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

\(^1\)Equal contribution \(^1\) University at Buffalo, Buffalo, NY, USA
\(^2\)University of Pennsylvania, Philadelphia, PA, USA. Correspondence to: Marco Gaboardi <gaboardi@buffalo.edu>.

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: \textit{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; \textit{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, \textit{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: \textit{LocalNoiseIND} (see supplementary file), \textit{LocalGenRRIND}, and \textit{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 \textit{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; Kairouz 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, \textit{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 \textit{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 \(\epsilon\)) one mechanism has higher \textit{utility} than other mechanisms and for large privacy regimes (large \(\epsilon\)) 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 \(x = (x_1, \ldots, x_n) \in \mathcal{X}^n\) 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 \(x, x' \in \mathcal{X}^n\) are neighboring if they differ in at most one element, i.e. \(\exists i \in [n] \text{ such that } x_i \neq x'_i \text{ and } \forall j \neq i, x_j = x'_j\).

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

Definition 3.2 (Bun & Steinke (2016)). An algorithm \(\mathcal{M} : \mathcal{X}^n \rightarrow \mathcal{Y}\) is \(z\)-zero-mean concentrated differentially private (zCDP) if for all neighboring datasets
We then define when the data was actually generated from it, by at most 
As was studied in (Gaboardi et al., 2016), (Wang et al., 2016) tests so that they can reject 
and always achieve this is easy, because we can just ignore the data 

From this we have that a \( L_{\varrho} \) algorithm is also \( \varrho \)-LDP Algorithm that all the data is stored in a central location and the algorithm \( M \) 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 \( x \), a local randomizer oracle \( LR_x ((\cdot, \cdot) \) takes as input an index \( i \in [n] \) and an \( \epsilon \)-DP algorithm \( R \), and outputs \( y \in Y \) chosen according to the distribution of \( R(x_i) \), i.e. \( LR_x (i, R) = R(x_i) \).

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

From this we have that a \((\epsilon, \delta)\)-LDP algorithm is also \((\epsilon, 0)\)-DP. Note that these definitions can be extended to include \( \rho \)-local \( z \)-CDP (LzCDP) where each local randomizer is \( \rho_j \)-\( z \)-CDP 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 \( M : X^n \rightarrow Y \) is \((\epsilon, 0)\)-LDP then it is also \( \epsilon^2/2\)-LzCDP. If \( M \) is \( \rho \)-LzCDP, then it is also \( \left( \rho + \sqrt{2 \rho \ln(2/\delta)} \right) \)-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 \( \alpha \), 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 \( X_i \) for \( i \in [n] \) is sampled i.i.d. from \( \text{Multinomial}(1, p) \) where \( p \in \mathbb{R}^d_{>0} \) and \( p^T \cdot \mathbf{1} = 1 \). The classical chi-square hypothesis test (without privacy) forms the histogram \( \bar{H} = (H_1, \cdots, H_d) = \sum_{i=1}^n X_i \) and computes the chi-square statistic \( T = \sum_{j=1}^d \frac{(h_j - np^0_j)^2}{np^0_j} \).

The reason for using this statistic is that it converges in distribution to \( \chi^2_{d-1} \) as more data is collected, i.e. \( n \rightarrow \infty \), when \( H_0 : p = p^0 \) holds. Hence, we can ensure the probability of false discovery to be close to \( \alpha \) as long as we only reject \( H_0 \) when \( T > \chi^2_{d-1, 1-\alpha} \) 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 \( Z \) from distribution \( D^n \) to the histogram of counts \( \hat{H} = H + Z \), which ensures \( \rho \)-zCDP when \( D = N(0, 1/\rho) \) and \( \epsilon \)-DP when \( D = \text{Lap}(2/\epsilon) \). The chi-square statistic then becomes

\[
\bar{T}(D) = \sum_{i=1}^d \frac{(H_i + Z_i - np^0_i)^2}{np^0_i} \quad \text{where} \quad Z \sim D^n. \tag{1}
\]

The previous works then show that this statistic converges in distribution to a linear combination of chi-squared variables, when \( D \sim N(0, 1/\rho) \) and \( \rho \) 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 \( \Pi \) defined as \( (I_d - \frac{1}{d} \mathbf{1} \mathbf{1}^T) \), middle matrix \( M_\sigma = \Pi (\text{Diag}(p^0 + \sigma) - p^0 (p^0)^T)^{-1} \Pi \), and sample noise \( Z \sim D^n \), with \( \tilde{H} = H + Z \)

\[
T^{(n)}_{K\Pi}(\sigma; D) = n \left( \frac{\tilde{H}}{n} - p^0 \right)^T M_\sigma \left( \frac{\tilde{H}}{n} - p^0 \right) \tag{2}
\]

