Differentially Private Chi-squared Test by Unit Circle Mechanism

Kazuya Kakizaki [1] **Kazuto Fukuchi** [1] **Jun Sakuma** [1 2 3]

Abstract

This paper develops differentially private mechanisms for χ^2 test of independence. While existing works put their effort into properly controlling the type-I error, in addition to that, we investigate the type-II error of differentially private mechanisms. Based on the analysis, we present *unit circle mechanism*: a novel differentially private mechanism based on the geometrical property of the test statistics. Compared to existing output perturbation mechanisms, our mechanism improves the dominated term of the type-II error from $O(1)$ to $O(\exp(-\sqrt{N}))$ where N is the sample size. Furthermore, we introduce novel procedures for multiple χ^2 tests by incorporating the unit circle mechanism into the sparse vector technique and the exponential mechanism. These procedures can control the family-wise error rate (FWER) properly, which has never been attained by existing mechanisms.

1. Introduction

Hypothesis testing is a statistical framework to ascertain whether or not given samples follow a specific model or not in a systematic manner. For this study, we presume that highly sensitive information might be included in the samples. A typical situation is genome-wide association studies (GWAS). Homer et al. reported that a patient's disease status could be inferred from aggregate statistics collected for GWAS (Homer et al., 2008).

Differential privacy (Dwork et al., 2006) is a recent notion of privacy that is tailored to a privacy-preserving release of aggregate statistics. As described in this paper, we give specific consideration to the differential privacy of χ^2 test of independence. Two lines of studies exist in differentially private χ^2-tests of independence: input perturbation and output perturbation. The input perturbation method ensures differential privacy by randomizing each count in the contingency table (Uhler et al., 2013; Johnson & Shmatikov, 2013; Wang et al., 2015; Gaboardi et al., 2016). The output perturbation method randomizes the test statistic to satisfy differential privacy (Fienberg et al., 2011; Yu et al., 2014; Uhler et al., 2013; Wang et al., 2015). Also recently, there is a study for the differential privacy of χ^2 test of identity, which ensures type-I and typ-II error (Cai et al., 2017).

The contribution of this study is three-fold. First, we analyze the type-II error of the differentially private mechanism for output perturbation. In principle, (non-privacy-preserving) hypothesis test procedures are preferred to achieve a greater power with keeping the significance at a prescribed level. In the case of differentially private hypothesis test, the same is required under the constraint of differential privacy. Much efforts have been devoted to controlling the significance level properly (Wang et al., 2015; Gaboardi et al., 2016) whereas little attention has been paid to the power analysis. We analyze the type-II error of the differentially private mechanism for χ^2-test, and show that it is upper-bounded by two terms which are mechanism-dependent. The bound indicates that a mechanism with a lower sensitivity and lower γ error (defined in Section 4) achieves a greater power.

Second, we investigate the geometrical property of the test statistic of the χ^2-test. Then, we propose the unit circle mechanism: a novel differentially private mechanism based on the geometrical property. In existing mechanisms based on output perturbation, the sensitivity is $O(1)$ in terms of N (Fienberg et al., 2011; Yu et al., 2014; Uhler et al., 2013). We demonstrate that the sensitivity of the unit circle mechanism is $O(1/\sqrt{N})$, which achieves lower type-II error (Theorem 7).

Third, we present two procedures for differentially private multiple χ^2 tests of independence that can control the familywise error rate (FWER). Actually, FWER can be controlled properly using existing DP mechanisms for χ^2-test by repeatedly using (Uhler et al., 2013), for example. However, such a naive construction consumes the privacy budget, which is linear to the number of all the considered

[1]Department of Computer Science, University of Tsukuba, 1-1-1 Tennohdai, Tsukuba, Ibaraki, Japan [2]JST CREST [3]RIKEN Center for Advanced Intelligence Project. Correspondence to: Kazuya Kakizaki <kazuya@mdl.cs.tsukuba.ac.jp>, Kazuto Fukuchi <kazuto@mdl.cs.tsukuba.ac.jp>, Jun Sakuma <jun@cs.tsukuba.ac.jp>.

Proceedings of the 34^{th} International Conference on Machine Learning, Sydney, Australia, PMLR 70, 2017.

tests and which would produce useless test results. There exist several studies that work with a lower privacy budget in the multiple testing setting (Johnson & Shmatikov, 2013). However, to the best of our knowledge, no differentially private multiple testing procedure that can control FWER properly has been presented. We introduce two novel procedures for multiple χ^2 tests by incorporating the unit circle mechanism into the sparse vector technique (SVT) (Dwork & Roth, 2014) and exponential mechanism (McSherry & Talwar, 2007). The exponential mechanism based procedure works in the non-interactive setting, in which the set of hypotheses considered needs to be prescribed before starting the testing procedure. This mechanism first selects top-k significant hypotheses and then performs the statistical test for each hypothesis. The SVT based procedure works in the interactive setting, in which the hypothesis to be considered at each round can be interactively chosen after observing the result of the tests in the previous rounds. The SVT based procedure controls FWER by adjusting the threshold of SVT using Monte Carlo sampling. Simultaneously, we show the privacy budget consumed by the both procedures is not dependent on the total number of the hypotheses considered, but only on the prescribed maximum number of hypothesizes to be accepted.

