Local Private Hypothesis Testing: Chi-Square Tests

Marco Gaboardi [1] **Ryan Rogers** [2]

Abstract

The local model for differential privacy is emerging as the reference model for practical applications of collecting and sharing sensitive information while satisfying strong privacy guarantees. In the local model, there is no *trusted* entity which is allowed to have each individual's raw data as is assumed in the traditional *curator* model. Individuals' data are usually perturbed before sharing them. We explore the design of private hypothesis tests in the local model, where each data entry is perturbed to ensure the privacy of each participant. Specifically, we analyze locally private chi-square tests for goodness of fit and independence testing.

1. Introduction

Hypothesis testing is a widely applied statistical tool used to test whether given models should be rejected, or not, based on sampled data from a population. Hypothesis testing was initially developed for scientific and survey data, but today it is also an essential tool to test models over collections of social network, mobile, and crowdsourced data (American Statistical Association, 2014; Hunter et al., 2008; Steele et al., 2017). Collected data samples may contain highly sensitive information about the subjects, and the privacy of individuals can be compromised when the results of a data analysis are released. A way to address this concern is by developing new techniques to support privacy-preserving data analysis. Among the different approaches, differential privacy (Dwork et al., 2006b) has emerged as a viable solution: it provides strong privacy guarantees and it allows to release accurate statistics. A standard way to achieve differential privacy is by injecting some statistical noise in the computation of the data analysis. When the noise is carefully chosen, it helps to protect the individual privacy without compromising the utility of the data analysis. Several recent works have studied differentially private hypothesis tests that can be used in place of the standard, non-private hypothesis tests (Uhler et al., 2013; Yu et al., 2014; Sheffet, 2015; Karwa & Slavković, 2016; Wang et al., 2015; Gaboardi et al., 2016; Kifer & Rogers, 2017; Cai et al., 2017). These tests work in the *curator model* of differential privacy. In this model, the data is centrally stored and the curator carefully injects noise in the computation of the data analysis in order to satisfy differential privacy.

In this work we instead address the *local model* of privacy, formally introduced by Raskhodnikova et al. (2008). The first differentially private algorithm called *randomized response* – in fact it predates the definition of differential privacy by more than 40 years – guarantees differential privacy in the local model (Warner, 1965). In this model, there is no *trusted* centralized entity that is responsible for the noise injection. Instead, each individual adds enough noise to guarantee differential privacy for their own data, which provides a stronger privacy guarantee than the curator model. The data analysis is then run over the collection of the individually sanitized data. The local model of differential privacy is a convenient model for several applications: for example it is used to collect statistics about the activity of the Google Chrome Web browser users (Erlingsson et al., 2014), and to collect statistics about the typing patterns of Apple's iPhone users (Apple Press Info, 2016). Despite these applications, the local model has received far less attention than the centralized curator model. This is in part due to the more firm requirements imposed by this model, which make the design of effective data analysis harder.

Our main contribution is in designing chi-square hypothesis tests for the local model of differential privacy. Similar to previous works we focus on goodness of fit and independence hypothesis tests. Most of the private chi-square tests proposed so far are based on mechanisms that add noise in some form to the aggregate data, e.g. the cells of the contingency tables, or the resulting chi-square statistics value. These approaches cannot be used in the local model, since noise needs to be added at the individual's data level. We then consider instead general privatizing techniques in the local model, and we study how to build new hypothesis tests with them. Each test we present is characterized by a specific local model mechanism. The main technical challenge for designing each test is to create statistics, which incorporate the local model mechanisms, that converge as

[*] Equal contribution [1] University at Buffalo, Buffalo, NY, USA [2] University of Pennsylvania, Philadelphia, PA, USA. Correspondence to: Marco Gaboardi <gaboardi@buffalo.edu>.

Proceedings of the 35^{th} International Conference on Machine Learning, Stockholm, Sweden, PMLR 80, 2018. Copyright 2018 by the author(s).

we collect more data to a chi-square distribution, as in the classical chi-square tests. We then use these statistics to find the critical value to correctly bound the Type I error.

We present three different goodness of fit tests: `LocalNoiseGOF` presents a statistic that guarantees the convergence to a chi-square distribution under the null hypothesis so that we can use the correct critical values when local (concentrated) differential privacy is guaranteed by adding Laplace or Gaussian noise to the individual data; `LocalGenRRGOF` also provides a statistic that converges to a chi-square under the null hypothesis when a private value for each individual is selected by using a generalized form of randomized response, which can also be thought of as an instantiation of the exponential mechanism (McSherry & Talwar, 2007); finally, `LocalBitFlipGOF` introduces a statistic that converges to a chi-square distribution when the data is privatized using a bit flipping algorithm (Bassily & Smith, 2015), which provide better accuracy for higher dimensional data. Further, we develop corresponding independence tests: `LocalNoiseIND` (see supplementary file), `LocalGenRRIND`, and `LocalBitFlipIND`. For all these tests we study their asymptotic behavior. A desiderata for private hypothesis tests is to have a guaranteed upper bound on the probability of a false discovery (or Type I error) – rejecting a null hypothesis or model when the data was actually generated from it – and to minimize the probability of a Type II error, which is failing to reject the null hypothesis when the model is indeed false. This latter criteria corresponds to the *power* of the statistical test. We then present experimental results showing the power of the different tests which demonstrates that no single local differentially private algorithm is best across all data dimensions and privacy parameter regimes. However, this evaluation also shows a relation between the power of the test and the noncentral parameter of the test statistic that is used. This suggests that besides looking at the parameters of the test, a data analyst may need also to consider which test statistic results in the largest noncentral parameter.

2. Related Works

There have been several works in developing private hypothesis test for categorical data, but all look at the traditional model of (concentrated) differential privacy instead of the local model, which we consider here. Several works have explored private statistical inference for GWAS data, (Uhler et al., 2013; Yu et al., 2014; Johnson & Shmatikov, 2013). Following these works, there has also been general work in private chi-square hypothesis tests, where the main tests are for goodness of fit and independence testing, although some do extend to more general tests (Wang et al., 2015; Gaboardi et al., 2016; Kifer & Rogers, 2017; Cai et al., 2017; Kakizaki et al., 2017). Among these, the works most related to ours are the ones by Gaboardi et al. (2016); Kifer & Rogers (2017). One of our mechanisms, `LocalNoiseGOF`, can be seen as an adaptation of their techniques to the local model. However, the other mechanisms we introduce differ substantially and require novel asymptotic analyses. There has also been work in private hypothesis testing for ordinary least squares regression (Sheffet, 2015).

Duchi et al. (2013b;a) focus on controlling disclosure risk in statistical estimation and inference by ensuring the analysis satisfies local differential privacy. In their work, they show that a generalized version of randomized response gives optimal sample complexity for estimating the multinomial probability vector. We use this idea as the basis for our hypothesis test `LocalBitFlipGOF`. Kairouz et al. (2014) also considers hypothesis testing in the local model, although they measure utility in terms of f-divergences and do not give a decision rule, i.e. when to reject a given null hypothesis. We provide statistics whose distributions asymptotically follow a chi-square distribution, which allows for approximating statistical p-values that can be used in a decision rule. We consider their *extremal* mechanisms and empirically confirm their result that for small privacy regimes (small ϵ) one mechanism has higher *utility* than other mechanisms and for large privacy regimes (large ϵ) a different mechanism outperforms the other. However, we measure utility in terms of the power of a locally private hypothesis test subject to a given Type I error bound. Other notable works in the local privacy model include Pastore & Gastpar (2016); Kairouz et al. (2016); Ye & Barg (2017)

Independent of this work, another paper (Sheffet, 2018) has addressed local private hypothesis testing. Sheffet (2018) considers finite sample complexity by showing certain test quantities take different values under the null- and alternative-hypothesis. In this work, we design and analyze asymptotic statistical tests and empirically evaluate the performance of each test for finite samples.

