RESEARCH REPORT SERIES
(Statistics #2019-04)

Minimax Randomized Response Methods for Providing Local Differential Privacy

Jichong Chai[1],
Tapan K. Nayak[2]

[1]Department of Statistics, George Washington University
[2]Center for Statistical Research and Methodology, U.S. Census Bureau,
and Department of Statistics, George Washington University

Center for Statistical Research & Methodology
Research and Methodology Directorate
U.S. Census Bureau
Washington, D.C. 20233

Report Issued: March 20, 2019

Minimax Randomized Response Methods for Providing Local Differential Privacy

Jichong Chai* and Tapan K. Nayak†‡

Abstract

Randomized response (RR) is a common privacy protection tool. It perturbs true responses using a probabilistic mechanism. Local differential privacy (LDP) is a rigorous privacy protection criterion that demands a guarantee that no intruder will get much new information about any respondent's true value from its perturbed value. Considering linear unbiased estimation of multinomial probabilities under LDP and squared error loss, we derive minimax RR methods. We address optimal choices for both the RR mechanism (or design) and the estimator. We obtain a minimax design, which has a specific structure and is termed a t-subset design. We describe and study properties of t-subset designs including their practical implementation. We also study mixtures of t-subset designs and examine the RAPPOR method, which is used notably by Google and Apple. We note inadmissibility of the RAPPOR design and offer some suggestions for improving both the design and the customary estimator.

Key words and Phrases: Admissibility; categorical data; linear unbiased estimator; RAPPOR algorithm; squared error loss; transition probability matrix.

*Department of Statistics, George Washington University, Washington, DC 20052.

†Center for Statistical Research and Methodology, U.S. Census Bureau, Washington, DC 20233 and Department of Statistics, George Washington University, Washington, DC 20052.

‡The views expressed in this article are those of the authors and not necessarily those of the U.S. Census Bureau.

1. Introduction

Warner (1965) proposed the first randomized response (RR) method for protecting respondents' privacy in interview surveys of sensitive dichotomous variables. Greenberg et al. (1969) proposed another method that uses unrelated non-sensitive questions. Numerous papers have extended and generalized the basic ideas of these two papers to propose various RR methods for using in surveys of categorical and quantitative variables. The books Chaudhuri and Mukerjee (1988), Chaudhuri (2011) and Fox (2016) review various methods and cite important works.

We shall consider only categorical survey variables. Thus, consider a categorical survey variable X and let $\mathcal{S}_X = \{c_1, \ldots, c_k\}$ denote its sample space. An RR procedure converts each true response into a randomized output. The randomization mechanism is predetermined and is applied independently to each true response. Conceptually, it consists of an output space and a set of probabilities for changing the true responses. For an RR procedure, let Z and $\mathcal{S}_Z = \{d_1, \ldots, d_m\}$ denote the output variable and its sample space, respectively. Also, let $p_{ij} = P(Z = d_i | X = c_j), i = 1, \ldots, m, j = 1, \ldots, k$, denote the transition probabilities, which are preset and are embedded in the randomization mechanism. Thus, for a true response c_j, the RR method outputs d_i with probability p_{ij}. The resultant values of Z constitute the data. Any RR mechanism is characterized by its transition probability matrix (TPM), to be denoted $P = ((p_{ij}))$, as it determines all effects of randomization on both privacy and data utility. Thus, P represents the RR design. In general, the two sample spaces $\mathcal{S}_X$ and $\mathcal{S}_Z$ need not be identical, or even have the same cardinality. Thus, P may not be a square matrix. For example, $m = 2^k$ in the RAPPOR algorithm of Erlingsson et al. (2014), further discussed in Section 4.

Let $\pi_i = P(X = c_i), i = 1, \ldots, k$, and $\pi = (\pi_1, \ldots, \pi_k)'$, which are unknown. Typically, the primary goal of a survey is to make inferences about π. We shall denote the sample size by n and the sample frequency of d_i by S_i, for $i = 1, \ldots, m$. We shall assume random sampling, in which case the frequency vector $S = (S_1, S_2, \ldots, S_m)'$ has a multinomial distribution, viz.

$S \sim mult(n, \lambda)$, where

$$\lambda_{m\times 1} = P_{m\times k}\pi_{k\times 1}. \tag{1.1}$$

The relative frequency vector $\hat{\lambda} = S/n$ is a natural (and method of moments) estimator of λ. Usually, inferences about λ can be translated into inferences about π, via (1.1). For example, if P is square and nonsingular, an estimator $\tilde{\lambda}(S)$ of λ gives the estimator $\tilde{\pi} = P^{-1}\tilde{\lambda}(S)$ for π. If $rank(P) < k$, then the distribution of S is not identifiable with respect to π and hence π is not estimable. So, for estimability, we shall require that $m \geq k$ and $rank(P) = k$.