2. Preliminaries

2.1. Differential Privacy

Let $S = \{x_i | x_i \in \mathcal{X}, i = 1, \ldots, N\}$ be a set of records. Presume that an analyst holds S and wishes to release a result of statistical analysis $f(S)$ to the public, where $f : \mathcal{S} \to \mathcal{Y}$ is a statistical query. $\mathcal{S}$ and $\mathcal{Y}$ respectively denote the input domain and output domain. Differential privacy (DP) is a privacy definition that restricts privacy breach of any element in S caused by releasing $y = f(S)$ (Dwork et al., 2006). Let $d(S, S') = |\{i : x_i \neq x'_i, x_i \in S, x'_i \in S'\}|$ denote the Hamming distance between two sets. When $h(S, S') = 1$, we say that S and S' are adjacent, or that $S \sim S'$. Ensuring DP requires randomization of outputs, by definition. Let $\mathcal{M}_f(S)$ denote a randomization mechanism of f. $\mathcal{M}_f$ is differentially private if it satisfies the following definition:

Definition 1 (ϵ-DP (Dwork et al., 2006))**.** *Mechanism $\mathcal{M} : \mathcal{S} \to \mathcal{Y}$ provides ϵ-DP if, for any $S \sim S'$ and $Y \subseteq \mathcal{Y}$,*

$$\Pr[\mathcal{M}(S) \in Y] \leq \exp(\epsilon) \Pr[\mathcal{M}(x') \in Y].$$

We introduce a basic mechanism that ensures DP for scalar outputs. For query f, the following randomization by Laplace mechanism

$$\mathcal{M}(S) = f(S) + \mathrm{Lap}(\frac{\Delta_f}{\epsilon})$$

guarantees ϵ-DP, where $\mathrm{Lap}(b)$ denotes a random value generated from the Laplace distribution for which the scale parameter is b (Dwork et al., 2006). Δ_f denotes the sensitivity of the query, which is defined as

$$\Delta_f = \max_{S \sim S'} |f(S) - f(S')|.$$

We remark that no post-processing of outputs obtained from any DP mechanism changes the guarantee on DP (Dwork & Roth, 2014).

2.2. Answering Multiple Queries

When we obtain multiple outputs from a mechanism computed on disjoint data subsets, the following composition theorem is applied.

Theorem 1 (Composition theorem (McSherry & Talwar, 2007))**.** *Let $M_i : S \to \mathcal{Y}_i$ be ϵ-DP mechanism for $i = 1, \cdots, K$. Let $\hat{M} : S \to \mathcal{Y}_i \times, \cdots, \times \mathcal{Y}_K$ be the mechanism that outputs $(M_1, \cdots, M_K)$. Then, $\hat{M}$ is $K\epsilon$-differential private.*

Let $f_1, \ldots, f_K$ be a query sequence. Suppose analyst wishes to know whether or not $f_k(S) > \tau$ holds for all k where τ is a threshold. If the analyst obtains $\{f_k(S)\}_{k=1}^K$, the privacy budget for each query needs to be set to $1/K\epsilon$ to ensure ϵ-DP, which would result in extremely unuseful outputs when K is large. Suppose the analyst is interested only in whether or not each output is above a threshold value. If the analyst has a good reason to believe that only $s(\ll K)$ answers would exceed the threshold value, the sparse vector technique (SVT) (Dwork & Roth, 2014) helps to reduce consumption of the privacy budget drastically. Let $\top$ mean that $f_k(S) > \tau$. Assuming SVT is terminated after answering at most s outputs, it outputs $\top$ only when the following holds for $k = 1, \ldots$:

$$f_k(S) + Lap(4\Delta s/\epsilon) \geq \tau + \rho \tag{1}$$

where $\rho = Lap(2\Delta s/\epsilon)$ and Δ is the sensitivity of f_k[1]. Otherwise, it outputs nothing. It is proved that SVT guarantees ϵ-DP. What is remarkable about SVT is that the privacy budget is independent on the total number of the queries, but only on s. See Appendix. A for the detail .

3. Differentially Private χ^2-Test

3.1. χ^2-Test of Independence

Let X_0 and X_1 be discrete random variables and presume that we are interested in the independence between the two random variables. In this study, we suppose the random variables are binary. For the statistical test of independence, the null hypothesis is that X_0 and X_1 are independent. Let $S = \{x_1, \cdots, x_N\}$ be a set of samples drawn

[1]We require that all the queries have the same sensitivity

Table 1. Contingency table T of two binary varaibles

	$X_1 = 1$	$X_1 = 0$	
$X_0 = 1$	c_{11}	c_{10}	M_1
$X_0 = 0$	c_{01}	c_{00}	M_0
	N_1	N_0	N

from the two random variables where $x_n^T = (x_{n,0}, x_{n,1}) \in \{0,1\}^2$ are realization of X_0 and X_1. Given S, Table 1 denotes the 2×2 contingency table w.r.t. X_0 and X_1 where c_{pq} denotes the number of samples in S such that $X_0 = p$ and $X_1 = q$. The marginals N_1, N_0, M_1, and M_0 are defined as in Table 1. The test statistic for χ^2-test of independence is given as

$$\chi^2(S) = \frac{(c_{11}c_{00} - c_{10}c_{01})^2 N}{(c_{11}+c_{10})(c_{11}+c_{01})(c_{10}+c_{00})(c_{01}+c_{00})}. \quad (2)$$

Under the null hypothesis H_0, the χ^2-test statistics is known to follow the χ^2 distribution with one degree of freedom, asymptotically (Bishop et al., 1975). Given S and *significance level* α, the χ^2-test of independence is run as

$$\chi^2\text{-test}(S, \alpha) = \begin{cases} \mathsf{rej} & \text{if } \chi^2(S) > \tau_\alpha, \\ \mathsf{acc} & \text{otherwise.} \end{cases}$$

Therein, τ_α is a threshold determined such that $\int_\tau^\infty \chi^2_{(1)}(z)dz = \alpha$. Here, rej and acc respectively indicate that H_0 is rejected and accepted. We learn that there exists evidence that X_0 and X_1 are dependent on the significance level of α if H_0 rejected. The type-1 error of χ^2 test is equivalent to the significance level:

$$\alpha = \Pr[\chi^2(S) > \tau_\alpha | H_0 \text{ is true}].$$

The *power* is defined by the probability of rejecting the null hypothesis when the alternative hypothesis is true:

$$1 - \beta = \Pr[\chi^2(S) > \tau_\alpha | H_1 \text{ is true}]$$

Therein, β denotes the type-II error.

