}$, which uses the true (unknown) parameter in the denominator and then show that closeness between the distributions of T_{n_0} and T'_{n_0} $\Rightarrow$ closeness between $M(T_{n_0})$ and $M(T'_{n_0})$ $\Rightarrow$ closeness between $\phi'_{n_0}(M(T_{n_0}))$ and $\phi'_{n_0}(M(T'_{n_0}))$ $\Rightarrow$ closeness between $\phi_{n_0}(M(T_{n_0}))$ and $\phi_{n_0}(M(T'_{n_0}))$ with high probability.

Since ϕ' is now just a quartic polynomial in its input, following the discussion in Section 6.1, we can define H_ℓ to be the set of bounded functions that are ℓ-times continuously differentiable with bounded derivatives and G_ℓ to be the quartic polynomials. One starting point is to require the stronger convergence condition:

$$d_{H_\ell G_\ell}(T_{n_0}, T'_{n_0}) \equiv \sup_{h\in H_\ell, g\in G_\ell} E[h(g(T_{n_0}))] - E[h(g(T'_{n_0}))] \leq \delta$$

(note that establishing such type of multivariate convergence results is a current area of research [Gaunt, 2015]). Lemma 6.2 then implies that

$$d_{H_\ell G_\ell}(M(T_{n_0}), M(T'_{n_0})) \leq \delta$$

and

$$d_{H_\ell}(\phi'_{n_0}(M(T_{n_0})), \phi'_{n_0}(M(T'_{n_0}))) \leq \delta.$$

Now note that under the null hypothesis of independence between rows and columns, $\tilde{E}[i,j]/n_0$ converges to $\vec{\theta}[i,j]$ almost surely so that there exist small constants $\gamma_{n_0} > 0$ and $\beta_{n_0} > 0$ such that $\tilde{E}[i,j]/n_0 \in [(1-\gamma_{n_0})\vec{\theta}[i,j], (1+\gamma_{n_0})\vec{\theta}[i,j]]$ for all i,j with probability at least $1-\beta_{n_0}$. Thus, assuming $\vec{\theta}[i,j] > 0$ for all i,j, the ratio $\phi'_{n_0}(\tilde{T}_{n_0})/\phi_{n_0}(\tilde{T}_{n_0})$ is bounded by $1\pm\gamma_{n_0}$ with probability at least $1-\beta_{n_0}$. Thus the following result will establish closeness between $\phi_{n_0}(M(T_{n_0}))$ and $\phi_{n_0}(M(T'_{n_0}))$ with high probability (i.e. conditioned on the high probability event that $\tilde{E}[i,j]/n_0$ is close to $\vec{\theta}[i,j]$).

Lemma 6.5. *Let H be a subset of the 1-Lipschitz continuous functions and let X and Y be random variables and ϕ' be a positive function such that $d_H(\phi'(X), \phi'(Y)) \equiv \sup_{h\in H} E[h(\phi'(X))] - E[h(\phi'(Y))] \leq \delta$ and $\max(E[|\phi'(X)|], E[|\phi'(Y)|]) \leq \mu$ for some μ. Let ϕ be another function with the same domain as ϕ'. For any $\gamma \in (0,1)$, define $B_\gamma = \left\{\vec{t} \,:\, \phi(\vec{t})/\phi'(\vec{t}) \in [1-\gamma, 1+\gamma]\right\}$ and let $\beta_1 \geq \max(P(X \notin B_\gamma),\ P(Y \notin B_\gamma))$ and $\beta_2 \geq \max(E[|\phi'(X)|1_{[X\notin B_\gamma]}], E[|\phi'(Y)|1_{[Y\notin B_\gamma]}])$. Then when we condition on X and Y being in B_γ, $d_H(\phi(X) \mid B_\gamma,\ \phi(Y) \mid B_\gamma) \leq \frac{2\gamma\mu}{1-\beta_1} + \frac{\delta}{1-\beta_1} + \frac{2\beta_2}{1-\beta_1}$.*

The proof of Lemma 6.5 is in Appendix A.8.

6.5. **Kolmogorov-Smirnov Test with ϵ-Differential Privacy.** In the one sample Kolmogorov-Smirnov test, we have i.i.d. samples $x_1, \cdots, x_{n_0}$ from some unknown distribution F, a known continuous distribution F_0 and would like to test the hypothesis $F = F_0$.

Define an empirical c.d.f. by $F_{n_0}(x) = \frac{1}{n_0}\sum_{i=1}^{n_0} I(x_i \leq x)$, then in the non-private world, the test statistic is computed as $\sqrt{n_0}\sup_x |F_{n_0}(x) - F(x)|$. Under the null hypothesis $F = F_0$, the test statistic converges in distribution to the Kolmogorov-Smirnov distribution when n_0 goes to infinity. The c.d.f. of KS distribution is $H(t) = 1 - 2\sum_{i=1}^{\infty}(-1)^{i-1}e^{-2i^2t}$. We then compute a p-value. A small p-value (e.g., < 0.01) indicates strong evidence against the null hypothesis.

Under the private setting, we let $S_{n_0} = \sup_x |F_{n_0}(x) - F(x)|$. A change in the value of one record increases or decreases $F_{n_0}(x)$ (for any x) by at most $1/n_0$, so the sensitivity of S_{n_0} is $2/n_0$ since the maximum could be achieved at a different value of x. To achieve ϵ-differential privacy, a privacy mechanism M could add a random variable Y from the distribution $\text{Lap}(2/(n_0\epsilon))$ to S_{n_0} to obtain $\tilde{S}_{n_0}$, as justified by Lemma 3.3. The statistic we are interested in is $\phi_{n_0}(S_{n_0}) = \sqrt{n_0}S_{n_0}$. We then define $h_n(S_n) = \sqrt{n}S_n$ to indicate we are interested in the Kolmogorov-Smirnov approximation.