For many years, mainly statisticians worked on RR theory and methods and for protecting privacy in interview surveys. More recently, privacy concerns have expanded significantly, largely in reaction to pervasive data collection from surveys, administrative records, customer information, on-line activities etc. That has stimulated strong interest in privacy research in other fields, especially in computer science; see e.g., Agrawal and Srikant (2000), Rizvi and Harista (2002), Aggarwal and Yu (2008), Chen et al. (2009) and Fung et al. (2010). In particular, the interest in RR methods has increased and for a wider range of applications. Evfimievski et al. (2004), Aggarwal et al. (2009) and Erlingsson et al. (2014) and others have suggested RR methods for addressing privacy challenges in emerging contexts such as privacy preserving data mining and on-line data collection.

Designing an RR mechanism reduces to choosing m and P suitably, taking both privacy protection and accuracy of inferences about π from RR data into account. However, many papers compared RR methods by comparing only the variance of estimators. That is misleading. A fair comparison should also require a common level of privacy protection. Some authors did that, but mainly for binary characteristics, see e.g., Anderson (1976), Fligner et al. (1977), Nayak (1994) and Nayak and Adeshiyan (2009). For a general categorical variable, the RR literature does not give much guidance on how to choose m and P. We believe that the choice of P could not be addressed properly because privacy measures and precise privacy protection goals were not

developed until very recently.

Remark 1.1. *The RR theory needs to address two questions: (i) how to randomize true responses? and (ii) how to make inferences from RR data? We shall use "RR mechanism" and "RR process" to refer to how randomization is done. Its mathematical crux is P, the transition probability matrix, which we shall call the "RR design." We shall call a pair of P and an estimator an "RR method" or "RR strategy" (analogous to sampling strategy). Evidently, only P affects privacy whereas estimation accuracy is affected by both P and the estimator.*

The following is a recently introduced privacy criterion (see, Duchi et al., 2018).

Definition 1.1. *An RR design provides ϵ-differentially local privacy (ϵ-DLP), for $\epsilon > 0$, if*

$$\sup_{B \subseteq \mathcal{S}_Z} \sup_{c_i, c_j \in \mathcal{S}_X} \frac{P(Z \in B | X = c_i)}{P(Z \in B | X = c_j)} \leq exp(\epsilon). \tag{1.2}$$

This criterion has been investigated and used by Kairouz et al. (2016b), Wang et al. (2016), Duchi et. al. (2018), Ye and Barg (2018) and others. A helpful interpretation of ϵ-DLP comes from an equivalency between ϵ-DLP and the *strict information privacy* (SIP) criterion of Chai and Nayak (2018). To describe SIP, we let $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_k)$ denote an intruder's (subjective) prior distribution for a respondent's true value of X, and for any $Q \subseteq \mathcal{S}_X$, let $P_\alpha(Q)$ and $P_\alpha(Q|d_i)$ denote, respectively, the intruder's prior and posterior probabilities of $\{X \in Q\}$, given $Z = d_i$.

Definition 1.2. *(Chai and Nayak, 2018). Let h_l and h_u be two functions from $[0,1]$ to $[0,1]$ such that $0 \leq h_l(a) \leq a \leq h_u(a) \leq 1$ for all $0 \leq a \leq 1$. An RR design is said to provide strict information privacy (SIP) with respect to h_l and h_u if for all $\alpha, Q \subseteq \mathcal{S}_X$ and $i = 1, \ldots, m$,*

$$h_l(P_\alpha(Q)) \leq P_\alpha(Q|d_i) \leq h_u(P_\alpha(Q)). \tag{1.3}$$

The SIP criterion formalizes the notion that an RR procedure should guarantee that any

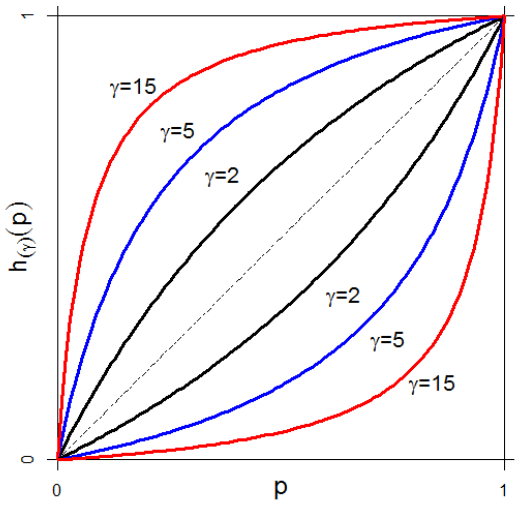

Figure 1: Plots of privacy breach boundaries

posterior probability would be "close" to the corresponding prior probability, as specified by h_l and h_u. The ρ_1-to-ρ_2 privacy (Evfimievski et al., 2003) and β-factor privacy (Nayak et al., 2015) are special cases of Definition 1.2. A relevant result that follows from Chai and Nayak (2018) is that an RR procedure provides ϵ-DLP if and only if