Let $\mathcal{M}$ be a randomization mechanism for χ^2 testing. Under the constraints that $\mathcal{M}$ ensures DP and that type-I error is preserved at most α, we evaluate the utility of the mechanism by $1 - \beta$, the power of the mechanism.

3.2. Output Perturbation Method

The most straightforward method to ensure the DP of χ^2 test is the randomization of the test statistic (Fienberg et al., 2011; Yu et al., 2014; Uhler et al., 2013; Wang et al., 2015). Application of the Laplace mechanism to the χ^2 statistic immediately ensures DP, as

$$\widehat{\chi^2}(S) = \chi^2(S) + Lap(\frac{\Delta}{\epsilon}),$$

where Δ is the sensitivity of the test statistic. For example, given N_0 and N_1 are released to the public, (Yu et al., 2014) derived the following sensitivity:

$$\Delta_Y = \frac{N^2}{N_0 N_1}\left(\frac{\max\{N_0, N_1\}}{\max\{N_0, N_1\}+1}\right). \quad (3)$$

Other sensitivity analyses are provided elsewhere in the relevant literature (Fienberg et al., 2011; Wang et al., 2015). In principle, these sensitivities are in $O(1)$ with respect to N assuming $N_0 \simeq N_1$.

3.3. Input Perturbation Method

Given a contingency table, the input perturbation method first randomizes each cell of the contingency table independently as[2] $\hat{c}_{pq} = c_{pq} + Lap(\frac{2}{\epsilon})$, and then evaluates the test statistic with the randomized table (Johnson & Shmatikov, 2013). After randomization, each cell can take a negative value. (Barak et al., 2007; Li et al., 2010; Hardt et al., 2012; Li & Miklau, 2012; Gaboardi et al., 2014) have suggested methods of calculating a differentially private contingency table while avoiding this problem. For independence testing, (Gaboardi et al., 2016) modified the contingency table by using the work of (Lee et al., 2015) based on constraint optimization problem so that the total number of the samples is N and each cell has a positive value. The DP of the test statistic is readily ensured because of the post-processing theorem. In input perturbation, given $N_0 \simeq N_1$, (Gaboardi et al., 2016) and (Wang et al., 2015) reported by experiments that the accuracy of the test results can be improved when the number of samples increases.

3.4. Significance Level of DP χ^2 Test

The threshold for non-privacy-preserving χ^2 test is determined based on the fact that $\chi^2(S)$ follows the χ^2 distribution asymptotically when H holds. However, $\widehat{\chi^2}(S)$ does not follow the χ^2 distribution anymore, even when H holds. For the randomization mechanism to keep the type-I error α, the threshold τ_α needs to be adjusted so that $\alpha = \Pr[\widehat{\chi^2}(S) > \hat{\tau}_\alpha | H_0 \text{ is true}]$ holds. Letting $\widehat{\chi^2_{(1)}}$ denote the distribution that $\widehat{\chi^2}(S)$ follows, then $\hat{\tau}_\alpha$ is determined so that $\int_{\hat{\tau}_\alpha}^\infty \widehat{\chi^2_{(1)}}(z)dz = \alpha$.

For the input perturbation case, (Uhler et al., 2013) demonstrated that the test static arising from randomized cell counts asymptotically approximates the χ^2 distribution. However, the sample distribution can deviate considerably from the χ^2 distribution when the sample size is small. A study by (Gaboardi et al., 2016) also proved that the ex-

[2] When one can assume that the marginal of the table (N_0 and N_1) is known publicly, the global sensitivity of the count query is 1.

pectation of $\widehat{\chi^2}(S)$ generated by input perturbation is biased. (Gaboardi et al., 2016) and (Wang et al., 2015) independently presented methods to correct the bias by adjusting the threshold, respectively using Monte Carlo sampling and a permutation test.

For the output perturbation case, (Uhler et al., 2013) derived a distribution of perturbed χ^2 statistic when the test statistic is randomized with the Laplace distribution. In this case, the threshold can be adjusted accurately using the derived distribution. Although not described in the literature above specifically, the threshold for test statistics generated by output perturbation can be adjusted using Monte Carlo sampling in a method similar to one reported in (Gaboardi et al., 2016).

4. Power of DP χ^2 Test

The type-I error can be correctly controlled by adjusting the threshold appropriately as discussed in Section 3.4 whereas theoretical analysis on the power has never been intensively investigated in existing works. For the power analysis, we derive the following upper bound.

Theorem 2. *Let β_τ denote the type-II error when one use threshold τ for the (non-privacy-preserving) χ^2 test. Let $\mathcal{P} = \{P : H_1 \textit{is true}\}$ be the set of distributions of sample sets. Let $\mathcal{M}$ be a differentially private mechanism for χ^2 test and $\hat{\tau}_\alpha > \tau_\alpha$ be the threshold for $\mathcal{M}$ that is determined so that the type-I error of $\mathcal{M}$ becomes α. Then for any $\gamma > 0$, the upper bound of the type-II error of $\mathcal{M}$ is*

$$\Pr[\mathcal{M}(S,\hat{\tau}_\alpha) = \mathsf{acc} | H_1 \textit{ is true}] \leq \sup_{P\in\mathcal{P}} \left\{ \Pr_{S\sim P}[\mathcal{M}(S,\hat{\tau}_\alpha) = \mathsf{acc} | \chi^2(S) > \hat{\tau}_\alpha + \gamma] + \beta_{\hat{\tau}_\alpha+\gamma} \right\}.$$

The proof is shown in Appendix B. The upper bound consists of the probability term and the type-II error term. We discuss the behavior of these terms w.r.t. N.