To compute the asymptotic approximation to $\phi_{n_0}(\tilde{S}_{n_0})$, in our proposed scheme, we take the limit of $\phi_{n_0}(M(h_{n_0}^{-1}(h_n(S_n))))$. We break the computation into simple steps:

- $h_n(S_n) = \sqrt{n}\sup_x |F_n(x) - F(x)|$.
- $h_{n_0}^{-1}(h_n(S_n)) = \frac{\sqrt{n}}{\sqrt{n_0}}\sup_x |F_n(x) - F(x)|$.
- $M(h_{n_0}^{-1}(h_n(S_n))) = \frac{\sqrt{n}}{\sqrt{n_0}}\sup_x |F_n(x) - F(x)| + Y$.
- $\phi_{n_0}(M(h_{n_0}^{-1}(h_n(S_n)))) = \sqrt{n}\sup_x |F_n(x) - F(x)| + \sqrt{n_0}Y$.

Now, taking the limit as $n \to \infty$, we get

$$\lim_{n\to\infty} \phi_{n_0}(M(h_{n_0}^{-1}(h_n(S_n)))) = \lim_{n\to\infty} \sqrt{n}\sup_x |F_n(x) - F(x)| + \sqrt{n_0}Y = K + \sqrt{n_0}Y$$

where K is a Kolmogorov-Smirnov random variable and Y is a $\text{Lap}(2/(n_0\epsilon))$ random variable.

As with all output perturbation methods, Corollary 5.3 and Lemma 5.6 show that if the sampling distribution of the test statistic (under the null hypothesis) is well approximated (either in KS distance or Wasserstein distance) by the KS distribution, then the noisy test statistic is approximated just as well by the approximation proposed by the recipe.

7. Conclusions

In this paper we studied a simple recipe for approximating the distribution of a statistic that is perturbed for privacy reasons. For the approximating distributions, we studied relative convergence guarantees of the form: if the non-private sampling distribution is well approximated by a classical statistical distribution, then how well (relative to the non-private case) is the sampling distribution of the privacy-preserving statistic approximated by the distribution proposed by the recipe?

In general, output perturbation privacy mechanisms preserve distances between distributions. However, input perturbation methods followed by nonlinear postprocessing require stronger convergence guarantees between the underlying data and their classical statistical approximations. Multivariate convergence results of this form are an active area of research.

References

J. Acharya, Z. Sun, and H. Zhang. Differentially private testing of identity and closeness of discrete distributions. https://arxiv.org/abs/1707.05128, 2017.

J. Awan and A. Slavković. Differentially private uniformly most powerful tests for binomial data. In *NIPS*, 2018.

A. C. Berry. The accuracy of the gaussian approximation to the sum of independent variates. *Transactions of the American Mathematical Society*, 49(1):122–136, 1941.

M. Bun and T. Steinke. Concentrated differential privacy: Simplifications, extensions, and lower bounds. In *TCC*, 2016.

B. Cai, C. Daskalakis, and G. Kamath. Priv'IT: private and sample efficient identity testing. In *ICML*, 2017.

A.-S. Charest. How can we analyze differentially-private synthetic datasets? *Journal of Privacy and Confidentiality*, 2(2), 2011.

K. Chaudhuri and D. Hsu. Convergence rates for differentially private statistical estimation. In *ICML*, 2012.

Y. Chen, A. Machanavajjhala, J. P. Reiter, and A. F. Barrientos. Differentially private regression diagnostics. In *IEEE 16th International Conference on Data Mining (ICDM)*, 2016.

V. D'Orazio, J. Honaker, and G. King. Differential privacy for social science inference. Sloan Foundation Economics Research Paper No. 2676160. Available at SSRN: http://ssrn.com/abstract=2676160, 2015.

J. C. Duchi, M. I. Jordan, and M. J. Wainwright. Minimax optimal procedures for locally private estimation. *Journal of the American Statistical Association*, 113(521):182–201, 2018.

C. Dwork and J. Lei. Differential privacy and robust statistics. In *STOC*, 2009.

C. Dwork, F. McSherry, K. Nissim, and A. Smith. Calibrating noise to sensitivity in private data analysis. In *TCC*, 2006.

C. Dwork, W. Su, and L. Zhang. Private false discovery rate control. *arXiv:1511.03803*, 2015.

D. Edwards. On the kantorovich–rubinstein theorem. *Expositiones Mathematicae*, 29(4):387 – 398, 2011.

W. Feller. *An Introduction to Probability Theory and its Applications*, volume 2. John Wiley & Sons, 2nd edition, 1971.

T. S. Ferguson. *A Course in Large Sample Theory*. Chapman & Hall, 1996.

S. E. Fienberg, A. Rinaldo, and X. Yang. Differential privacy and the risk-utility tradeoff for multi-dimensional contingency tables. In *PSD*, 2010.

M. Gaboardi, H. W. Lim, R. Rogers, and S. Vadhan. Differentially private chi-squared hypothesis testing: Goodness of fit and independence testing. In *ICML*, 2016.

R. E. Gaunt. Stein's method for functions of multivariate normal random variables. arXiv:1507.08688, 2015.

H. Goldstein and N. Shlomo. A probabilistic procedure for anonymisation and analysis of perturbed datasets. http://hummedia.manchester.ac.uk/institutes/cmist/archive-publications/working-papers/2018/A%20Probabilistic%20Procedure%20for%20Anonymisation%20and%20Analysis%20of%20Perturbed%20Datasets.pdf, 2018.

A. Johnson and V. Shmatikov. Privacy-preserving data exploration in genome-wide association studies. In *KDD*, 2013.

V. Karwa. Differentially private statistical inference and hypothesis testing. In *JSM*, 2016.

V. Karwa and A. Slavkovic. Inference using noisy degrees: Differentially private β-model and synthetic graphs. *The Annals of Statistics*, 44(1):87–112, 2016.