We use \( D = \text{Lap}(2/\epsilon) \) with \( \sigma = \frac{8}{\sqrt{n}} \) for an \( \epsilon \)-DP claim or \( D = N(0, 1/\rho) \) with \( \sigma = \frac{\rho}{np} \) for a \( \rho \)-zCDP claim. When comparing the power of all our tests, we will consider the alternate \( H_1 : p = p^n \) where \( p^n_i = p^0 + \frac{\Delta_i}{\sqrt{n}} \) where \( \mathbf{1}^T \Delta = 0 \).
Further, under the alternate hypothesis $H_1 : p = p^1$, the statistic $T_{KR}^{(n)}(1/\rho; N(0, 1/\rho))$ given in (2) for $\rho > 0$ converges in distribution to $\chi^2_{d-1}$. Further, under the alternate hypothesis $H_1 : p = p^1$, the resulting asymptotic distribution is a noncentral chi-square random variable with $d - 1$ degrees of freedom and noncentral parameter $\Delta = (\text{Diag}(p^0) - \text{Diag}(p^1))^\top + 1/\rho I_d$.

When $D = \text{Lap}(2/\epsilon)$, 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 $\alpha$ 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 $Z_i \sim N(0, 1/\rho I_d)$ independent noise to each individual’s data $X_i$ to ensure $\rho$-LzCDP or $Z_i \sim \text{Lap}(2/\epsilon)$ independent noise to $X_i$ to ensure $\epsilon$-LDP. In either case, the resulting noisy histogram $\tilde{H} = H + Z$ where $Z = \sum_{i=1}^n Z_i$ 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, $\tilde{H} = H + Z$ where $Z \sim N(0, \frac{1}{\rho} I_d)$. Thus, we can use either statistic $T_{KR}^{(n)}(\rho/n)$ or $T_{KR}^{(n)}(\rho/n)$, 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 $\rho$-LzCDP when $D = N(0, 1/\rho)$ and $\epsilon$-LDP when $D = \text{Lap}(2/\epsilon)$.

Although we cannot guarantee the probability of a Type I error at most $\alpha$ 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 $\tau$ that ensures the probability of a Type I error is at most $\alpha$ if we reject $H_0$ when the statistic is larger than $\tau$. For the local setting, if each individual perturbs each coordinate by adding $\text{Lap}(2/\epsilon)$ this will ensure our test is $\epsilon$-LDP. However, the sum of independent Laplace random variables is not Laplace, so we will need to estimate a sum of 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 can apply 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 $\{e_1, \ldots, e_d\}$, where $e_j \in \mathbb{R}^d$ is the standard basis element with a 1 in the $j$th 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 $M_{\text{GenRR}}$ is $\epsilon$-DP.

We have the following result when we use $M_{\text{GenRR}}$ on each data entry to obtain a private histogram.
Further, with alternate $H$, we can then form a chi-square statistic using the histogram $H$ which will have the correct asymptotic distribution.

**Theorem 5.3.** Let $H \sim \text{Multinomial}(n,p)$ and $\hat{H}$ be given in Theorem 5.2 with privacy parameter $\epsilon > 0$. Under the null hypothesis $H_0 : p = p^0$, we have for $p^0 = \frac{1}{e^\epsilon + d - 1} \left( e^\epsilon p^1 + (1 - p^0) \right)$,

$$T_{\text{GenRR}}^{(n)}(\epsilon) = \frac{d}{n} \left( H_j - n\hat{p}_j^0 \right)^2 D \rightarrow \chi_d^2. \quad (4)$$

Further, with alternate $H_1 : p = p^1$, the resulting asymptotic distribution is a noncentral chi-square distribution with $d - 1$ degrees of freedom and noncentral parameter,

$$\left( \frac{e^{-1}}{e^\epsilon + d - 1} \right)^2 \sum_{j=1}^d \frac{\Delta_j^2}{p_j^0}.$$