3. Preliminaries

We consider datasets $\boldsymbol{x} = (x_1, \cdots, x_n) \in \mathcal{X}^n$ in some data universe $\mathcal{X}$, typically $\mathcal{X} = \{0,1\}^d$ where d is the dimensionality. We first present the standard definition of differential privacy, as well as its variant *concentrated differential privacy*. We say that two datasets $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}^n$ are *neighboring* if they differ in at most one element, i.e. $\exists i \in [n]$ such that $x_i \neq x_i'$ and $\forall j \neq i$, $x_j = x_j'$.

Definition 3.1 (Dwork et al. (2006b;a)). *An algorithm $\mathcal{M} : \mathcal{X}^n \to \mathcal{Y}$ is (ϵ, δ)-differentially private (DP) if for all neighboring datasets $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}^n$ and for all outcomes $S \subseteq \mathcal{Y}$, we have* $\Pr[\mathcal{M}(\boldsymbol{x}) \in S] \leq e^{\epsilon} \Pr[\mathcal{M}(\boldsymbol{x}') \in S] + \delta$.

Definition 3.2 (Bun & Steinke (2016)). *An algorithm $\mathcal{M} : \mathcal{X}^n \to \mathcal{Y}$ is ρ-zero-mean concentrated differentially private (zCDP) if for all neighboring datasets*

$\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}^n$, *we have the following bound for all* $t > 0$ *where the expectation is over outcomes* $y \sim \mathcal{M}(\boldsymbol{x})$, $\mathbb{E}\left[\exp\left(t\left(\ln\left(\frac{\Pr[\mathcal{M}(\boldsymbol{x})=y]}{\Pr[\mathcal{M}(\boldsymbol{x}')=y]}\right) - \rho\right)\right)\right] \leq e^{t^2\rho}$.

Note that in both of these privacy definitions, it is assumed that all the data is stored in a central location and the algorithm $\mathcal{M}$ can access all the data. Most of the work in differential privacy has been in this *trusted curator model*. We then define *local* differential privacy, formalized by Raskhodnikova et al. (2008) and Dwork & Roth (2014), which does not require the subjects to release their raw data, rather each data entry is perturbed to prevent the true entry from being stored. Thus, local differential privacy ensures one of the strongest privacy guarantees.

Definition 3.3 (LR Oracle)**.** *Given a dataset* $\boldsymbol{x}$*, a* local randomizer oracle $LR_{\boldsymbol{x}}(\cdot,\cdot)$ *takes as input an index* $i \in [n]$ *and an* ϵ*-DP algorithm* R*, and outputs* $y \in \mathcal{Y}$ *chosen according to the distribution of* $R(x_i)$*, i.e.* $LR_{\boldsymbol{x}}(i, R) = R(x_i)$*.*

Definition 3.4 (Raskhodnikova et al. (2008))**.** *An algorithm* $\mathcal{M} : \mathcal{X}^n \to \mathcal{Y}$ *is* (ϵ, δ)-local differentially private *(LDP) if it accesses the input database* $\boldsymbol{x}$ *via the LR oracle* $LR_{\boldsymbol{x}}$ *with the following restriction: if* $LR(i, R_j)$ *for* $j \in [k]$ *are* $\mathcal{M}$*'s invocations of* $LR_{\boldsymbol{x}}$ *on index* i*, then each* R_j *for* $j \in [k]$ *is* (ϵ_j, δ_j)*- DP and* $\sum_{j=1}^k \epsilon_j \leq \epsilon$*,* $\sum_{j=1}^k \delta_j \leq \delta$*.*

From this we have that a (ϵ, δ)-LDP algorithm is also (ϵ, δ)-DP. Note that these definitions can be extended to include ρ-local zCDP (LzCDP) where each local randomizer is ρ_j-zCDP and $\sum_{j=1}^k \rho_j \leq \rho$. We point out the following connection between $LzCDP$ and LDP, which follows directly from results in (Bun & Steinke, 2016)

Lemma 3.5. *If* $\mathcal{M} : \mathcal{X}^n \to \mathcal{Y}$ *is* $(\epsilon, 0)$*-LDP then it is also* $\epsilon^2/2$*-LzCDP. If* $\mathcal{M}$ *is* ρ*-LzCDP, then it is also* $\left(\left(\rho + \sqrt{2\rho\ln(2/\delta)}\right), \delta\right)$*-LDP for any* $\delta > 0$*.*

4. Chi-Square Hypothesis Tests

As was studied in (Gaboardi et al., 2016), (Wang et al., 2015), and (Kifer & Rogers, 2017), we will study hypothesis tests with categorical data. A null hypothesis, or model H_0 is how we might expect the data to be generated. The goal for hypothesis testing is to reject the null hypothesis if the data is not likely to have been generated from the given model. As is common in statistical inference, we want to design hypothesis tests to bound the probability of a false discovery (or Type I error), i.e. rejecting a null hypothesis when the data was actually generated from it, by at most some amount α, such as 5%. However, designing tests that achieve this is easy, because we can just ignore the data and always *fail to reject* the null hypothesis, i.e. have an inconclusive test. Thus, we want additionally to design our tests so that they can reject H_0 if the data was not actually generated from the given model. We then want to minimize the probability of a Type II error, which is failing to reject H_0 when the model is false, subject to a given Type I error.

For goodness of fit testing, we assume that each individual's data $\boldsymbol{X}_i$ for $i \in [n]$ is sampled i.i.d. from $\mathrm{Multinomial}(1, \boldsymbol{p})$ where $\boldsymbol{p} \in \mathbb{R}^d_{>0}$ and $\boldsymbol{p}^\intercal \cdot \boldsymbol{1} = 1$. The classical chi-square hypothesis test (without privacy) forms the histogram $\boldsymbol{H} = (H_1, \cdots, H_d) = \sum_{i=1}^n \boldsymbol{X}_i$ and computes the chi-square statistic $\mathrm{T} = \sum_{j=1}^d \frac{\left(H_j - np_j^0\right)^2}{np_j^0}$. The reason for using this statistic is that it converges in distribution to χ^2_{d-1} as more data is collected, i.e. $n \to \infty$, when $\mathrm{H}_0 : \boldsymbol{p} = \boldsymbol{p}^0$ holds. Hence, we can ensure the probability of false discovery to be close to α as long as we only reject H_0 when $\mathrm{T} > \chi^2_{d-1,1-\alpha}$ where the *critical value* $\chi^2_{d-1,1-\alpha}$ is defined as the following quantity $\Pr\left[\chi^2_{d-1} > \chi^2_{d-1,1-\alpha}\right] = \alpha$.

Prior Private Chi-square Tests in the Curator Model. One approach for chi-square private hypothesis tests is to add noise (Gaussian or Laplace) directly to the histogram to ensure privacy and then use the classical test statistic (Gaboardi et al., 2016; Wang et al., 2015) . Note that the resulting asymptotic distribution needs to be modified for such changes to the statistic – it is no longer a chi-square random variable. To introduce the different statistics, we will consider goodness of fit testing after adding noise $\boldsymbol{Z}$ from distribution $\mathcal{D}^n$ to the histogram of counts $\widetilde{\boldsymbol{H}} = \boldsymbol{H} + \boldsymbol{Z}$, which ensures ρ-zCDP when $\mathcal{D} = \mathrm{N}(0, 1/\rho)$ and ϵ-DP when $\mathcal{D} = \mathrm{Lap}(2/\epsilon)$. The chi-square statistic then becomes

$$\widetilde{\mathrm{T}}(\mathcal{D}) = \sum_{i=1}^d \frac{\left(H_i + Z_i - np_i^0\right)^2}{np_i^0} \quad \text{where } \boldsymbol{Z} \sim \mathcal{D}^n. \quad (1)$$

The previous works then show that this statistic converges in distribution to a linear combination of chi-squared variables, when $\mathcal{D} \sim \mathrm{N}(0, 1/\rho)$ and ρ is also decreasing with n.