V. Karwa and S. Vadhan. Finite sample differentially private confidence intervals. In *LIPIcs-Leibniz International Proceedings in Informatics*, volume 94. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2018.

D. Kifer and A. Machanavajjhala. Pufferfish: A framework for mathematical privacy definitions. *ACM Trans. Database Syst.*, 39(1):3:1–3:36, 2014.

A. Machanavajjhala and D. Kifer. Designing statistical privacy for your data. *Commun. ACM*, 58(3):58–67, 2015.

S. Polettini and S. Arima. Small area estimation with covariates perturbed for disclosure limitation. *Statistica*, 75(1):57–72, 2015.

Y. Rinott, C. M. O'Keefe, N. Shlomo, C. Skinner, et al. Confidentiality and differential privacy in the dissemination of frequency tables. *Statistical Science*, 33(3):358–385, 2018.

R. Rogers and D. Kifer. A new class of private chi-square hypothesis tests. In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, (AISTATS)*, 2017.

N. Ross. Fundamentals of stein's method. *Probability Surveys*, 8:210—293, 2011.

O. Sheffet. Differentially private ordinary least squares: t-values, confidence intervals and rejecting null-hypotheses. http://arxiv.org/abs/1507.02482, 2015.

A. Smith. Privacy-preserving statistical estimation with optimal convergence rates. In *STOC*, 2011.

E. Solea. Differentially private hypothesis testing for normal random variables. Master's thesis, Penn State University, 2014.

C. Stein. A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, volume 2, pages 583–602. University of California Press, 1972.

C. Uhler, A. Slavkovic, and S. E. Fienberg. Privacy-preserving data sharing for genome-wide association studies. *Journal of Privacy and Confidentiality*, 5(1), 2013.

D. Vu and A. Slavkovic. Differential privacy for clinical trial data: Preliminary evaluations. In *ICDM Workshops*, 2009.

Y. Wang, J. Lee, and D. Kifer. Differentially private hypothesis testing, revisited. *arXiv:1511.03376*, 2015.

L. Wasserman and S. Zhou. A statistical framework for differential privacy. *Journal of the American Statistical Association*, 105(489):375–389, 2010.

Y. M. J. Woo and A. Slavkovic. Generalised linear models with variables subject to post randomization method. *Statistica Applicata-Italian Journal of Applied Statistics*, 24(1): 29–56, 2015.

F. Yu, S. E. Fienberg, A. B. Slavković, and C. Uhler. Scalable privacy-preserving data sharing methodology for genome-wide association studies. *Journal of biomedical informatics*, 50: 133–141, 2014.

Appendix A. Proofs

A.1. Proof of Theorem 5.1.

Theorem 5.1. *Let $\mathbf{X}$ and $\mathbf{Y}$ be two random variables using a common measure μ, so that $F_{\mathbf{X}}(\vec{t}) = \int_{\vec{x} \preceq \vec{t}} f_{\mathbf{X}}(\vec{x})\ d\mu(\vec{x})$ and $F_{\mathbf{Y}}(\vec{y}) = \int_{\vec{y} \preceq \vec{t}} f_{\mathbf{Y}}(\vec{y})\ d\mu(\vec{y})$. Let $M : \mathbb{R}^k \to \mathbb{R}^\ell$ be a randomized function associated with conditional probability $g(\cdot \mid \cdot)$ (so that $P(M(\vec{x}) \in S) = \int_{\vec{z} \in S} g(\vec{z} \mid \vec{x})\ d\nu(\vec{z})$). Then*

$$d_{KS}(M(\mathbf{X}), M(\mathbf{Y})) \leq 2 d_V(\mathbf{X}, \mathbf{Y})$$

where the probability is over the randomness in the function M and the random variables $\mathbf{X}$ and $\mathbf{Y}$.

Proof.

$$\begin{aligned}
&|P(M(\mathbf{X}) \preceq \vec{t}) - P(M(\mathbf{Y}) \preceq \vec{t})| \\
=&\left| \int_{\vec{z} \preceq \vec{t}} \int_{\vec{x} \in \mathbb{R}^k} g(\vec{z} \mid \vec{x})\, (f_{\mathbf{X}}(\vec{x}) - f_{\mathbf{Y}}(\vec{x}))\ d\mu(\vec{x})\ d\nu(\vec{z}) \right| \\
\leq& \int_{\vec{z} \preceq \vec{t}} \int_{\vec{x} \in \mathbb{R}^k} g(\vec{z} \mid \vec{x}) \Big| (f_{\mathbf{X}}(\vec{x}) - f_{\mathbf{Y}}(\vec{x})) \Big|\ d\mu(\vec{x})\ d\nu(\vec{z}) \\
=& \int_{\vec{x} \in \mathbb{R}^k} \left(\int_{\vec{z} \preceq \vec{t}} g(\vec{z} \mid \vec{x})\ d\nu(\vec{z}) \right) \Big| (f_{\mathbf{X}}(\vec{x}) - f_{\mathbf{Y}}(\vec{x})) \Big|\ d\mu(\vec{x}) \\
\leq& \int_{\vec{x} \in \mathbb{R}^k} \Big| (f_{\mathbf{X}}(\vec{x}) - f_{\mathbf{Y}}(\vec{x})) \Big|\ d\mu(\vec{x}) \\
=& 2 d_V(\mathbf{X}, \mathbf{Y}).
\end{aligned}$$

Taking the supremum over all $\vec{t}$ yields the result. □

A.2. Proof of Theorem 5.2.

To prove this theorem, first we need an intermediate result (a special case of multidimensional integration by parts).