The first probability term represents the probability that the privacy mechanism accepts the null hypothesis when the non-privacy-preserving test rejects it with threshold $\hat{\tau}_\alpha + \gamma$. For notational simplicity, we call the probability term the γ error:

$$E(\hat{\tau}_\alpha, \gamma, \mathcal{M}) = \sup_{P\in\mathcal{P}} \Pr_{S\sim P}[\mathcal{M}(S,\hat{\tau}_\alpha) = \mathsf{acc} | \chi^2(S) > \hat{\tau}_\alpha + \gamma].$$

The γ error measures how often the mechanism wrongly rejects the null hypothesis. The γ error thus depends on the mechanism $\mathcal{M}$. The analysis of this term will be discussed in the next subsection again.

The second term $\beta_{\hat{\tau}_\alpha+\gamma}$ is the type-II error of non-privacy-preserving test with threshold $\hat{\tau}_\alpha + \gamma$, which depends on the mechanism, too. This term becomes smaller if $\hat{\tau}_\alpha$ is closer to τ_α. This occurs when the distribution of $\widehat{\chi^2}(S)$ is close to the distribution of $\chi^2(S)$. In the case of output perturbation, this happens when the sensitivity of the mechanism decreases faster w.r.t. N. Thus, fixing sample size N, a greater power would be realized by employing a mechanism with a low sensitivity .

4.1. Power Analysis of Output Perturbation

First, we discuss the γ error of output perturbation.

Theorem 3. *Let $\mathcal{M}_\Delta$ be a ϵ-differentially private mechanism of output perturbation with sensitivity Δ. Then, the γ error is upper bounded by $\frac{1}{2}\exp\left(\frac{-\gamma\epsilon}{\Delta}\right)$.*

The proof is shown in Appendix C. Substituting the sensitivity derived in Eq.3 to this bound, we can confirm that the γ error of the mechanism of (Yu et al., 2014) is $O(1)$ in terms of N. The same conclusion is derived from the mechanisms of (Fienberg et al., 2011) and (Wang et al., 2015). The sensitivities employed by these mechanisms are also $O(1)$. A mechanism with a lower γ error and a lower sensitivity is needed to achieve greater power with ensuring DP.

4.2. Power Analysis of Input Perturbation

In the input perturbation method, randomization is applied to each cell. So the analysis of the γ error and the sensitivity cannot be appropriately derived. Because of this difficulty of analysis, we will evaluate the power of input perturbation numerically in Section 7.

5. Unit Circle Mechanism

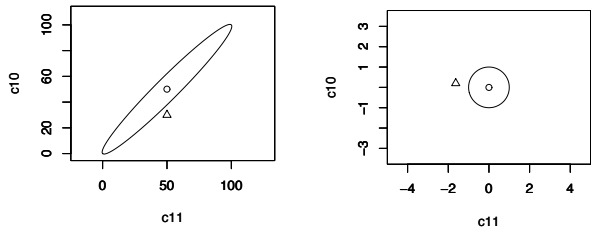
This section introduces a variant of output perturbation: the unit circle mechanism. This section first investigates the geometrical property. Then a novel mechanism is designed based on the property. We also show that the γ error and the sensitivity asymptotically vanish in the limit of N. Because of this property, the proposed mechanism achieves better power compared to existing output perturbation mechanisms.

5.1. Geometrical Interpretation of χ^2 Test

Given (c_{11}, c_{10}) and the marginals N_0 and N_1 in Table 1, the test statistic of Eq. 2 is represented as

$$\chi^2(c_{11}, c_{10}) = \frac{(c_{11}N_0 - c_{10}N_1)^2 N}{(c_{11}+c_{10})(N-c_{11}-c_{10})N_1N_0}. \quad (4)$$

Letting $\chi^2(c_{11}, c_{10}) = \tau_\alpha$ and rearranging Eq. 4 with respect to c_{11} and c_{10}, we have a quadratic form. The following lemma provides a geometrical interpretation of the test



(a) Before trans. (b) After trans.

Figure 1. Plot of contingency tables on the (c_{11}, c_{10})-plane before and after transformation. Circles and triangles respectively denote contingency tables with which the null hypothesis is rejected and accepted.

statistic of χ^2-test on the (c_{11}, c_{10})-plane.

Lemma 1. *Given $\tau_\alpha > 0$, for any $N_1 > 0, N_0 > 0$, and $N > 0$, $\chi^2(c_{11}, c_{10}) = \tau_\alpha$ forms an ellipse on the (c_{11}, c_{10})-plane.*

The proof is presented in Appendix D. Recall that the null hypothesis is rejected at significance level α if $\chi^2(S) > \tau_\alpha$. Using this, we can infer the following theorem.

Theorem 4. *Given Table 1 specified by S and a threshold $\tau_\alpha > 0$, H_0 is rejected if and only if (c_{11}, c_{10}) exists outside of the ellipse $\chi^2(c_{11}, c_{10}) = \tau_\alpha$.*

The proof, although it might be readily apparent, is omitted here. Figure 1(a) present plots of contingency tables with the ellipse. For the convenience of theoretical analysis, we introduce an affine transformation V, which transforms the ellipse to the unit circle. The formulation of the affine transformation V is shown in Appendix E.

Using this transformation, χ^2 test can be conducted by using $||V((c_{11}, c_{10})^t)||_2$ as the test statistic.

Theorem 5. *Given Table 1 specified by S and a threshold $\tau_\alpha > 0$, H_0 is rejected if and only if $||V((c_{11}, c_{10})^t)||_2 > 1$, i.e., $V((c_{11}, c_{10})^t)$ exists outside of the unit circle.*

The proof is presented in the Appendix F. Figure. 1(b) presents the plot of contingency tables and the ellipse after transformation.

5.2. Privacy Analysis of Unit Circle Mechanism

The input perturbation ensures DP by randomizing c_{10} and c_{11} independently by the Laplace or Gaussian mechanism. In the unit circle view, equivalent results are obtainable by randomizing the distance from the origin, $||V((c_{11}, c_{10})^t)||_2$, and judging whether the randomized distance is greater than 1 or not. The global sensitivity of $||V((c_{11}, c_{10})^t)||_2$ is derived as described below.