Kifer & Rogers (2017) showed that modifying the chi-square statistic to account for the additional noise leads to tests with better empirical power. The *projected* statistic from Kifer & Rogers (2017) is the following where we use projection matrix $\Pi \overset{\text{defn}}{=} \left(I_d - \frac{1}{d}\boldsymbol{1}\boldsymbol{1}^\intercal\right)$, middle matrix $M_\sigma = \Pi\left(\mathrm{Diag}\left(\boldsymbol{p}^0 + \sigma\right) - \boldsymbol{p}^0\left(\boldsymbol{p}^0\right)^\intercal\right)^{-1}\Pi$, and sample noise $\boldsymbol{Z} \sim \mathcal{D}^n$, with $\widehat{\boldsymbol{H}} = \boldsymbol{H} + \boldsymbol{Z}$

$$\mathrm{T}^{(n)}_{KR}(\sigma; \mathcal{D}) = n\left(\frac{\widehat{\boldsymbol{H}}}{n} - \boldsymbol{p}^0\right)^\intercal M_\sigma \left(\frac{\widehat{\boldsymbol{H}}}{n} - \boldsymbol{p}^0\right) \quad (2)$$

We use $\mathcal{D} = \mathrm{Lap}(2/\epsilon)$ with $\sigma = \frac{8}{n\epsilon^2}$ for an ϵ-DP claim or $\mathcal{D} = \mathrm{N}(0, 1/\rho)$ with $\sigma = \frac{1}{n\rho}$ for a ρ-zCDP claim. When comparing the power of all our tests, we will be considering the alternate $\mathrm{H}_1 : \boldsymbol{p} = \boldsymbol{p}^1_n$ where $\boldsymbol{p}^1_n = \boldsymbol{p}^0 + \frac{\boldsymbol{\Delta}}{\sqrt{n}}$ where $\boldsymbol{1}^\intercal\boldsymbol{\Delta} = 0$.

Theorem 4.1 (Kifer & Rogers (2017)). *Under the null hypothesis* $\mathrm{H}_0 : \boldsymbol{p} = \boldsymbol{p}^0$*, the statistic* $\mathrm{T}^{(n)}_{KR}\left(\frac{1}{n\rho}; \mathrm{N}\left(0, 1/\rho\right)\right)$ *given in* (2) *for* $\rho > 0$ *converges in distribution to* χ^2_{d-1}*. Further, under the alternate hypothesis* $\mathrm{H}_1 : \boldsymbol{p} = \boldsymbol{p}^1_n$*, the resulting asymptotic distribution is a noncentral chi-square random variable with* $d-1$ *degrees of freedom and noncentral parameter* $\boldsymbol{\Delta}^\intercal \left(\mathrm{Diag}(\boldsymbol{p}^0) - \boldsymbol{p}^0 \left(\boldsymbol{p}^0\right)^\intercal + 1/\rho I_d\right)^{-1} \boldsymbol{\Delta}$

When $\mathcal{D} = \mathrm{Lap}(2/\epsilon)$, Gaboardi et al. (2016) showed that we can still obtain the null hypothesis distribution using Monte Carlo simulations to estimate the critical value, since the asymptotic distribution will no longer be chi-square. That is, we can obtain m samples from the statistic under the null hypothesis with Laplace noise added to the histogram of counts. We can then guarantee that the probability of a false discovery is at most α as long as $m > \lceil 1/\alpha \rceil$.

5. Local Private Goodness of Fit

We now turn to designing local private goodness of fit tests. We first show how the existing statistics from the previous section can be adapted to the local setting and then develop new tests based on the generalized randomized response mechanism that returns one of $d > 1$ categories and bit flipping (Bassily & Smith, 2015). Each test is locally private because it perturbs each individual's data through a local randomizer. However, each of them has a different asymptotic behavior and so we need different analyses to identify the different critical values. We empirically check the power of each test to see which tests outperform others in different parameter regimes. An interesting result of this analysis is that the power of a test is directly related to the size of the noncentral parameter of the chi-square statistic under the alternate distribution.

Testing with Noise Addition. In the local model we can add $\boldsymbol{Z}_i \sim \mathrm{N}\left(\boldsymbol{0}, \frac{1}{\rho} I_d\right)$ independent noise to each individual's data $\boldsymbol{X}_i$ to ensure ρ-LzCDP or $\boldsymbol{Z}_i \overset{i.i.d.}{\sim} \mathrm{Lap}\left(\frac{2}{\epsilon}\right)$ independent noise to $\boldsymbol{X}_i$ to ensure ϵ-LDP. In either case, the resulting noisy histogram $\widehat{\boldsymbol{H}} = \boldsymbol{H} + \boldsymbol{Z}$ where $\boldsymbol{Z} = \sum_i \boldsymbol{Z}_i$ will have variance that scales with n for fixed privacy parameters $\epsilon, \rho > 0$. Consider the case where we add Gaussian noise, which results in the following histogram, $\widehat{\boldsymbol{H}} = \boldsymbol{H} + \boldsymbol{Z}$ where $\boldsymbol{Z} \sim \mathrm{N}\left(\boldsymbol{0}, \frac{n}{\rho} I_d\right)$. Thus, we can use either statistic $\widetilde{\mathrm{T}}\left(\rho/n\right)$ or $\mathrm{T}^{(n)}_{KR}\left(\rho/n\right)$, with the latter statistic typically having better empirical power (Kifer & Rogers, 2017). We then give our first local private hypothesis test in Algorithm 1.

Theorem 5.1. `LocalNoiseGOF` *is* ρ*-LzCDP when* $\mathcal{D} = \mathrm{N}\left(0, 1/\rho\right)$ *and* ϵ*-LDP when* $\mathcal{D} = \mathrm{Lap}(2/\epsilon)$.

Algorithm 1 Locally Private GOF Test:`LocalNoiseGOF`

```
Input: $\boldsymbol{x} = (\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n)$, $\rho$, $\alpha$, $\mathrm{H}_0 : \boldsymbol{p} = \boldsymbol{p}^0$.
  Let $\boldsymbol{H} = \sum_{\ell=1}^n \boldsymbol{x}_\ell$
  if $\mathcal{D} = \mathrm{N}\left(0, n/\rho\right)$ then
    Set $q = \mathrm{T}^{(n)}_{KR}\left(n/\rho; \mathcal{D}\right)$ given in (2).
    if $q > \chi^2_{d-1,1-\alpha}$ Decision ← Reject.
    else Decision ← Fail to Reject.
  end if
  if $\mathcal{D} = \sum_{i=1}^n \mathrm{Lap}(2/\epsilon)$ then
    Set $q = \mathrm{T}^{(n)}_{KR}\left(8n/\epsilon^2; \mathcal{D}\right)$ given in (2).
    Sample $m > \lceil 1/\alpha \rceil$ from the distribution of
    $\mathrm{T}^{(n)}_{KR}\left(8n/\epsilon^2; \mathcal{D}\right)$ assuming $H_0$
    Set $\tau$ to be the $\lceil (m+1)(1-\alpha) \rceil$th largest sample.
    if $q > \tau$ Decision ← Reject.
    else Decision ← Fail to Reject.
  end if
Output: Decision
```

Although we cannot guarantee the probability of a Type I error at most α due to the fact that we use the asymptotic distribution (as in the tests from prior work and the classical chi-square tests without privacy), we expect the Type I errors to be similar to those from the nonprivate test. Note that the test can be modified to accommodate arbitrary noise distributions, e.g. Laplace to ensure differential privacy. In this case, we can use a Monte Carlo (MC) approach to estimate the critical value τ that ensures the probability of a Type I error is at most α if we reject H_0 when the statistic is larger than τ. For the local setting, if each individual perturbs each coordinate by adding $\mathrm{Lap}\left(2/\epsilon\right)$ then this will ensure our test is ϵ-LDP. However, the sum of independent Laplace random variables is not Laplace, so we will need to estimate a sum of n independent Laplace random variables using MC. We can do this by sampling m entries from the exact distribution under H_0 to find the critical value. In the experiments section we will use this method to compare the power of the other local private tests with the one of the version of `LocalNoiseGOF` using Laplace noise, which has a better power than the one using Gaussian noise.