Theorem A.1 . *Let μ and ν be two σ-finite nonnegative measures over $\mathbb{R}^k$ and let f and g be functions such that $\int_{\mathbb{R}^k} |f(\vec{x})|\ d\mu(\vec{x}) < \infty$ and $\int_{\mathbb{R}^k} |g(\vec{s})|\ d\nu(\vec{s}) < \infty$.[8] Let $F(\vec{t}) = \int_{\vec{x} \preceq \vec{t}} f(\vec{x})\ d\mu(\vec{x})$ and $\overline{G}(\vec{t}) = \int_{\vec{s} \succeq \vec{t}} g(\vec{s})\ d\nu(\vec{s})$. Then*

$$\int_{\mathbb{R}^k} f(\vec{x}) \overline{G}(\vec{x})\ d\mu(\vec{x}) = \int_{\mathbb{R}^k} F(\vec{s}) g(\vec{s})\ d\nu(\vec{s}).$$

Proof. Using Fubini's theorem to switch order of integration (since f and g are absolutely integrable),

$$\begin{aligned}
&\int_{\mathbb{R}^k} f(\vec{x}) \overline{G}(\vec{x})\ d\mu(\vec{x}) \\
=& \int_{\mathbb{R}^k} f(\vec{x}) \left(\int_{\vec{s} \succeq \vec{x}} g(\vec{s})\ d\nu(\vec{s}) \right)\ d\mu(\vec{x})
\end{aligned}$$

[8] Note that this includes counting measures, so if ν is a point mass on elements of some countable set $S \subset \mathbb{R}^k$ then $\int_{\mathbb{R}^k} |g(\vec{s})|\ d\nu(\vec{s}) = \sum_{\vec{s} \in S} |g(\vec{s})|$. This notation allows us to combine proofs for discrete and continuous cases.

$$
\begin{aligned}
&= \int_{\mathbb{R}^k} \int_{\mathbb{R}^k} \mathbb{1}_{\{\vec{s} \succeq \vec{x}\}} f(\vec{x}) g(\vec{s}) \; d\nu(\vec{s}) \; d\mu(\vec{x}) \\
&= \int_{\mathbb{R}^k} \left(\int_{\mathbb{R}^k} \mathbb{1}_{\{\vec{s} \succeq \vec{x}\}} f(\vec{x}) g(\vec{s}) \; d\mu(\vec{x}) \right) \; d\nu(\vec{s}) \\
&= \int_{\mathbb{R}^k} \left(\int_{\vec{s} \succeq \vec{x}} f(\vec{x}) \; d\mu(\vec{x}) \right) g(\vec{s}) \; d\nu(\vec{s}) \\
&= \int_{\mathbb{R}^k} F(\vec{s}) g(\vec{s}) \; d\nu(\vec{s}).
\end{aligned}
$$

□

Theorem 5.2. *Let $\mathcal{L} \subseteq \mathbb{R}^k$. Let $\mathbf{X}$ and $\mathbf{Y}$ be random variables over $\mathbb{R}^k$ whose cumulative distribution functions can be written as $\int_{\vec{x} \preceq \vec{t}} f_{\mathbf{X}}(\vec{x}) \; d\mu_1(\vec{x})$ and $\int_{\vec{y} \preceq \vec{t}} f_{\mathbf{Y}}(\vec{y}) \; d\mu_2(\vec{y})$, respectively.*[9] *Let $M : \mathbb{R}^k \to \mathbb{R}^\ell$ be a randomized function associated with conditional probability $g(\cdot \mid \cdot)$ satisfying the following conditions:*

- *Denote $H(\vec{t} \mid \vec{x}) = \int_{\vec{z} \preceq \vec{t}} g(\vec{z} \mid \vec{x}) \; d\nu(\vec{z})$, where ν is the measure used to integrate over the output space of M, and suppose there is a function $\bar{g}^\dagger$ such that $H(\vec{t} \mid \vec{x}) = \int_{\vec{s} \succeq \vec{x}} \bar{g}^\dagger(\vec{t} \mid \vec{s}) \; d\mu_3(\vec{s})$ for some measure μ_3.*
- *$\int_{\vec{s} \in \mathbb{R}^k} \mathbb{1}_{\{\vec{s} \notin \mathcal{L}\}} \; d\mu_3(\vec{s}) = 0$.*

Then for all $\vec{t} \in \mathbb{R}^\ell$:

$$d_{KS}(M(\mathbf{X}), M(\mathbf{Y})) \leq d_{KS(\mathcal{L})}(\mathbf{X}, \mathbf{Y}) \sup_{\vec{t} \in \mathcal{L}} ||\bar{g}^\dagger(\vec{t} \mid \cdot)||_1,$$

where $||\bar{g}^\dagger(\vec{t} \mid \cdot)||_1 = \int_{\vec{s} \in \mathbb{R}^k} |\bar{g}^\dagger(\vec{t} \mid \vec{s})| \; d\mu_3(\vec{s})$.

Proof. Pick a $\vec{t} \in \mathbb{R}^\ell$. With two applications of Theorem A.1 in the third equality below,