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

**Algorithm 2 Generalized Randomized Response: $M_{\text{GenRR}}$**

**Input:** $x \in \{e_1, \cdots, e_d\}, \epsilon$

Let $q(x, z) = 1\{x = z\}$

Select $\tilde{x}$ with probability $\frac{\exp[\epsilon q(x,\tilde{x})]}{e^\epsilon + d - 1}$

**Output:** $\tilde{x}$

**Lemma 5.2.** If we have histogram $H = \sum_{i=1}^n X_i$, where $\{X_i\}_{i=1}^n \sim \text{Multinomial}(1, p)$ and we write $H = \sum_{i=1}^n M_{\text{GenRR}}(X_i, \epsilon)$ for each $i \in [n]$, then $H \sim \text{Multinomial}(n, \tilde{p})$ where

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

Once we have $\hat{H}$, we can create a chi-square statistic by subtracting $H$ by its expectation and dividing the difference by the expectation. Hence testing $H_0 : p = p^0$ after the generalized randomized response mechanism is equivalent to testing $H_0 : p = \tilde{p}^0$ with data $\hat{H}$.

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

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

**Theorem 5.5.** The algorithm $M_{\text{bin}}$ is $\epsilon$-DP.

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 $\left( \frac{e^{-1}}{e^\epsilon + d - 1} \right)^2$. Thus, for large dimensional data the generalized randomized response cannot reject a false null hypothesis. We next consider another differentially private algorithm $M : \{e_1, \cdots, e_d\} \to \{0, 1\}^d$, given in Algorithm 4 used in (Bassily & Smith, 2015) that flips each bit with some biased probability.

**Algorithm 4 Bit Flip Local Randomizer: $M_{\text{bin}}$**

**Input:** $x \in \{e_1, \cdots, e_d\}, \epsilon$

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:** $z$

Further, with alternate $H_1 : p = p^1$, the resulting asymptotic distribution is a noncentral chi-square distribution with $d - 1$ degrees of freedom and noncentral parameter,

$$\left( \frac{e^{-1}}{e^\epsilon + d - 1} \right)^2 \sum_{j=1}^d \frac{\Delta_j^2}{p_j^0}.$$.

The histogram $\tilde{H} = \sum_{i=1}^n M_{\text{bin}}(X_i)$ has the following asymptotic distribution $\sqrt{n}(\tilde{H} - \bar{p}) \overset{D}{\to} N(0, \Sigma(p))$. Further, $\Sigma(p)$ is invertible for any $\epsilon > 0$ and $p > 0$.

Following a similar analysis in (Kifer & Rogers, 2017), we can form the following statistic for null hypothesis $H_0 : p = p^0$ in terms of the histogram $H$ and projection matrix $\Pi = I_d - \frac{1}{n}11^T$, as well as the covariance $\Sigma = \Sigma(p^0)$ and mean $\bar{p}^0$ both given in (5) where we replace $p$ with $p^0$:

$$T_{\text{BinFlip}}^{(n)}(\epsilon) = n \left( \frac{\tilde{H}}{n} - \bar{p}^0 \right)^T \Pi \Sigma^{-1} \Pi \left( \frac{\tilde{H}}{n} - \bar{p}^0 \right) \quad (6)$$

We can then design a hypothesis test based on the outputs from $M_{\text{bin}}$ in Algorithm 5

**Theorem 5.7.** $\text{LocalBinFlipGOF}$ is $\epsilon$-LDP.
Noncentral Parameter Coefficient

We set \( \rho \) presented in Algorithms 1, 3 and 5. We consider the null hypothesis \( H_0 : p = p^0 \).

The larger this coefficient is, the better the power will be on Sept. 13, 2016.

\( \Delta \Delta \Delta \) differentially private algorithms used in practice.

\( \Delta \Delta \Delta \) noise does not satisfy \( \Delta \Delta \Delta \) distribution for each local private test presented thus far.

\( \Delta \Delta \Delta \) hypothesis in 1000 independent trials for various sample sizes \( n \) and privacy parameters \( \epsilon \). From Figure 2, we can see that the test statistics that have the largest noncentral parameter for a particular dimension \( d \) and privacy parameter \( \epsilon \) 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 \( \epsilon \), the null hypothesis \( p^0 \), and alternate \( p = p^0 + \Delta \). 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.

Algorithm 5 Local DP GOF Test: LocalBitFlipGOF

**Input:** \( x = (x_1, \cdots, x_n) \), \( \epsilon, \alpha \), \( H_0 : p = p^0 \).

Let \( \overline{H} = \sum_{i=1}^{n} M_{bit}(x_i, \epsilon) \).