Testing with Generalized Randomized Response. Rather than having to add noise to each component of the original histogram, we consider applying randomized response to obtain a LDP hypothesis test. We will use a generalized form of randomized response given in Algorithm 2 which takes a single data entry from the set $\{\boldsymbol{e}_1, \cdots, \boldsymbol{e}_d\}$, where $\boldsymbol{e}_j \in \mathbb{R}^d$ is the standard basis element with a 1 in the jth coordinate and is zero elsewhere, and reports the original entry with probability slightly more than uniform and otherwise reports a different element with equal probability. Note that $\mathcal{M}_{\texttt{GenRR}}$ is ϵ-DP.

We have the following result when we use $\mathcal{M}_{\texttt{GenRR}}$ on each data entry to obtain a private histogram.

Algorithm 2 Generalized Randomized Response: $\mathcal{M}_{\texttt{GenRR}}$

```
Input: $\boldsymbol{x} \in \{\boldsymbol{e}_1, \cdots, \boldsymbol{e}_d\}$, $\epsilon$.
  Let $q(\boldsymbol{x}, \boldsymbol{z}) = \mathbb{1}\{\boldsymbol{x} = \boldsymbol{z}\}$
  Select $\check{\boldsymbol{x}}$ with probability $\frac{\exp[\epsilon\, q(\boldsymbol{x}, \check{\boldsymbol{x}})]}{e^{\epsilon}-1+d}$
Output: $\check{\boldsymbol{x}}$
```

Lemma 5.2. *If we have histogram* $\boldsymbol{H} = \sum_{i=1}^n \boldsymbol{X}_i$, *where* $\{\boldsymbol{X}_i\} \overset{i.i.d.}{\sim} \text{Multinomial}(1, \boldsymbol{p})$ *and we write* $\check{\boldsymbol{H}} = \sum_{i=1}^n \mathcal{M}_{\texttt{GenRR}}(\boldsymbol{X}_i, \epsilon)$ *for each* $i \in [n]$, *then* $\check{\boldsymbol{H}} \sim \text{Multinomial}(n, \check{\boldsymbol{p}})$ *where*

$$\check{\boldsymbol{p}} = \boldsymbol{p}\left(\frac{e^{\epsilon}}{e^{\epsilon}+d-1}\right) + (1-\boldsymbol{p})\left(\frac{1}{e^{\epsilon}+d-1}\right). \quad (3)$$

Once we have $\check{\boldsymbol{H}}$, we can create a chi-square statistic by subtracting $\check{\boldsymbol{H}}$ by its expectation and dividing the difference by the expectation. Hence testing $\text{H}_0 : \boldsymbol{p} = \boldsymbol{p}^0$ after the generalized randomized response mechanism, is equivalent to testing $\text{H}_0 : \boldsymbol{p} = \check{\boldsymbol{p}}^0$ with data $\check{\boldsymbol{H}}$.

We can then form a chi-square statistic using the histogram $\check{\boldsymbol{H}}$ which will have the correct asymptotic distribution.

Theorem 5.3. *Let* $\boldsymbol{H} \sim \text{Multinomial}(n, \boldsymbol{p})$ *and* $\check{\boldsymbol{H}}$ *be given in Theorem 5.2 with privacy parameter* $\epsilon > 0$. *Under the null hypothesis* $\text{H}_0 : \boldsymbol{p} = \boldsymbol{p}^0$, *we have for* $\check{\boldsymbol{p}}^0 = \frac{1}{e^{\epsilon}+d-1}\left(e^{\epsilon}\boldsymbol{p}^0 + (1-\boldsymbol{p}^0)\right)$,

$$\text{T}^{(n)}_{\text{GenRR}}(\epsilon) = \sum_{j=1}^{d} \frac{(\check{H}_j - n\check{p}^0_j)^2}{n\check{p}^0_j} \xrightarrow{D} \chi^2_{d-1}. \quad (4)$$

Further, with alternate $\text{H}_1 : \boldsymbol{p} = \boldsymbol{p}^1_n$, *the resulting asymptotic distribution is a noncentral chi-square distribution with* $d-1$ *degrees of freedom and noncentral parameter,* $\left(\frac{e^{\epsilon}-1}{e^{\epsilon}+d-1}\right)^2 \sum_{j=1}^d \frac{\Delta_j^2}{\check{p}^0_j}$.

We then base our LDP goodness of fit test on this result to obtain the correct critical value to reject the null hypothesis based on a chi-square distribution. The test is presented in Algorithm 3. The following result is immediate from the generalized randomized response mechanism being ϵ-DP and the fact that we use it as a local randomizer.

Algorithm 3 Local DP GOF Test: `LocalGenRRGOF`

```
Input: $\boldsymbol{x} = (\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n)$, $\epsilon$, $\alpha$, $\text{H}_0 : \boldsymbol{p} = \boldsymbol{p}^0$.
  Let $\check{\boldsymbol{p}}^0 = \frac{1}{e^{\epsilon}+d-1}\left(e^{\epsilon}\boldsymbol{p}^0 + (1-\boldsymbol{p}^0)\right)$.
  Let $\check{\boldsymbol{H}} = \sum_{i=1}^n \mathcal{M}_{\texttt{GenRR}}(\boldsymbol{x}_i, \epsilon)$.
  Set $q = \sum_{j=1}^d \frac{(\check{H}_j - n\check{p}^0_j)^2}{n\check{p}^0_j}$
  if $q > \chi^2_{d-1,1-\alpha}$ Decision ← Reject.
  else Decision ← Fail to Reject.
Output: Decision
```

Theorem 5.4. `LocalGenRRGOF` *is* ϵ*-LDP.*

Testing with Bit Flipping. Note that the noncentral parameter in Theorem 5.3 goes to zero as d grows large due to the coefficient being $\left(\frac{e^{\epsilon}-1}{e^{\epsilon}+d-1}\right)^2$. Thus, for large dimensional data the generalized randomized response cannot reject a false null hypothesis. We next consider another differentially private algorithm $\mathcal{M} : \{\boldsymbol{e}_1, \cdots, \boldsymbol{e}_d\} \to \{0, 1\}^d$, given in Algorithm 4 used in (Bassily & Smith, 2015) that flips each bit with some biased probability. [1]

Algorithm 4 Bit Flip Local Randomizer: $\mathcal{M}_{\text{bit}}$

```
Input: $\boldsymbol{x} \in \{\boldsymbol{e}_1, \cdots, \boldsymbol{e}_d\}$, $\epsilon$.
  for $j \in [d]$ do
    Set $z_j = x_j$ with probability $\frac{e^{\epsilon/2}}{e^{\epsilon/2}+1}$, otherwise $z_j = (1 - x_j)$.
  end for
Output: $\boldsymbol{z}$
```

Theorem 5.5. *The algorithm* $\mathcal{M}_{bit}$ *is* ϵ*-DP.*

We then want to form a statistic based on the output $\boldsymbol{z} \in \{0, 1\}^d$ that is asymptotically distributed as a chi-square under the null hypothesis. We defer the proof to the supplementary material.

Lemma 5.6. *Consider* $\boldsymbol{X}_i \sim \text{Multinomial}(1, \boldsymbol{p})$ *for each* $i \in [n]$. *We define the following covariance matrix* $\Sigma(\boldsymbol{p})$ *and mean vector* $\widetilde{\boldsymbol{p}} = \frac{\left[(e^{\epsilon/2}-1)\boldsymbol{p}+1\right]}{e^{\epsilon/2}+1}$, *in terms of* $\alpha_\epsilon = \left(\frac{e^{\epsilon/2}-1}{e^{\epsilon/2}+1}\right)$

$$\Sigma(\boldsymbol{p}) = \alpha_\epsilon^2 \left[\text{Diag}(\boldsymbol{p}) - \boldsymbol{p}(\boldsymbol{p})^{\intercal}\right] + \frac{e^{\epsilon/2}}{\left(e^{\epsilon/2}+1\right)^2} I_d \quad (5)$$

The histogram $\widetilde{\boldsymbol{H}} = \sum_{i=1}^n \mathcal{M}_{bit}(\boldsymbol{X}_i)$ *has the following asymptotic distribution* $\sqrt{n}\left(\frac{\widetilde{\boldsymbol{H}}}{n} - \widetilde{\boldsymbol{p}}\right) \xrightarrow{D} \text{N}(\boldsymbol{0}, \Sigma(\boldsymbol{p}))$. *Further,* $\Sigma(\boldsymbol{p})$ *is invertible for any* $\epsilon > 0$ *and* $\boldsymbol{p} > \boldsymbol{0}$.