Lemma 2 (Sensitivity of $||V((c_{11}, c_{10})^t)||_2$)**.** *Given Table 1 specified by S and threshold $\tau_\alpha > 0$, the sensitivity of $||T((c_{11}, c_{10})^t)||_2$ is given as*

$$\Delta_{V,\alpha}(N_0, N_1) = 2\sqrt{\frac{(N_0^2 + N_1^2)N + 2\tau_\alpha N_0 N_1}{\tau_\alpha N_0 N_1 N^2}}. \quad (5)$$

The proof is presented in the Appendix G. This sensitivity analysis immediately derives the following Laplace mechanism:

$$\hat{d}(S) = ||V_{\tau_\alpha}((c_{11}, c_{10})^t)||_2 + \text{Lap}(\frac{\Delta_{V,\alpha}(N_0, N_1)}{\epsilon}). \quad (6)$$

The mechanism releasing $\hat{d}(S)$ ensures ϵ-DP whereas, as we already discussed in Section 3.4, if we use $\hat{d}(S)$ as the test statistic, the type-I error cannot be properly controlled. To maintain type-I error as α or less, we want $\hat{d}(S) < 1$ to hold with the probability of at least $1 - \alpha$ if H_0 holds. To attain this state, we propose the unit circle mechanism in Algorithm 1 that controls the type-I error by generating the finite sample distribution of $\hat{d}(S)$ with Monte-Carlo sampling. At line 2, the randomized test statistics is evaluated. The for-loop starting from line 4 generates the sample distribution of the randomized test statistics when the null hypothesis is true. At line 8, the p-value of S is evaluated with the sample distribution. The null hypothesis is rejected at line 10 if the p-value is less than the significance level.

Theorem 6. *Algorithm 1 ensures ϵ-DP.*

If marginals M_0, M_1, N_0, N_1 are public, then computation with S^k does not consume the privacy budget because S^k are samples that are artificially generated using the distribution specified by the public marginals. All operations after line 2 are attributable to post processing and therefore do not consume the privacy budget. Consequently, the privacy budget is consumed at line 2 only. Thus, the proof is immediately obtained by the privacy guarantee of the mechanism based on the global sensitivity(Dwork et al., 2006).

5.3. Utility Analysis of Unit Circle Mechanism

From the discussion in Section 4, the power of test mechanisms can be investigated by analyzing the γ error and the sensitivity. The γ error of Algorithm 1 is obtained using the following theorem.

Theorem 7. *The γ error of Algorithm 1 is*

$$\begin{aligned} & E(\hat{\tau_\alpha}, \gamma, M_{\Delta_{V,\alpha}(N_0,N_1)}) \\ \leq\ & \frac{1}{2}\exp\left(\frac{\epsilon N}{2}\left(1 - \sqrt{1 + \frac{4\gamma M_1 M_0}{\hat{\tau_\alpha} N^2}}\right)\right. \\ & \left.\cdot\sqrt{\frac{\hat{\tau_\alpha} N_1 N_0}{(N_1^2 + N_0^2)N + 2\hat{\tau_\alpha} N_1 N_0}}\right). \end{aligned} \quad (7)$$

The proof is shown in Appendix H. Eq. 7 is $O(\exp(-\sqrt{N}))$, and asymptotically vanishing as the sample size increases. Furthermore, the sensitivity of the unit

Algorithm 1 Unit Circle Mechanism

Require: Sample set S, sig level α, privacy budget ϵ
Evaluate contingency table from S
$\hat{d}(S) = ||V((c_{11}, c_{10})^t)||_2 + \text{Lap}(\frac{\Delta_{V,\alpha}(N_0,N_1)}{\epsilon})$
for $k = 1$ to m **do**
  $S^k \sim mult(\frac{N_1 M_1}{N^2}, \frac{N_0 M_1}{N^2}, \frac{N_1 M_0}{N^2}, \frac{N_0 M_0}{N^2})$
  Evaluate contingency table from S^k
  $\hat{d}(S^k) = ||V((c^k_{11}, c^k_{10})^t)||_2 + Lap(\frac{\Delta_{V,\alpha}(N^k_0,N^k_1)}{\epsilon})$
end for
$p = \frac{|\{i:\hat{d}(S)^k \geq \hat{d}(S)\}|}{m}$
if $p < \alpha$ **then**
  Return rej
else
  Return acc
end if

circle mechanism is $O(\frac{1}{\sqrt{N}})$ as derived from Eq. 5. Thus, from the discussion in Section 4, the upper bound of the type-II error of Algorithm 1 is expected to decrease with a faster rate than existing output perturbation method.

6. Differentially Private Multiple χ^2-test

Presume that we are interested in the independence between random variables X_0 and other K random variables, $X_1, X_2, \ldots, X_K$. The objective is investigation of the independence between X_0 and X_k for $k = 1, \ldots, K$. We denote the null hypothesis that X_0 and X_k are independent by H^k_0. The test statistic for independence between X_0 and X_k is calculated with the set of samples S^k. We suppose N_0, N_1, and N are the same for all S_k. We can verify the independence of each random variable pair by evaluating Eq. 2 with S^k for $k = 1, \ldots, K$ in turn. In multiple hypothesis testing, we consider to control the familywise error rate (FWER), the probability that the null hypothesis is rejected mistakenly at least once among K tests. If the significance of each test is kept 1- α, the FWER of the K tests in this setting is given as $1 - (1 - \alpha)^K \simeq \alpha K$, which increases as K increases. We use Bonferroni correction to correct the significance level so that the FWER for the entire test set is kept less than α (Bonferroni, 1936).