$$
\begin{aligned}
&\left| P(M(\mathbf{X}) \preceq \vec{t}) - P(M(\mathbf{Y}) \preceq \vec{t}) \right| \\
=&\left| \int_{\vec{z} \preceq \vec{t}} \int_{\vec{x} \in \mathbb{R}^k} f_{\mathbf{X}}(\vec{x}) g(\vec{z}|\vec{x}) d\mu_1(\vec{x}) d\nu(\vec{z}) - \int_{\vec{z} \preceq \vec{t}} \int_{\vec{y} \in \mathbb{R}^k} f_{\mathbf{Y}}(\vec{y}) g(\vec{z}|\vec{y}) d\mu_2(\vec{y}) d\nu(\vec{z}) \right| \\
=&\left| \int\limits_{\vec{x} \in \mathbb{R}^k} f_{\mathbf{X}}(\vec{x}) H(\vec{t}|\vec{x}) d\mu_1(\vec{x}) - \int\limits_{\vec{y} \in \mathbb{R}^k} f_{\mathbf{Y}}(\vec{y}) H(\vec{t}|\vec{y}) d\mu_2(\vec{y}) \right| \\
=&\left| \int\limits_{\vec{s} \in \mathbb{R}^k} F_{\mathbf{X}}(\vec{s}) \bar{g}^\dagger(\vec{t}|\vec{s}) d\mu_3(\vec{s}) - \int\limits_{\vec{s} \in \mathbb{R}^k} F_{\mathbf{Y}}(\vec{s}) \bar{g}^\dagger(\vec{t}|\vec{s}) d\mu_3(\vec{s}) \right| \\
\leq& \int_{\vec{s} \in \mathbb{R}^k} |F_{\mathbf{X}}(\vec{s}) - F_{\mathbf{Y}}(\vec{s})| * \left| \bar{g}^\dagger(\vec{t}|\vec{s}) \right| \; d\mu_3(\vec{s}) \\
=& \int_{\vec{s} \in \mathcal{L}} |F_{\mathbf{X}}(\vec{s}) - F_{\mathbf{Y}}(\vec{s})| * \left| \bar{g}^\dagger(\vec{t}|\vec{s}) \right| \; d\mu_3(\vec{s}) \\
\leq& d_{KS(\mathcal{L})}(\mathbf{X}, \mathbf{Y}) * ||\bar{g}^\dagger(\vec{t} \mid \cdot)||_1 \\
\leq& d_{KS(\mathcal{L})}(\mathbf{X}, \mathbf{Y}) \sup_{\vec{t} \in \mathcal{L}} ||\bar{g}^\dagger(\vec{t} \mid \cdot)||_1.
\end{aligned}
$$

[9] Note that μ_1 and μ_2 can be different measures, so, for example $\mathbf{X}$ can be discrete and $\mathbf{Y}$ can be continuous.

□

A.3. **Proof of Corollary 5.3.**

Corollary 5.3. *Using the notation of Theorem 5.2, if the conditional CDF $g(\cdot \mid \cdot)$ of M is translation invariant (i.e. $g(\vec{z}+\vec{t} \mid \vec{x}+\vec{t}) = g(\vec{z} \mid \vec{x})$) and uses the same base measure as $\mathbf{Y}$ (i.e. $P(M(\vec{x}) \preceq \vec{t}) = \int_{\vec{z} \preceq \vec{t}} g(\vec{z} \mid \vec{x})\ d\mu_2(\vec{z})$). Then*

$$d_{KS(\mathcal{L})}(M(\mathbf{X}), M(\mathbf{Y})) \leq d_{KS(\mathcal{L})}(\mathbf{X}, \mathbf{Y}).$$

Proof. The complication is that $\mathbf{X}$ and $\mathbf{Y}$ may be defined with respect to different measures. For example, in the case of the Central Limit Theorem for Binomial random variables, $\mathbf{X}$ would be discrete and $\mathbf{Y}$ would be continuous. We will use Theorem 5.2. Note that due to translation invariance,

$$\begin{aligned}
H(\vec{t} \mid \vec{x}) &= \int_{\vec{z} \prec \vec{t}} g(\vec{z} \mid \vec{x})\ d\mu_2(\vec{z}) \\
&= \int_{\vec{z} \preceq \vec{t}} g(\vec{z} - \vec{x} \mid \vec{0})\ d\mu_2(\vec{z}) \\
&= \int_{\vec{y} \preceq \vec{t} - \vec{x}} g(\vec{y} \mid \vec{0})\ d\mu_2(\vec{y}) \\
&= \int_{\vec{w} \preceq -\vec{x}} g(\vec{w} + \vec{t} \mid \vec{0})\ d\mu_2(\vec{w}) \\
&= \int_{-\vec{w} \succeq \vec{x}} g(\vec{w} + \vec{t} \mid \vec{0})\ d\mu_2(\vec{w}) \\
&= -\int_{\vec{v} \succeq \vec{x}} g(-\vec{v} + \vec{t} \mid \vec{0})\ d\mu_2(\vec{v}) \\
&= -\int_{\vec{v} \succeq \vec{x}} g(\vec{t} \mid \vec{v})\ d\mu_2(\vec{v})
\end{aligned}$$

so the desired $\bar{g}^\dagger(\vec{t} \mid \vec{s})$ is $-g(\vec{t} \mid \vec{s})$. Next note that

$$\begin{aligned}
||\bar{g}^\dagger(\vec{t} \mid \cdot)||_1 &= \int |\bar{g}^\dagger(\vec{t} \mid \vec{s})|\ d\mu(\vec{s}) \\
&= \int g(\vec{t} \mid \vec{s})\ d\mu(\vec{s}) \\
&= \int g(\vec{t} - \vec{s} \mid \vec{0})\ d\mu(\vec{s}) \\
&= 1.
\end{aligned}$$

□

A.4. **Proof of Lemma 5.6.**

Lemma 5.6. *Let M be a mechanism that satisfies ϵ-differential privacy and that works as follows. On input $\mathbf{X} \in \mathbb{R}^k$, M samples a finite-dimensional random vector $\mathbf{B}$ from a distribution μ_b and then returns $f(\mathbf{X}, \mathbf{B})$ for some function $f : \mathbb{R}^k \times \mathbb{R}^b \to \mathbb{R}^\ell$. Let $\mathbf{X}$ and $\mathbf{Y}$ be random vectors in $\mathbb{R}^k$ having distributions μ_1 and μ_2, respectively. If all of the following conditions are satisfied:*

- *f is L-Lipschitz continuous in its first argument (i.e. $\mathbf{X}$)*
- *$E[||f(\vec{0}, \mathbf{B})||_2] < \infty$*
- *$E[||\mathbf{X}||_2] < \infty$*
- *$E[||\mathbf{Y}||_2] < \infty$*

then the following holds: $d_W(\mathbf{X}, \mathbf{Y}) \leq \delta \Rightarrow d_W(M(\mathbf{X}), M(\mathbf{Y})) \leq L\delta$.