Following a similar analysis in (Kifer & Rogers, 2017), we can form the following statistic for null hypothesis $\text{H}_0 : \boldsymbol{p} = \boldsymbol{p}^0$ in terms of the histogram $\widetilde{\boldsymbol{H}}$ and projection matrix $\Pi = I_d - \frac{1}{d}\boldsymbol{1}\boldsymbol{1}^{\intercal}$, as well as the covariance $\Sigma = \Sigma\left(\boldsymbol{p}^0\right)$ and mean $\widetilde{\boldsymbol{p}}^0$ both given in (5) where we replace $\boldsymbol{p}$ with $\boldsymbol{p}^0$:

$$\text{T}^{(n)}_{\text{BitFlip}}(\epsilon) = n\left(\frac{\widetilde{\boldsymbol{H}}}{n} - \widetilde{\boldsymbol{p}}^0\right)^{\intercal} \Pi\Sigma^{-1}\Pi \left(\frac{\widetilde{\boldsymbol{H}}}{n} - \widetilde{\boldsymbol{p}}^0\right) \quad (6)$$

We can then design a hypothesis test based on the outputs from $\mathcal{M}_{\text{bit}}$ in Algorithm 5

Theorem 5.7. `LocalBitFlipGOF` *is* ϵ*-LDP.*

[1] Special thanks to Adam Smith for recommending to use this particular algorithm.

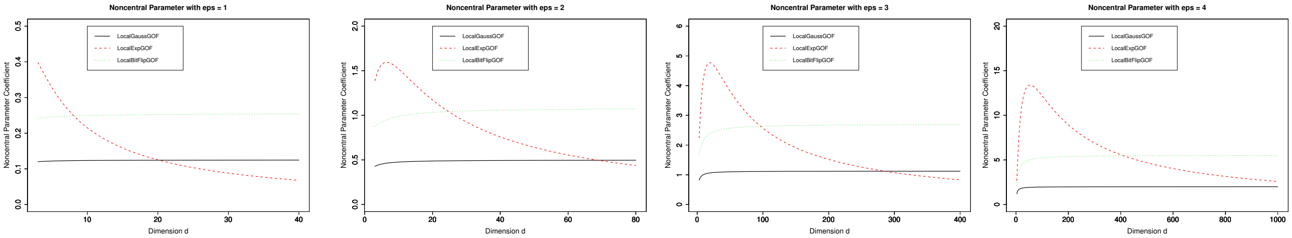

Figure 1: The coefficient on $\boldsymbol{\Delta}^\intercal\boldsymbol{\Delta}$ in the noncentral parameter for the local private tests where “LocalGaussGOF" is `LocalNoiseGOF` with Gaussian noise, `LocalGenRRGOF`, and `LocalBitFlipGOF` for various dimensions d and ϵ.

Algorithm 5 Local DP GOF Test: `LocalBitFlipGOF`

```
Input: $\boldsymbol{x} = (\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n)$, $\epsilon$, $\alpha$, $\mathrm{H}_0 : \boldsymbol{p} = \boldsymbol{p}^0$.
    Let $\widetilde{\boldsymbol{H}} = \sum_{i=1}^n \mathcal{M}_{\text{bit}}(\boldsymbol{x}_i, \epsilon)$.
    Set $q = \mathrm{T}^{(n)}_{\text{BitFlip}}(\epsilon)$
    if $q > \chi^2_{d-1,1-\alpha}$ Decision ← Reject.
    else Decision ← Fail to Reject.
Output: Decision
```

We now show that the statistic in (6) is asymptotically distributed as χ^2_{d-1}, with proof in the supplementary file.

Theorem 5.8. *If the null hypothesis* $\mathrm{H}_0 : \boldsymbol{p} = \boldsymbol{p}^0$ *holds, then the statistic* $\mathrm{T}^{(n)}_{\mathit{BitFlip}}(\epsilon)$ *is asymptotically distributed as a chi-square, i.e.* $\mathrm{T}^{(n)}_{\mathit{BitFlip}}(\epsilon) \xrightarrow{D} \chi^2_{d-1}$. *Further, if we consider the alternate* $\mathrm{H}_1 : \boldsymbol{p} = \boldsymbol{p}^1$ *then* $\mathrm{T}^{(n)}_{\mathit{BitFlip}}(\epsilon)$ *converges in distribution to a noncentral chi-square with* $d-1$ *degrees of freedom and noncentral parameter* $\left(\frac{e^{\epsilon/2}-1}{e^{\epsilon/2}+1}\right)^2 \cdot \boldsymbol{\Delta}^\intercal \Sigma(\boldsymbol{p}^0)^{-1}\boldsymbol{\Delta}$.

Comparison of Noncentral Parameters. We now compare the noncentral parameters of the three local private tests we presented in Algorithms 1, 3 and 5. We consider the null hypothesis $\boldsymbol{p}^0 = (1/d, \cdots, 1/d)$ for $d > 2$, and alternate $\mathrm{H}_1 : \boldsymbol{p} = \boldsymbol{p}^0 + \frac{\boldsymbol{\Delta}}{\sqrt{n}}$. In this case, we can easily compare the various noncentral parameters for various privacy parameters and dimensions d. In Figure 1 we give the coefficient to the term $\boldsymbol{\Delta}^\intercal\boldsymbol{\Delta}$ in the noncentral parameter of the asymptotic distribution for each local private test presented thus far. The larger this coefficient is, the better the power will be for any alternate $\boldsymbol{\Delta}$ vector. Note that in `LocalNoiseGOF`, we set $\rho = \epsilon^2/8$ which makes the variance the same as for a random variable distributed as $\mathrm{Lap}(2/\epsilon)$ for an ϵ-DP guarantee – recall that `LocalNoiseGOF` with Gaussian noise does not satisfy ϵ-DP for any $\epsilon > 0$. We give results for $\epsilon \in \{1, 2, 3, 4\}$ which are all in the range of privacy parameters that have been considered in actual locally differentially private algorithms used in practice.[2] From the plots, we see how `LocalGenRRGOF` may outperform `LocalBitFlipGOF` depending on the privacy parameter and dimension of the data. We can use these plots to determine which test to use given ϵ and the dimension of data d. When H_0 is not uniform, we can use the noncentral parameters given for each test to find the test with the largest noncentral parameter for a particular privacy budget ϵ.

Empirical Results. We then empirically compare the power between `LocalNoiseGOF` with Laplace noise in Algorithm 1, `LocalGenRRGOF` in Algorithm 3, and `LocalBitFlipGOF` in Algorithm 5. Recall that all three of these tests have the same privacy benchmark of local differential privacy. For `LocalNoiseGOF` with Laplace noise, we will use $m = 999$ samples in our Monte Carlo simulations. In our experiments we fix $\alpha = 0.05$ and $\epsilon \in \{1, 2, 4\}$. We then consider null hypotheses of the form $\boldsymbol{p}^0 = (1/d, 1/d, \cdots, 1/d)$ and alternate $\mathrm{H}_1 : \boldsymbol{p} = \boldsymbol{p}^0 + \eta(1, -1, \cdots, 1, -1)$ for some $\eta > 0$. In Figure 2, we plot the number of times our tests correctly rejects the null hypothesis in 1000 independent trials for various sample sizes n and privacy parameters ϵ. From Figure 2, we can see that the test statistics that have the largest noncentral parameter for a particular dimension d and privacy parameter ϵ will have the best empirical power. When $d = 4$, we see that `LocalGenRRGOF` performs the best. However, for $d = 40$ it is not so clear cut. When $\epsilon = 4$, we can see that `LocalGenRRGOF` does the best, but then when $\epsilon = 2$, `LocalBitFlipGOF` does best. Thus, the best Local DP Goodness of Fit test depends on the noncentral parameter, which is a function of ϵ, the null hypothesis $\boldsymbol{p}^0$, and alternate $\boldsymbol{p} = \boldsymbol{p}^0 + \boldsymbol{\Delta}$. Note that the worst local DP test also depends on the privacy parameter and the dimension d. Based on our empirical results, we see that no single locally private test is best for all data dimensions. However, knowing the corresponding noncentral parameter for a given problem is useful in determining which tests to use. Indeed, the larger the noncentral parameter is the higher the power will be.