Set \( q = T_{\text{BitFlip}}^{(n)}(\epsilon) \).

**if** \( q > \chi^2_{d-1, 1-\alpha} \) \( \text{Decision} \leftarrow \text{Reject.} \)

**else** \( \text{Decision} \leftarrow \text{Fail to Reject.} \)

**Output:** Decision

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

**Theorem 5.8. If the null hypothesis \( H_0 : p = p^0 \) holds, then the statistic \( T_{\text{BitFlip}}^{(n)}(\epsilon) \) is asymptotically distributed as a chi-square, i.e. \( T_{\text{BitFlip}}^{(n)}(\epsilon) \xrightarrow{D} \chi^2_{d-1} \). Further, if we consider the alternate \( H_1 : p = p^1 \) then \( T_{\text{BitFlip}}^{(n)}(\epsilon) \) converges in distribution to a noncentral chi-square with \( d-1 \) degrees of freedom and noncentral parameter \( (\frac{\epsilon}{\alpha}^2 - 1)^2 \cdot \Delta^\top \Sigma(p^0)^{-1} \Delta \).**

**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 \( p^0 = (1/d, \cdots, 1/d) \) for \( d > 2 \), and alternate \( H_1 : p = p^1 + \Delta \). 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 \( \Delta^\top \Delta \) 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 \( \Delta \) vector. Note that in LocalNoiseGOF, we set \( \rho = \epsilon^2/8 \) which makes the variance the same as for a random variable distributed as \( \text{Lap}(2/\epsilon) \) for an \( \epsilon \)-DP guarantee – recall that LocalNoiseGOF with Gaussian noise does not satisfy \( \epsilon \)-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.\[2\]

\[\text{In (Erlingsson et al., 2014), we know that Google uses } \epsilon = \ln(3) \text{ in RAPPOR and from Aleksandra Korolova’s Twitter post on Sept. 13, 2016 https://twitter.com/korolova/status/775801259504734208, Apple uses } \epsilon = 1.4.\]
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 $\{U_i\}_{i=1}^n \sim \text{Multinomial}(1, \pi(i))$ for $\pi(i) \in \mathbb{R}^r$, $\{V_i\}_{i=1}^n \sim \text{Multinomial}(1, \pi(2))$ for $\pi(2) \in \mathbb{R}^c$ and no component of $\pi(1)$ or $\pi(2)$ 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 $H$ whose $(i, j)$th coordinate is $H_{i,j} = \sum_{\ell=1}^n 1\{U_{\ell,i} = 1 \& V_{\ell,j} = 1\}$. Note that $H$ is then the contingency table over the joint outcomes.

Under the null hypothesis of independence between $\{U_i\}_{i=1}^n$ and $\{V_i\}_{i=1}^n$, for probability vector $p(\pi(1), \pi(2)) = \pi(1)^T \pi(2)$, we have $H \sim \text{Multinomial}(n, \pi(1), \pi(2))$.

What makes this test difficult is that the analyst does not know the data distribution $p(\pi(1), \pi(2))$ 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 $X_\ell$ which would be 1 in exactly one location. Thus we can alternatively write the contingency table as $H = \sum_{\ell=1}^n X_\ell$. 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 $M_{\text{GenRR}}$ given in Algorithm 2. In this case our contingency table can be written as $H \sim \text{Multinomial}(n, \tilde{p}(\pi(1), \pi(2)))$ where $\beta_\epsilon = \frac{1}{e^\epsilon + \epsilon r - 1}$ and we use (3) to get

$$\tilde{p}(\pi(1), \pi(2)) = \beta_\epsilon \left((e^\epsilon - 1) \pi(1)^T \pi(2) + 1\right)$$  (7)

We then obtain an estimate for the unknown parameters,

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

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

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

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

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

We then use this result to design Algorithm 6.

Theorem 6.2. LocalGenRRIND is $\epsilon$-LDP.

Testing with Bit Flipping. Lastly, we design an independence test when the data is reported via $M_{\text{Bit}}$ in Algorithm 4. Assuming that $H = \sum_{\ell=1}^n X_\ell \sim$
We present the test in Algorithm 7. The following result follows from same privacy analysis as before.

**Theorem 6.4.** *LocalBitFlipIND is $\epsilon$-LDP.*
Acknowledgements

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

References


Local Private Hypothesis Testing: Chi-Square Tests