Proof. First consider the function $h(\vec{x}) \equiv E_{\mathbf{B}\sim\mu_b}[g(f(\vec{x}, \mathbf{B}))]$, where g is 1-Lipschitz continuous. The following calculations show that h is finite and L-Lipschitz continuous (for the finite part, set $\vec{y} = \vec{0}$):

$$\begin{aligned} L||\vec{x} - \vec{y}||_2 &= E_{\mathbf{B}\sim\mu_b}[L||\vec{x} - \vec{y}||_2] \\ &\geq E_{\mathbf{B}\sim\mu_b}[||f(\vec{x}, \mathbf{B}) - f(\vec{y}, \mathbf{B})||_2] \\ &\geq E_{\mathbf{B}\sim\mu_b}[||g(f(\vec{x}, \mathbf{B})) - g(f(\vec{y}, \mathbf{B}))||_2] \\ &\geq ||E_{\mathbf{B}\sim\mu_b}[g(f(\vec{x}, \mathbf{B})) - g(f(\vec{y}, \mathbf{B}))]||_2 \\ &= ||h(\vec{x}) - h(\vec{y})||_2. \end{aligned}$$

Now, for any $\alpha > 0$, let Ω^ℓ_α be the set of α-Lipschitz continuous functions from $\mathbb{R}^\ell \to \mathbb{R}$.

$$\begin{aligned} d_W(M(\mathbf{X}), M(\mathbf{Y})) &= \sup_{g\in\Omega^\ell_1} E[g(M(\mathbf{X}))] - E[g(M(\mathbf{Y}))] \\ &= \sup_{g\in\Omega^\ell_1} E[g(f(\mathbf{X}, \mathbf{B}))] - E[g(f(\mathbf{Y}, \mathbf{B}))] \\ &= \sup_{g\in\Omega^\ell_1} E_{\mathbf{X}\sim\mu_1}\Big[E_{\mathbf{B}\sim\mu_b}[g(f(\mathbf{X}, \mathbf{B}))]\Big] - E_{\mathbf{Y}\sim\mu_2}\Big[E_{\mathbf{B}\sim\mu_b}[g(f(\mathbf{Y}, \mathbf{B}))]\Big] \\ &\leq \sup_{h\in\Omega^\ell_L} E_{\mathbf{X}\sim\mu_1}\Big[h(\mathbf{X})\Big] - E_{\mathbf{Y}\sim\mu_2}\Big[h(\mathbf{Y})\Big] \\ &\leq L\delta. \end{aligned}$$

□

A.5. **Proof of Theorem 5.9.**

Theorem 5.9. *Let $\mathbf{X} = (X_1, \ldots, X_k)$ and $\mathbf{Y} = (Y_1, \ldots, Y_k)$ be k-dimensional random variables and suppose $d_{KS(\mathcal{L})}(\mathbf{X}, \mathbf{Y}) \leq \delta$ then:*

1. *If ϕ is the coordinate projection operator that selects a fixed subset of the components of a vector (e.g., $\phi(\vec{t}) = (t_2, t_4, t_5)$) then $d_{KS(\mathcal{L})}(\phi(\mathbf{X}), \phi(\mathbf{Y})) \leq \delta$.*
2. *If ϕ is the sum of coordinates (i.e. $\phi(\vec{t}) = t_1 + t_2 + \cdots + t_k$) then there exist $\mathbf{X}$ and $\mathbf{Y}$ such that $d_{KS(\mathcal{L})}(\mathbf{X}, \mathbf{Y}) \leq \delta$ but $d_{KS(\mathcal{L})}(\phi(\mathbf{X}), \phi(\mathbf{Y})) = 1$.*
3. *If $\phi(\vec{t}) \equiv c\vec{t} + \vec{a}$, where c is a scalar, then $d_{KS(\mathcal{L})}(\phi(\mathbf{X}), \phi(\mathbf{Y})) \leq \delta$.*

4. *Even if ϕ is continuous, is one-to-one, and preserves partial orders (i.e. $\vec{t} \preceq \vec{s} \Rightarrow \phi(\vec{t}) \preceq \phi(\vec{s})$) then there still exist $\mathbf{X}$ and $\mathbf{Y}$ such that $d_{KS(\mathcal{L})}(\mathbf{X}, \mathbf{Y}) \leq \delta$ but $d_{KS(\mathcal{L})}(\phi(\mathbf{X}), \phi(\mathbf{Y})) = 1$.*

Proof. For (1), pick a $\vec{t} = (t_1, \ldots, t_k)$ and define the vector $\vec{s} = (s_1, \ldots, s_k)$ such that $s_i = t_i$ when i is one of the coordinates selected by ϕ and $s_i = \infty$ otherwise. Then $P(\phi(\mathbf{X}) \preceq \phi(\vec{t})) = P(\mathbf{X} \preceq \vec{s})$ and the conclusion follows.
For (2), we construct a pair $\mathbf{X}$ and $\mathbf{Y}$ for which the distance grows large. let $n = \lceil 1/\delta \rceil$ so that $1/n \leq \delta$. Define a two-dimensional random variable $\mathbf{Y}$: the support of $\mathbf{Y}$ is the set of pairs $(i + \frac{1}{4}, n - i)$ for $i = 1, \ldots, n$ and $\mathbf{Y}$ is uniformly distributed on its support. Define $\mathbf{X}$ as follows: the support of $\mathbf{X}$ is $(i, n - i)$ for $i = 1, \ldots, n$ and $\mathbf{X}$ is uniformly distributed on its support. There are several facts to note:

- Rectangles are convex, so if a rectangle contains the points $(i, n - i)$ and $(j, n - j)$ then it also contains the points $(\ell, n - \ell)$ for $\ell = i, \ldots, j$.
- If a rectangle contains the points $(i + \frac{1}{4}, n - i)$ and $(j + \frac{1}{4}, n - j)$ then it also contains the points $(\ell + \frac{1}{4}, n - \ell)$ for $\ell = i, \ldots, j$.
- If a rectangle contains $(i, n - i)$ and $(i + 1, n - (i + 1))$ then it contains $(i + \frac{1}{4})$.
- If a rectangle contains $(i+\frac{1}{4}, n-i)$ and $(i+1+\frac{1}{4}, n-(i+1))$ then it contains $(i+1, n-(i+1))$.