We can realize DP multiple χ^2-test using the DP χ^2 test mechanisms repeatedly. However, the privacy guarantee weakens as the number of hypothesis K increases by Theorem 1. (Fienberg et al., 2011) presented a multiple testing procedure using output perturbation. This method consumes a privacy budget that is proportional to K, which makes it almost impossible to obtain useful results under a meaningful privacy guarantee. (Johnson & Shmatikov, 2013; Yu et al., 2014; Simmons & Berger, 2016) presented a method which outputs the top s_1 significant random variable pairs using the exponential mechanism (McSherry & Talwar, 2007). This method outputs s_1 pairs even if all pairs are not significant. For that reason, naive application of exponential mechanism cannot control FWER.

In this section, we provide two differentially private multiple hypothesis testing methods that can conserve the privacy budget even with large K and controls FWER properly. The exponential mechanism based procedure works in the non-interactive setting, in which the set of hypotheses considered needs to be prescribed before starting the testing procedure. The SVT based procedure works in the interactive setting (Lyu et al., 2017), in which the hypothesis to be considered at each round can be interactively chosen after observing the result of the tests in the previous rounds (Webb & Petitjean, 2016).

6.1. Unit Circle Mechanism + SVT

We first show the SVT-based procedure with the unit circle mechanism (UCM+SVT). Recall that the unit circle mechanism verifies the result by checking if $||V((c_{11}, c_{10})^t)||_2 > 1$. Also, provided all the marginals are public information and N_0 and N_1 are the same for all S_k, the sensitivity of $||V((c_{11}, c_{10})^t)||_2$ is the same for any contingency table. Thus, the unit circle mechanism can be naturally incorporated into SVT.

Application of the unit circle mechanism (i.e., thresholding by Eq. 6) to SVT immediately guarantees differential privacy. However, the type-I error cannot be properly controlled by the naive combination. Recall that SVT adds noise to the threshold to compare in Eq. 1. In order to control the type-I error, the threshold of SVT should be properly adjusted considering the effect of the additive noise. We employ Monte Carlo sampling to determine the threshold that properly controls the type-I error. Also, we use Bonferoni correction to control the FWER.

The UCM+SVT takes as input sample sets $S^1, \cdots, S^K$, the significance level α, the privacy budget ϵ, and stop parameters $s_1 \leq s_2$. The UCN+SVT is terminated if (1) it rejects at most s_1 null hypotheses, or (2) it outputs s_2 test results. Here, we remark the stop parameter s_2 is not used in the regular SVT. In our setting, we need to apply Bonferroni correction to control the FWER. We thus use s_2 to upper bound the maximum number of hypotheses considered. The settings and the algorithm are detailed in Appendix I.

In the experimental results of single hypothesis test in Section 7, input perturbation+MC also achieves good performance. However, when input perturbation is used, the sensitivity of the test statistic is not uniform and thus it cannot be used with SVT.

6.2. Unit Circle Mechanism + EM

We introduce another DP multiple test procedure using the exponential mechanism (UCM+EM). By simple application of the exponential mechanism, we can get the top s_1 significant random variable pairs. However, this method outputs s_1 pairs even if all pairs are not significant. For that reason, it cannot control FWER. By adding s_1 significant dummy pairs of random variables, we can avoid accepting non-significant random variable pairs. However, this method does not necessarily control FWER properly because we cannot add appropriate dummy pairs to control FWER without knowing the scores related the p-values of non-significant pairs. In order to control FWER properly, we can use the exponential mechanism to select candidate pairs, and then apply the unit circle mechanism to the candidate pairs so that the significance level of each test is properly controlled. The settings and the algorithm is detailed in Appendix J.

7. Experiment

7.1. Experiments for Single Testing

In this section, we evaluate the significance and the power of the respective mechanisms, input perturbation (Gaboardi et al., 2016), output perturbation (Yu et al., 2014), and unit circle mechanism for single hypothesis testing. Both the output perturbation and the unit circle mechanism use N_1 and N_0 as public information. If these are public, the sensitivity of the input perturbation can be made 1. For controlling the significance of input perturbation, we use $\mathrm{MCIndep}_{\mathrm{Lap}}$ with the Laplace distribution in (Gaboardi et al., 2016). The significance of the unit circle mechanism and output perturbation is controlled by adjusting τ_α by Monte Carlo sampling. For Monte Carlo sampling, we set the number of sampling as 1000 for $\mathrm{MCIndep}_{\mathrm{Lap}}$ with the laplace distribution, and 10000 for the other methods. For output perturbation, we used the sensitivity derived in Eq.3.

Significance. To evaluate the significance, we sample 1000 contingency tables from $mult(0.25, 0.25, 0.25, 0.25)$ so that H_0 is true. $mult(\cdot,\cdot,\cdot,\cdot)$ denotes the multinomial distribution. Then, we assess the proportion that the mechanism outputs acc correctly. We set the privacy parameter as $\epsilon = 0.1$ and significance level as $\alpha = 0.05$.

In Figure 2, the significance of mechanisms are shown. If Monte Carlo sampling is not used (Figure 2(a)-2(c)), the significance remains poor for small samples. The significance of output perturbation unchanged even when the sample size increases. However, the input perturbation and the unit circle mechanism improve the significance as the sample size increases. The controlled significance version of these mechanisms can properly control the significance at 0.95 for any sample size (Figure 2(d)-2(f)).

Power. Evaluation of the power of the respective mechanisms controlling significance is presented. We sample 1000 contingency tables from $mult(0.25 + 0.01, 0.25 - 0.01, 0.25 - 0.01, 0.25 + 0.01)$ so that H_1 is true. To evaluate the power, we assess the rate at which the mechanism outputs rej correctly when H_1 is true. We set the privacy parameter $\epsilon = 0.1$ and significance level $\alpha = 0.05$.

Figure 3 shows that the unit circle mechanism with Monte Carlo sampling (UCM+MC) has similar power to that of the input perturbation with Monte Carlo sampling (IP+MC). In addition, compared with output perturbation with Monte Carlo sampling (OP+MC), UCM+MC quickly improves the power as the number of samples increases.