[2] In (Erlingsson et al., 2014), we know that Google uses $\epsilon = \ln(3)$ in RAPPOR and from Aleksandra Korolova’s Twitter post on Sept. 13, 2016 `https://twitter.com/korolova/status/775801259504734208`, Apple uses $\epsilon = 1, 4$.

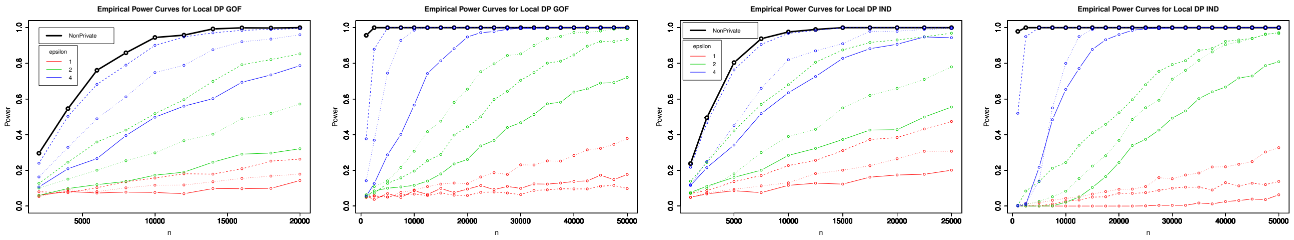

Figure 2: Plot 1 & 2: comparison of empirical power among the classical non-private test and the local private tests: `LocalNoiseGOF` with Laplace noise (solid line), `LocalGenRRGOF` (dashed line), and `LocalBitFlipGOF` (dotted line); in Plot 1 $d = 4$ and $\eta = 0.01$, in Plot 2 $d = 40$ and $\eta = 0.005$. Plot 3 & 4: comparison of empirical power among classical non-private test versus local private tests: adding Laplace noise (solid line) which is described in the supplementary file, `LocalGenRRIND` (dashed line), and `LocalBitFlipIND` (dotted line) for the contingency table data distribution given in (12) where Plot 3 $(r, c) = (2, 2)$ and $\eta = 0.01$, Plot 4 $(r, c) = (10, 4)$ and $\eta = 0.005$.

6. Local Private Independence Tests

Our techniques can be extended to include *composite* hypothesis tests, where we test whether the data comes from a whole family of probability distributions. We will focus on independence testing, but much of the theory can be extended to general chi-square tests. We will closely follow the presentation and notation as in (Kifer & Rogers, 2017).

We consider two multinomial random variables $\{\boldsymbol{U}_\ell\}_{\ell=1}^n \overset{i.i.d.}{\sim} \text{Multinomial}(1, \boldsymbol{\pi}^{(1)})$ for $\boldsymbol{\pi}^{(1)} \in \mathbb{R}^r$, $\{\boldsymbol{V}_\ell\}_{\ell=1}^n \overset{i.i.d.}{\sim} \text{Multinomial}(1, \boldsymbol{\pi}^{(2)})$ for $\boldsymbol{\pi}^{(2)} \in \mathbb{R}^c$ and no component of $\boldsymbol{\pi}^{(1)}$ or $\boldsymbol{\pi}^{(2)}$ is zero and each sums to 1. Without loss of generality, we will consider an individual to be in one of r groups who reports a data record that is in one of c categories. The collected data consists of n joint outcomes $\boldsymbol{H}$ whose (i, j)th coordinate is $H_{i,j} = \sum_{\ell=1}^n \mathbb{1}\{U_{\ell,i} = 1 \quad \& \quad V_{\ell,j} = 1\}$. Note that $\boldsymbol{H}$ is then the contingency table over the joint outcomes. Under the null hypothesis of independence between $\{\boldsymbol{U}_\ell\}_{\ell=1}^n$ and $\{\boldsymbol{V}_\ell\}_{\ell=1}^n$, for probability vector $\boldsymbol{p}(\boldsymbol{\pi}^{(1)}, \boldsymbol{\pi}^{(2)}) = \boldsymbol{\pi}^{(1)} \left(\boldsymbol{\pi}^{(2)}\right)^\intercal$, we have $\boldsymbol{H} \sim \text{Multinomial}\left(n, \boldsymbol{p}(\boldsymbol{\pi}^{(1)}, \boldsymbol{\pi}^{(2)})\right)$

What makes this test difficult is that the analyst does not know the data distribution $\boldsymbol{p}(\boldsymbol{\pi}^{(1)}, \boldsymbol{\pi}^{(2)})$ and so cannot simply plug it into the chi-square statistic. Rather, we use the data to estimate the *best guess* for the unknown probability distribution that satisfies the null hypothesis. Note that without privacy, each individual $\ell \in [n]$ is reporting a $r \times c$ matrix $\boldsymbol{X}_\ell$ which would be 1 in exactly one location. Thus we can alternatively write the contingency table as $\boldsymbol{H} = \sum_{\ell=1}^n \boldsymbol{X}_\ell$. We then use the three local private algorithms we presented earlier to see how we can form a private chi-square statistic for independence testing. We want to be able to ensure the privacy of both the group and the category that each individual belongs to. Due to space we will only cover private independence tests that use the generalized randomized response mechanism from Algorithm 2 and the bit flipping local randomizer from Algorithm 4. We defer our independence test with noise addition in the local setting to the supplementary file.

Testing with Generalized Randomized Response. We want to design an independence test when the data is generated from $\mathcal{M}_{\texttt{GenRR}}$ given in Algorithm 2. In this case our contingency table can be written as $\check{\boldsymbol{H}} \sim \text{Multinomial}\left(n, \check{\boldsymbol{p}}(\boldsymbol{\pi}^{(1)}, \boldsymbol{\pi}^{(2)})\right)$ where $\beta_\epsilon = \frac{1}{e^\epsilon + rc - 1}$ and we use (3) to get

$$\check{\boldsymbol{p}}(\boldsymbol{\pi}^{(1)}, \boldsymbol{\pi}^{(2)}) = \beta_\epsilon \left((e^\epsilon - 1)\, \boldsymbol{\pi}^{(1)} \left(\boldsymbol{\pi}^{(2)}\right)^\intercal + \mathbf{1} \right) \tag{7}$$

We then obtain an estimate for the unknown parameters,

$$\check{\boldsymbol{\pi}}^{(1)} = \frac{1}{\beta_\epsilon \left(e^\epsilon - 1\right)} \left(\frac{\check{H}_{i,\cdot}}{n} - c\beta_\epsilon : i \in [r] \right),$$

$$\check{\boldsymbol{\pi}}^{(2)} = \frac{1}{\beta_\epsilon \left(e^\epsilon - 1\right)} \left(\frac{\check{H}_{\cdot,j}}{n} - r\beta_\epsilon : j \in [c] \right)$$

$$\check{\mathcal{T}}^{(n)}_{\text{GenRR}}(\epsilon) = \sum_{i,j} \frac{\left(\check{H}_{i,j} - n\check{p}_{i,j}\left(\check{\boldsymbol{\pi}}^{(1)}, \check{\boldsymbol{\pi}}^{(2)}\right) \right)^2}{n\check{p}_{i,j}(\check{\boldsymbol{\pi}}^{(1)}, \check{\boldsymbol{\pi}}^{(2)})} \tag{8}$$

We can then prove the following result, where the full proof is in the supplementary file.