These facts mean that if a rectangle contains r points from the domain of $\mathbf{X}$ it also contains either $r - 1$, r, or $r + 1$ points from the domain of $\mathbf{Y}$. Since the probability of any point in the domain of its random variable is $1/n \leq \delta$, this means $\sup_{\vec{t}} |P(\mathbf{X} \preceq \vec{t}) - P(\mathbf{Y} \preceq \vec{t})| = \delta$. However, $\phi(\mathbf{X}) = n$ and $\phi(\mathbf{Y}) = n + \frac{1}{4}$ so $|P(\phi(\mathbf{X}) \leq n)| = 1$ while $|P(\phi(\mathbf{Y})) \leq n| = 0$.
For (3) note that $\phi(\vec{t}) \preceq \phi(\vec{s}) \Leftrightarrow \vec{t} \preceq \vec{s}$ so the result follows trivially.
For (4), let's define $\phi((t_1, \ldots, t_n)) \equiv (t_1, \ldots, t_{n-1}, \sum_i t_i)$. From part (1) we know that if $d_{KV}(\phi(\mathbf{X}), \phi(\mathbf{Y})) \leq \delta^*$, then projection on the last coordinate preserves distance but this projection equals the sum of the components of $\mathbf{X}$ and $\mathbf{Y}$. Thus if distributions of the sums of the components of $\mathbf{X}$ and $\mathbf{Y}$ have a distance of 1 then so does $d_{KV}(\phi(\mathbf{X}),\ \phi(\mathbf{Y}))$. It is trivial to extend the construction in part (2) to this case. □

A.6. **Proof of Lemma 6.2.**

Lemma 6.2. *Let M be a mechanism that adds some random variable $\mathbf{Z}$ to its input. Suppose that $\phi \in G_\ell$ and that G_ℓ is closed under translation of its input (i.e., if $g \in G$ then for every c, the function $x \to g(x + c)$ belongs to G_ℓ) then:*

$$d_{H_\ell G_\ell}(\mathbf{X}, \mathbf{Y}) \leq \delta \Rightarrow d_{H_\ell G_\ell}(M(\mathbf{X}), M(\mathbf{Y})) \leq \delta \Rightarrow d_{H_\ell}(\phi(M(\mathbf{X})), \phi(M(\mathbf{Y}))) \leq \delta.$$

Proof. The last implication, $d_{H_\ell G_\ell}(M(\mathbf{X}), M(\mathbf{Y})) \leq \delta \Rightarrow d_{H_\ell}(\phi(M(\mathbf{X})), \phi(M(\mathbf{Y}))) \leq \delta$ follows immediately from the requirement that $\phi \in G_\ell$. The first implication can be proved as follows. Let $\mathbf{Z}_1$ and $\mathbf{Z}_2$ be two independent random variables having the same distribution as $\mathbf{Z}$. Assume $\mathbf{Z}$ has density p without loss of generality (i.e. replace densities with probability mass functions for discrete variables, etc). For any given g and constant c, define g_c to be the function $g_c(x) = g(x + c)$. Now,

$$\begin{aligned} E[h(g(M(\mathbf{X})))] - E[h(g(M(\mathbf{Y})))] &= E[h(g(\mathbf{X} + \mathbf{Z}_1))] - E[h(g(\mathbf{Y} + \mathbf{Z}_2))] \\ &= \int E[h(g(\mathbf{X} + t))]\ p(t)dt - \int E[h(g(\mathbf{Y} + t))]\ p(t)dt \end{aligned}$$

$$
\begin{aligned}
&= \int \left(E[h(g_t(\mathbf{X}))] - E[h(g_t(\mathbf{Y}))]\right) p(t)\ dt \\
&\leq \sup_t E[h(g_t(\mathbf{X}))] - E[h(g_t(\mathbf{Y}))] \\
&\leq \sup_{h\in H_\ell} \sup_{g\in G_\ell} E[h(g(\mathbf{X}))] - E[h(g(\mathbf{Y}))] \\
&\leq d_{H_\ell G_\ell}(\mathbf{X}, \mathbf{Y}) \leq \delta.
\end{aligned}
$$

□

A.7. **Proof of Lemma 6.4.**

Lemma 6.4. *$Sen_1(\phi_{n_0}) = \frac{1}{n_0}\max_{u,v:u\neq v}\left(\frac{2n_0-1}{\vec{\theta}[u]} + \frac{2n_0+1}{\vec{\theta}[v]}\right)$ for the chi-squared goodness of fit test statistic ϕ_{n_0}.*