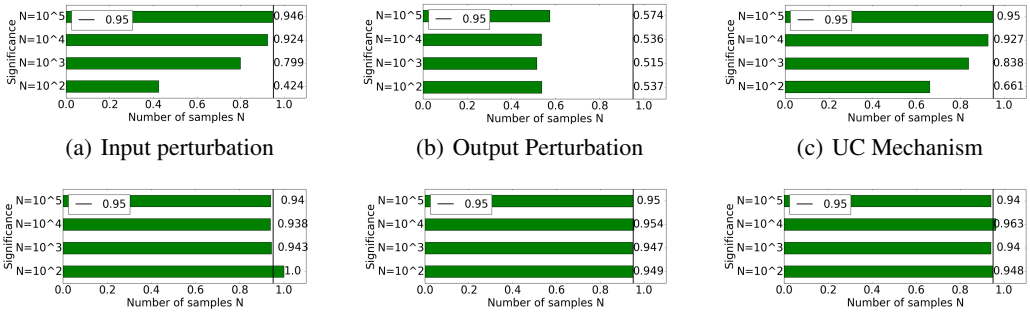
The γ-error of UCM+MC can be analyzed as discussed in Section 3.4. UCM+MC has a faster rate of the power than OP+MC because the γ-error of the UCM+MC decreases as the sample size increases. Unlike the UCM/OP+MC, the γ-error of IP+MC has never been analyzed. The UCM+MC is thus advantageous compared with the IP+MC in that the upper bound of type-II error is analyzed by Theorem 2. We remark that, looking at the results of the power, IP+MC and UCM+MC might have the same the γ-error rate. The analysis of IP+MC remains as future work.

7.2. Experiments for Multiple Testing

In this subsection, we experimentally evaluate FWER and utility of UCM+SVT and UCM+EM.

FWER. FWER can be properly controlled if the significance for a single test is exactly adjusted to α by using the Bonferroni correction. In UCM+EM, the significance of each test perfomed by UCM is properly controlled with a sufficiently large number of Monte Carlo samples. We can thus guarantee that FWER is properly controlled by applying Bonferroni correction. Since we can confirme this from Fig. 2, we skip experimental evaluation of the significance of UCM+EM. Similarly, UCM+SVT is expected to properly control FWER with a sufficiently large number of Monte Carlo samples. To confirm this, we evaluate the FWER of UCM+SVT experimentally.

In order to evaluate the FWER of UCM+SVT, all samples must be drawn from the null distribution. We artificially generated samples for evaluation in the following manner. We generated 1000 sample sets from $mult(0.25, 0.25, 0.25, 0.25)$ so that H_0 is true. Then, we assess the rate at which UCM+SVT outputs acc correctly. We set the privacy parameter as $\epsilon = 0.1$ and the significance level as $\alpha = 0.05$. We set the stop parameters as $s_1 = 1$ and $s_2 = 1$, and evaluate the significance $1 - \alpha$ for a single test instead of FWER. Note that we can control FWER by Bonferroni correction for multiple tests if the



(a) Input perturbation (b) Output Perturbation (c) UC Mechanism

(d) Input perturbation + MC (e) Output Perturbation+MC (f) UC Mechanism +MC

Figure 2. Plot of significance. The significance becomes 0.95 if properly controled.

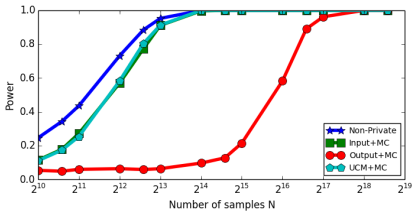


Figure 3. Change of the power with respect to the sample size N.

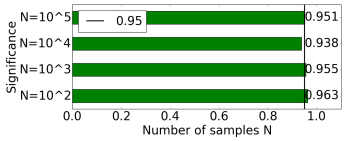


Figure 4. plot of FWER.

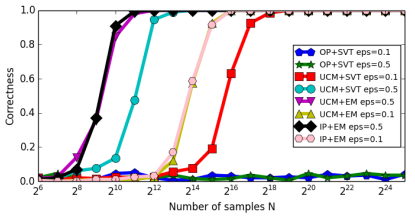


Figure 5. Plot of the correct rate.

mechanism can guarantee the significance of a single test. Also, we remark that this does not mean that our mechanism cannot work with multiple sample sets.

Figure 4 shows the significance level of UCM+SVT when the sample size is changed. As we can see from the figure, UCM+SVT can control the significance at the prescribed level (0.95) for any sample size.

Utility. Next, we compare the utility of UCM+SVT and UCM+EM with existing methods. We denote the SVT with the output perturbation by OP+SVT, the EM with the input perturbation by IP+EM. We compare the correctness of these algorithms with respect to the sample size.

For evaluation, we artificially generated $K = 10$ sample sets in the following method. We fix the marginals $N_1^k = \frac{N}{2}, N_0^k = \frac{N}{2}, M_1^k = \frac{N}{2}, M_0^k = \frac{N}{2}$, and then create the sample sets $S^1, \ldots, S^K$ so that the test statistics of the resulting sample sets become $\chi^2(S^k) \in \{1, 1, 1, 1, 1, 1, 1, 1, 30, 30\}$. Two of them reject the corresponding null hypotheses at significance level $\alpha = 0.05$ after Bonferroni correction (i.e., $\chi^2(S^k) = 30$). Regarding the remaining eight sample sets, the corresponding null hypotheses are accepted. We input S^k to the algorithms, and then evaluate the utility. For utility measure, we employ the correctness $\frac{r}{r'}$, where r is the number that the algorithm outputs rej correctly; r' is the number of the contingency tables rejected by non-private multiple χ^2 tests with Bonferroni correction. In our artificially generated sample sets, $r' = 2$. For UCM+SVT, we input the sample sets in a random order. We set the privacy parameter as $\epsilon = \{0.1, 0.5\}$, the significance levels as $\alpha = 0.05$, and the stop parameters as $s_1 = 2, s_2 = 10$.