Theorem 6.1. *Assuming $\boldsymbol{U}$ and $\boldsymbol{V}$ are independent with true probability vectors $\boldsymbol{\pi}^{(1)}, \boldsymbol{\pi}^{(2)} > 0$ respectively, then as $n \to \infty$ we have $\check{\mathcal{T}}^{(n)}_{\text{GenRR}}(\epsilon) \xrightarrow{D} \chi^2_{(r-1)(c-1)}$.*

We then use this result to design Algorithm 6.

Theorem 6.2. `LocalGenRRIND` *is ϵ-LDP.*

Testing with Bit Flipping. Lastly, we design an independence test when the data is reported via $\mathcal{M}_{\text{bit}}$ in Algorithm 4. Assuming that $\boldsymbol{H} = \sum_{\ell=1}^n \boldsymbol{X}_\ell \sim$

Algorithm 6 Local DP IND Test: `LocalGenRRIND`

```
Input: $\boldsymbol{x} = (\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n)$, $\epsilon$, $\alpha$, $\mathrm{H}_0 : \boldsymbol{p} = \boldsymbol{p}^0$.
    Let $\check{\boldsymbol{H}} = \sum_{i=1}^{n} \mathcal{M}_{\texttt{GenRR}}(\boldsymbol{x}_i, \epsilon)$.
    Set $q = \check{\boldsymbol{\mathcal{T}}}^{(n)}_{\mathrm{GenRR}}(\epsilon)$ from (8)
    if $q > \chi^2_{d-1,1-\alpha}$, Decision ← Reject.
    else Decision ← Fail to Reject.
Output: Decision
```

$\mathrm{Multinomial}\left(n, \boldsymbol{p}(\boldsymbol{\pi}^{(1)}, \boldsymbol{\pi}^{(2)})\right)$, then we know that replacing $\boldsymbol{p}^0$ with $\boldsymbol{p}(\boldsymbol{\pi}^{(1)}, \boldsymbol{\pi}^{(2)})$ in Section 5 gives us the following asymptotic distribution (treating the contingency table of values as a vector) with covariance matrix $\Sigma(\cdot)$ given in (5)

$$\sqrt{n}\left(\frac{\widetilde{\boldsymbol{H}}}{n} - \left[\underbrace{\left(\frac{e^{\epsilon/2}-1}{e^{\epsilon/2}+1}\right)\boldsymbol{\pi}^{(1)}\left(\boldsymbol{\pi}^{(2)}\right)^{\intercal} + \frac{1}{e^{\epsilon/2}+1}}_{\widetilde{\boldsymbol{p}}(\boldsymbol{\pi}^{(1)},\boldsymbol{\pi}^{(2)})}\right]\right) \xrightarrow{D} \mathrm{N}\left(\boldsymbol{0}, \Sigma\left(\boldsymbol{\pi}^{(1)}\left(\boldsymbol{\pi}^{(2)}\right)^{\intercal}\right)\right) \tag{9}$$

Similar to analysis for Theorem 6.1, we start with a rough estimate for the unknown parameters which converges in probability to the true estimates, so we use $\alpha_\epsilon = \left(\frac{e^{\epsilon/2}-1}{e^{\epsilon/2}+1}\right)$ to get

$$\widetilde{\boldsymbol{\pi}}^{(1)} = \left(\frac{1}{\alpha_\epsilon}\right)\left(\frac{\widetilde{H}_{i,\cdot}}{n} - \frac{c}{e^{\epsilon/2}+1} : i \in [r]\right)$$
$$\widetilde{\boldsymbol{\pi}}^{(2)} = \left(\frac{1}{\alpha_\epsilon}\right)\left(\frac{\widetilde{H}_{\cdot,j}}{n} - \frac{r}{e^{\epsilon/2}+1} : j \in [c]\right) \tag{10}$$

We then give the resulting statistic, parameterized by the unknown parameters $\boldsymbol{\pi}^{(\ell)}$, for $\ell \in \{1, 2\}$. For middle matrix $\widetilde{M} = \Pi\Sigma\left(\widetilde{\boldsymbol{\pi}}^{(1)}\left(\widetilde{\boldsymbol{\pi}}^{(2)}\right)^{\intercal}\right)^{-1}\Pi$, we have

$$\widetilde{\boldsymbol{\mathcal{T}}}^{(n)}_{\mathrm{BitFlip}}\left(\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}; \epsilon\right) = \frac{1}{n}\left(\widetilde{\boldsymbol{H}} - n\widetilde{\boldsymbol{p}}\left(\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}\right)\right)^{\intercal} \widetilde{M}\left(\widetilde{\boldsymbol{H}} - n\widetilde{\boldsymbol{p}}\left(\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}\right)\right) \tag{11}$$

Minimizing $\widetilde{\boldsymbol{\mathcal{T}}}^{(n)}_{\mathrm{BitFlip}}\left(\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}; \epsilon\right)$ over $(\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)})$ results in a statistic that is distributed as a chi-square random variable, we defer the full proof to the supplementary file.

Theorem 6.3. *Under the null hypothesis where* $\boldsymbol{U}$ *and* $\boldsymbol{V}$ *are independent with true probability vectors* $\boldsymbol{\pi}^{(1)}, \boldsymbol{\pi}^{(2)} > 0$ *respectively, then we have as* $n \to \infty$, $\min_{\boldsymbol{\theta}^{(1)},\boldsymbol{\theta}^{(2)}}\left\{\widetilde{\boldsymbol{\mathcal{T}}}^{(n)}_{\mathit{BitFlip}}\left(\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}; \epsilon\right)\right\} \xrightarrow{D} \chi^2_{(r-1)(c-1)}$.

We present the test in Algorithm 7. The following result follows from same privacy analysis as before.

Theorem 6.4. `LocalBitFlipIND` *is* ϵ*-LDP.*

Algorithm 7 Local DP IND Test: `LocalBitFlipIND`

```
Input: $(\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n)$, $\epsilon$, $\alpha$.
    Let $\widetilde{\boldsymbol{H}} = \sum_{i=1}^{n} \mathcal{M}_{\mathrm{bit}}(\boldsymbol{x}_i, \epsilon)$.
    $q = \min_{\boldsymbol{\pi}^{(1)},\boldsymbol{\pi}^{(2)}} \left\{\widetilde{\boldsymbol{\mathcal{T}}}^{(n)}_{\mathrm{BitFlip}}\left(\boldsymbol{\pi}^{(1)}, \boldsymbol{\pi}^{(2)}; \epsilon\right)\right\}$ from (11).
    if $q > \chi^2_{(r-1)(c-1),1-\alpha}$ Decision ← Reject.
    else Decision ← Fail to Reject.
Output: Decision
```

Empirical Results. As we did for the goodness of fit tests, we empirically compare the power for our various tests for independence. We consider the null hypothesis that the two sequences of categorical random variables $\{\boldsymbol{U}_\ell\}_{\ell=1}^n$ and $\{\boldsymbol{V}_\ell\}_{\ell=1}^n$ are independent of one another. Under an alternate hypothesis, we generate the contingency data according to a non-product distribution. We fix the distribution $\boldsymbol{p}^1$ for the contingency table to be of the following form, where $\boldsymbol{\pi}^{(1)} \in \mathbb{R}^r$ is the unknown distribution for $\{\boldsymbol{U}_\ell\}_{\ell=1}^n$, $\boldsymbol{\pi}^{(2)} \in \mathbb{R}^c$ is the unknown distribution for $\{\boldsymbol{V}_\ell\}_{\ell=1}^n$, and r, c are even

$$\boldsymbol{p}^1 = \boldsymbol{\pi}^{(1)}\left(\boldsymbol{\pi}^{(2)}\right)^{\intercal} + \eta(1, -1, \cdots, -1, 1)^{\intercal}(1, -1, \cdots, -1, 1) \tag{12}$$

Note that the hypothesis test does not know the underlying $\boldsymbol{\pi}^{(i)}$ for $i \in \{1, 2\}$, but to generate the data we must fix these distributions. We show power results when the marginal distributions satisfy $\boldsymbol{\pi}^{(1)} = (1/r, \cdots, 1/r)$ and $\boldsymbol{\pi}^{(2)} = (1/c, \cdots, 1/c)$. In Figure 2, we give results for various n and $\epsilon \in \{1, 2, 4\}$.