Proof. Without loss of generality, assume the value 1 from the vector $\vec{x}_i$ changes from index u to index v ($u \neq v$), and thus $\mathbf{S}_{n_0}[u] \geq 1$. The L_1 sensitivity can be computed as

$$
\begin{aligned}
&Sen_1(\phi_{n_0}) \\
=&\max_{u,v}\left|\frac{(\mathbf{S}_{n_0}[u] - n_0\vec{\theta}[u])^2}{n_0\vec{\theta}[u]} - \frac{(\mathbf{S}_{n_0}[u] - 1 - n_0\vec{\theta}[u])^2}{n_0\vec{\theta}[u]} + \frac{(\mathbf{S}_{n_0}[v] - n_0\vec{\theta}[v])^2}{n_0\vec{\theta}[v]} - \frac{(\mathbf{S}_{n_0}[v] + 1 - n_0\vec{\theta}[v])^2}{n_0\vec{\theta}[v]}\right| \\
=&\max_{u,v}\left|\frac{2(\mathbf{S}_{n_0}[u] - n_0\vec{\theta}[u]) - 1}{n_0\vec{\theta}[u]} + \frac{2(n_0\vec{\theta}[v] - \mathbf{S}_{n_0}[v]) - 1}{n_0\vec{\theta}[v]}\right| \\
=&\frac{1}{n_0}\max_{u,v}\left|\frac{2\mathbf{S}_{n_0}[u] - 1}{\vec{\theta}[u]} - \frac{2\mathbf{S}_{n_0}[v] + 1}{\vec{\theta}[v]}\right| \\
\leq&\frac{1}{n_0}\max_{u,v}\left|\frac{2\mathbf{S}_{n_0}[u] - 1}{\vec{\theta}[u]}\right| + \left|\frac{2\mathbf{S}_{n_0}[v] + 1}{\vec{\theta}[v]}\right| \\
\leq&\frac{1}{n_0}\max_{u,v}\left(\frac{2n_0 - 1}{\vec{\theta}[u]} + \frac{2n_0 + 1}{\vec{\theta}[v]}\right).
\end{aligned}
$$

□

A.8. **Proof of Lemma 6.5.**

Lemma 6.5. *Let H be a subset of the 1-Lipschitz continuous functions and let X and Y be random variables and ϕ' be a positive function such that $d_H(\phi'(X), \phi'(Y)) \equiv \sup_{h\in H} E[h(\phi'(X))] - E[h(\phi'(Y))] \leq \delta$ and $\max(E[|\phi'(X)|], E[|\phi'(Y)|]) \leq \mu$ for some μ. Let ϕ be another function with the same domain as ϕ'. For any $\gamma \in (0,1)$, define $B_\gamma = \left\{\vec{t} \ :\ \phi(\vec{t})/\phi'(\vec{t}) \in [1-\gamma, 1+\gamma]\right\}$ and let $\beta_1 \geq \max(P(X \notin B_\gamma),\ P(Y \notin B_\gamma))$ and $\beta_2 \geq \max(E[|\phi'(X)|1_{[X\notin B_\gamma]}], E[|\phi'(Y)|1_{[Y\notin B_\gamma]}])$. Then when we condition on X and Y being in B_γ, $d_H(\phi(X) \mid B_\gamma,\ \phi(Y) \mid B_\gamma) \leq \frac{2\gamma\mu}{1-\beta_1} + \frac{\delta}{1-\beta_1} + \frac{2\beta_2}{1-\beta_1}$.*

Proof.

$$\begin{aligned} d_H(\phi(X) \mid B_\gamma,\ \phi(Y) \mid B_\gamma) \leq & d_H(\phi(X) \mid B_\gamma,\ \phi'(X) \mid B_\gamma) \\ & + d_H(\phi'(X) \mid B_\gamma,\ \phi'(Y) \mid B_\gamma) \\ & + d_H(\phi(Y) \mid B_\gamma,\ \phi'(Y) \mid B_\gamma). \end{aligned}$$

Now we consider the first term $d_H(\phi(X) \mid B_\gamma,\ \phi'(X) \mid B_\gamma)$. Because we condition on B_γ, the ratio of ϕ and ϕ' is bounded by $1 \pm \gamma$. Using Lipschitz continuity of h, $\Big|E[h(\phi(X)) \mid B_\gamma] - E[h(\phi'(X)) \mid B_\gamma]\Big| \leq E[|h(\phi(X)) - h(\phi'(X)) \mid B_\gamma] \leq E\Big[|\phi(X) - \phi'(X)| \mid B_\gamma\Big] \leq \gamma E[|\phi'(X)| \mid B_\gamma] \leq \frac{\gamma}{1-\beta_1} E[|\phi'(X)|] \leq \frac{\gamma\mu}{1-\beta_1}$.

By similar reasoning, the third term is also bounded by $\frac{\gamma\mu}{1-\beta_1}$.

For the second term, for any $h \in H$, consider $E[h(\phi'(X)) \mid B_\gamma] - E[h(\phi'(Y)) \mid B_\gamma]$. Without loss of generality, due to the subtraction, we may assume $h(0) = 0$ and so due to 1-Lipschitz continuity, $h(x) \leq |x|$.

$$\begin{aligned} \Big|E[h(\phi'(X)) \mid B_\gamma] - E[h(\phi'(Y)) \mid B_\gamma]\Big| \leq & \left|\frac{1}{1-\beta_1} E[h(\phi'(X)) 1_{[X \in B_\gamma]}] - E[h(\phi'(Y)) 1_{[Y \in B_\gamma]}]\right| \\ \leq & \left|\frac{1}{1-\beta_1} E[h(\phi'(X))] - E[h(\phi'(Y))]\right| \\ & + \left|\frac{1}{1-\beta_1} E[h(\phi'(X)) 1_{[X \notin B_\gamma]}] - E[h(\phi'(Y)) 1_{[Y \notin B_\gamma]}]\right| \\ \leq & \frac{\delta}{1-\beta_1} + \frac{1}{1-\beta_1} \left(E[|\phi'(X)| 1_{[X \notin B_\gamma]}] + E[|\phi'(Y)| 1_{[X \notin B_\gamma]}]\right) \\ \leq & \frac{\delta}{1-\beta_1} + \frac{2\beta_2}{1-\beta_1}. \end{aligned}$$

Thus it follows that $d_H(\phi(X) \mid B_\gamma,\ \phi(Y) \mid B_\gamma) \leq \frac{2\gamma\mu}{1-\beta_1} + \frac{\delta}{1-\beta_1} + \frac{2\beta_2}{1-\beta_1}$. □