Figure 5 shows the correctness (average of 100 trials) of each algorithm with respect to the sample size. As we see from the figure, the correctness of UCM+SVT approaches to 1 when N increases while the correctness of OP+SVT does not. This is because the sensitivity of the unit circle mechanism decreases faster than that of output perturbation with respect to the sample size. Both UCM+EM and IP+EM have the comparable correctness. We remark that UCM is advantageous compared to IP in the sense that the type-II error of UCM is upper-bounded as we discuss in Section 5, while no theoretical guarantee on the type-II error of input perturbation is provided.

The algorithms using EM achieves better correctness than the algorithms using SVT. This difference of the correctness comes from the settings that SVT and EM can handle. As we discussed in Section 6, SVT can deal with the interactive setting whereas EM considers the non-interactive setting only. In this sense, the results of these two procedures are not directly comparable.

Acknowledgement

We thank anonymous reviewers for insightful reviews and discussions. We appreciate that one of the reviewers gives an important suggestion regarding the combination of the unit circle mechanism and exponential mechanism. We thank Prof. Hideitsu Hino and Dr. Masayuki Terada for important comments that greatly improved the manuscript. The research is partly supported by JST CREST Grant Number JPMJCR1302 and JSPS Grand-in-Aid for Scientific Research No. 16H02864.

References

Barak, Boaz, Chaudhuri, Kamalika, Dwork, Cynthia, Kale, Satyen, McSherry, Frank, and Talwar, Kunal. Privacy, accuracy, and consistency too: a holistic solution to contingency table release. In *Proceedings of the twenty-sixth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*, pp. 273–282, 2007.

Bishop, Yvonne M.M, Fienberg, Stephen E, and Holland, Paul W. Discrete multivariate analysis: Theory and practice. 1975.

Bonferroni, C.E. *Teoria statistica delle classi e calcolo delle probabilità*. Pubblicazioni del R. Istituto superiore di scienze economiche e commerciali di Firenze. Libreria internazionale Seeber, 1936.

Cai, Bryan, Daskalakis, Constantinos, and Kamath, Gautam. Priv'it: Private and sample efficient identity testing. In *Proceedings of The 34rd International Conference on Machine Learning*, pp. to appear, 2017.

Dwork, Cynthia and Roth, Aaron. The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science*, 9(3-4):211–407, 2014.

Dwork, Cynthia, McSherry, Frank, Nissim, Kobbi, and Smith, Adam. Calibrating noise to sensitivity in private data analysis. In *Proceedings of the Third Conference on Theory of Cryptography*, pp. 265–284, 2006.

Fienberg, Stephen E., Slavkovic, Aleksandra, and Uhler, Caroline. Privacy preserving gwas data sharing. In *Proceedings of the 2011 IEEE 11th International Conference on Data Mining Workshops*, pp. 628–635, 2011.

Gaboardi, Marco, Arias, Emilio Jesús Gallego, Hsu, Justin, Roth, Aaron, and Wu, Zhiwei Steven. Dual query: Practical private query release for high dimensional data. In *Proceedings of The 31rd International Conference on Machine Learning*, pp. 1170–1178, 2014.

Gaboardi, Marco, Lim, Hyun, Rogers, Ryan, and Vadhan, Salil. Differentially private chi-squared hypothesis testing: Goodness of fit and independence testing. In *Proceedings of The 33rd International Conference on Machine Learning*, pp. 2111–2120, 2016.

Hardt, Moritz, Ligett, Katrina, and McSherry, Frank. A simple and practical algorithm for differentially private data release. In *Advances in Neural Information Processing Systems 25*, pp. 2339–2347, 2012.

Homer, Nils, Szelinger, Szabolcs, Redman, Margot, Duggan, David, Tembe, Waibhav, Muehling, Jill, Pearson, John V, Stephan, Dietrich A, Nelson, Stanley F, and Craig, David W. Resolving individuals contributing trace amounts of dna to highly complex mixtures using high-density snp genotyping microarrays. *PLoS genetics*, 4 (8):e1000167, 2008.

Johnson, Aaron and Shmatikov, Vitaly. Privacy-preserving data exploration in genome-wide association studies. In *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1079–1087, 2013.

Lee, Jaewoo, Wang, Yue, and Kifer, Daniel. Maximum likelihood postprocessing for differential privacy under consistency constraints. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 635–644, 2015.

Li, Chao and Miklau, Gerome. An adaptive mechanism for accurate query answering under differential privacy. In *Proceedings of the VLDB Endowment*, volume 5, pp. 514–525, 2012.

Li, Chao, Hay, Michael, Rastogi, Vibhor, Miklau, Gerome, and McGregor, Andrew. Optimizing linear counting queries under differential privacy. In *Proceedings of the twenty-ninth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*, pp. 123–134, 2010.

Lyu, Min, Su, Dong, and Li, Ninghui. Understanding the sparse vector technique for differential privacy. In *Proceedings of the VLDB Endowment*, volume 10, pp. 637–648, 2017.

McSherry, Frank and Talwar, Kunal. Mechanism design via differential privacy. In *Proceedings of the 48th Annual IEEE Symposium on Foundations of Computer Science*, pp. 94–103, 2007.

Simmons, Sean and Berger, Bonnie. Realizing privacy preserving genome-wide association studies. *Bioinformatics*, 32(9):1293–1300, 2016.

Uhler, Caroline, Slavković, Aleksandra, and Fienberg, Stephen E. Privacy-preserving data sharing for genome-wide association studies. *The Journal of privacy and confidentiality*, 5(1):137, 2013.

Wang, Yue, Lee, Jaewoo, and Kifer, Daniel. Differentially private hypothesis testing, revisited. *arXiv preprint arXiv:1511.03376*, 2015.

Webb, Geoffrey I and Petitjean, François. A multiple test correction for streams and cascades of statistical hypothesis tests. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1255–1264, 2016.

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