7. Conclusion

We have designed several hypothesis tests, each depending on different local differentially private algorithms. We showed that each statistic has a noncentral chi-square distribution when the data is drawn from some alternate hypothesis H_1. Depending on the form of the alternate probability distribution, the dimension of the data, and the privacy parameter, either `LocalGenRRGOF` or `LocalBitFlipGOF` gave the best power. This corroborates the results from Kairouz et al. (2014) who showed that in hypothesis testing, different privacy regimes have different optimal local differentially private mechanisms, although utility in their work was in terms of KL divergence. Our results show that the power of the test is directly related to the noncentral parameter of the test statistic that is used. This requires the data analyst to carefully consider alternate hypotheses, as well as the data dimension and privacy parameter for a particular test and then see which test statistic results in the largest noncentral parameter.

Acknowledgements

Marco Gaboardi has been partially supported by NSF under grant TWC-1565365.

References

American Statistical Association. Discovery with Data: Leveraging Statistics with Computer Science to Transform Science and Society, 2014.

Apple Press Info. Apple previews ios 10, the biggest ios release ever, 2016.

Bassily, R. and Smith, A. Local, private, efficient protocols for succinct histograms. In *Proceedings of the Forty-seventh Annual ACM Symposium on Theory of Computing*, STOC '15, pp. 127–135, New York, NY, USA, 2015. ACM. ISBN 978-1-4503-3536-2. doi: 10.1145/2746539.2746632.

Bun, M. and Steinke, T. Concentrated differential privacy: Simplifications, extensions, and lower bounds. In *Theory of Cryptography - 14th International Conference, TCC 2016-B, Beijing, China, October 31 - November 3, 2016, Proceedings, Part I*, pp. 635–658, 2016. doi: 10.1007/978-3-662-53641-4_24.

Cai, B., Daskalakis, C., and Kamath, G. Priv'it: Private and sample efficient identity testing. In *International Conference on Machine Learning*, 2017.

Duchi, J. C., Jordan, M. I., and Wainwright, M. J. Local privacy and statistical minimax rates. In *54th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2013, 26-29 October, 2013, Berkeley, CA, USA*, pp. 429–438, 2013a. doi: 10.1109/FOCS.2013.53.

Duchi, J. C., Wainwright, M. J., and Jordan, M. I. Local privacy and minimax bounds: Sharp rates for probability estimation. In *Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems 2013. Proceedings of a meeting held December 5-8, 2013, Lake Tahoe, Nevada, United States.*, pp. 1529–1537, 2013b.

Dwork, C. and Roth, A. The algorithmic foundations of differential privacy. *Foundations and TrendsÂő in Theoretical Computer Science*, 9(3âĂŞ4):211–407, 2014. ISSN 1551-305X. doi: 10.1561/0400000042.

Dwork, C., Kenthapadi, K., McSherry, F., Mironov, I., and Naor, M. Our data, ourselves: Privacy via distributed noise generation. In *Advances in Cryptology - EUROCRYPT 2006, 25th Annual International Conference on the Theory and Applications of Cryptographic Techniques, St. Petersburg, Russia, May 28 - June 1, 2006, Proceedings*, pp. 486–503, 2006a. doi: 10.1007/11761679_29.

Dwork, C., Mcsherry, F., Nissim, K., and Smith, A. Calibrating noise to sensitivity in private data analysis. In *In Proceedings of the 3rd Theory of Cryptography Conference*, pp. 265–284. Springer, 2006b.

Erlingsson, U., Pihur, V., and Korolova, A. Rappor: Randomized aggregatable privacy-preserving ordinal response. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*, CCS '14, pp. 1054–1067, New York, NY, USA, 2014. ACM. ISBN 978-1-4503-2957-6. doi: 10.1145/2660267.2660348.

Gaboardi, M., Lim, H. W., Rogers, R., and Vadhan, S. P. Differentially private chi-squared hypothesis testing: Goodness of fit and independence testing. In *Proceedings of the 33rd International Conference on International Conference on Machine Learning - Volume 48*, ICML'16, pp. 2111–2120. JMLR.org, 2016.

Hunter, D. R., Goodreau, S. M., and Handcock, M. S. Goodness of fit of social network models. *Journal of the American Statistical Association*, 103:248–258, 2008.

Johnson, A. and Shmatikov, V. Privacy-preserving data exploration in genome-wide association studies. In *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '13, pp. 1079–1087, New York, NY, USA, 2013. ACM.

Kairouz, P., Oh, S., and Viswanath, P. Extremal mechanisms for local differential privacy. In Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N. D., and Weinberger, K. Q. (eds.), *Advances in Neural Information Processing Systems 27*, pp. 2879–2887. Curran Associates, Inc., 2014.

Kairouz, P., Bonawitz, K., and Ramage, D. Discrete distribution estimation under local privacy. In *Proceedings of the 33nd International Conference on Machine Learning, ICML 2016, New York City, NY, USA, June 19-24, 2016*, pp. 2436–2444, 2016.

Kakizaki, K., Fukuchi, K., and Sakuma, J. Differentially private chi-squared test by unit circle mechanism. In Precup, D. and Teh, Y. W. (eds.), *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pp. 1761–1770, International Convention Centre, Sydney, Australia, 06–11 Aug 2017. PMLR.

Karwa, V. and Slavković, A. Inference using noisy degrees: Differentially private β-model and synthetic graphs. *Ann. Statist.*, 44(P1):87–112, 02 2016.

Kifer, D. and Rogers, R. A New Class of Private Chi-Square Hypothesis Tests. In Singh, A. and Zhu, J. (eds.), *Proceedings of the 20th International Conference on Artificial*

Intelligence and Statistics, volume 54 of *Proceedings of Machine Learning Research*, pp. 991–1000, Fort Lauderdale, FL, USA, 20–22 Apr 2017. PMLR.

McSherry, F. and Talwar, K. Mechanism design via differential privacy. In *Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, Providence, RI, October 2007. IEEE.

Pastore, A. and Gastpar, M. Locally differentially-private distribution estimation. In *2016 IEEE International Symposium on Information Theory (ISIT)*, pp. 2694–2698, July 2016. doi: 10.1109/ISIT.2016.7541788.

Raskhodnikova, S., Smith, A., Lee, H. K., Nissim, K., and Kasiviswanathan, S. P. What can we learn privately? *2013 IEEE 54th Annual Symposium on Foundations of Computer Science*, 00:531–540, 2008. ISSN 0272-5428. doi: doi.ieeecomputersociety.org/10.1109/FOCS.2008.27.

Sheffet, O. Differentially private least squares: Estimation, confidence and rejecting the null hypothesis. *arXiv preprint arXiv:1507.02482*, 2015.

Sheffet, O. Locally private hypothesis testing. *CoRR*, abs/1802.03441, 2018. URL http://arxiv.org/abs/1802.03441.

Steele, J. E., Sundsøy, P. R., Pezzulo, C., Alegana, V. A., Bird, T. J., Blumenstock, J., Bjelland, J., Engø-Monsen, K., de Montjoye, Y.-A., Iqbal, A. M., Hadiuzzaman, K. N., Lu, X., Wetter, E., Tatem, A. J., and Bengtsson, L. Mapping poverty using mobile phone and satellite data. *Journal of The Royal Society Interface*, 14(127), 2017. ISSN 1742-5689. doi: 10.1098/rsif.2016.0690.

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

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

Warner, S. L. Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association*, 60:63–69, 1965.

Ye, M. and Barg, A. Optimal schemes for discrete distribution estimation under local differential privacy. In *2017 IEEE International Symposium on Information Theory (ISIT)*, pp. 759–763, June 2017. doi: 10.1109/ISIT.2017.8006630.

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