$$\frac{P_\alpha(Q)}{1+(\gamma-1)(1-P_\alpha(Q))} \leq P_\alpha(Q|d_i) \leq \frac{\gamma P_\alpha(Q)}{1+(\gamma-1)P_\alpha(Q)} \tag{1.4}$$

for all α, Q and d_i, where $\gamma = \exp(\epsilon)$. Figure 1 shows the upper and lower boundaries in (1.4) for some values of γ. For each γ, the points inside the region enclosed by the two curves satisfy privacy. Any prior-posterior pair falling outside the region constitutes a privacy breach. As expected, the privacy breach region increases as γ decreases. A useful characterization of all P that satisfy ϵ-DLP involves the following.

Definition 1.3. *(Nayak et al., 2015). The ith row parity of P is defined as*

$$\eta_i(P) = \max\left\{\frac{p_{ij}}{p_{il}} \;\middle|\; j,l = 1,\ldots,k\right\} = \frac{\max_j\{p_{ij}\}}{\min_j\{p_{ij}\}},$$

with the convention $0/0 = 1$ and $a/0 = \infty$ for any $a > 0$. Furthermore, the parity of P is defined as $\eta(P) = \max_i\{\eta_i(P)\}$.

It can be seen that an RR design P provides ϵ-DLP if and only if

$$\eta(P) \leq \gamma = exp(\epsilon). \tag{1.5}$$

The SIP criterion also yields the constraint in (1.5), because (1.3) is equivalent to (1.5) with a suitable γ, determined by h_l and h_u (see Chai and Nayak, 2018). Thus, to find an optimal ϵ-DLP satisfying RR procedure, we should maximize estimation accuracy subject to (1.5). In Section 2, we motivate and formulate a precise problem. Specifically, we shall consider finding an RR strategy that is minimax under squared error loss for estimating π, subject to (1.5), unbiasedness and linearity. This problem is solved in Section 3. An optimal design has a specific balanced structure, called a t-subset design, and an estimator that minimizes the risk under $\pi_1 = \cdots = \pi_k$ is a minimax estimator. In Section 4, we discuss a convenient approach for implementing our minimax strategy. We also examine mixtures of t-subset designs, which includes the RAPPOR method of Erlingsson et al. (2014) that is in use by Google and Apple. We derive minimax estimators under mixture designs and show that for RAPPOR design, it is uniformly better than the customary estimator. Thus, the RAPPOR method can be improved by changing the design and/or the estimator. Numerical calculations show that moderate to low privacy domain ($\gamma \geq 6$) statistical efficiency gains can be substantial. Section 5 contains some concluding remarks.

2. Optimality Criteria

To find an optimal RR design at a specified privacy level γ (> 1), one should try to find a P in $\{P : \eta(P) \leq \gamma\}$ that maximizes "data utility." To formulate this idea, one needs to specify a measure of data utility. Also, as the data may be used for various purposes by diverse users, one should use a widely suitable measure. Chai and Nayak (2018) explored a general criterion, based on Blackwell's (1951, 1953) concept of sufficiency of experiments.

Definition 2.1. *An RR design $P_{m\times k}$ is said to be sufficient for (or at least as informative as) another RR design $A_{r\times k}$, to be denoted $P \succeq A$, if there exists a transition probability matrix $C_{r\times m}$ such that $A = CP$.*

If $P \succeq A$ and also $A \succeq P$, then A and P are equivalent, and P is better than A if $P \succeq A$ but $A \not\succeq P$. Furthermore, P is said to be admissible if there does not exist any better design A.

The above criterion is appealing because $P \succeq A$ implies that under any loss function, for any inference rule δ based on the data from A, there exists a rule δ_* based on P such that the risk of δ_* is no larger than the risk of δ. In this sense, if $P \succeq A$, then P is universally at least as good as A. Two natural restrictions on P are: (i) each row of P must have at least one nonzero value (otherwise the corresponding response is irrelevant) and (ii) no two rows of P can be proportional to each other (see, Chai and Nayak, 2018). With these two conditions, a characterization of all admissible RR designs is:

Theorem 2.1. *(Chai and Nayak, 2018). For a given privacy level γ, an RR design P is admissible if and only if (i) $\eta_i(P) = \gamma$ for all i (i.e., each row parity is γ) and (ii) each row of P contains exactly two distinct values.*

Let $\mathcal{C}_\gamma^a$ denote the class of all admissible P at privacy level γ. Generally, $\mathcal{C}_\gamma^a$ is large and sufficiency does not yield a best procedure. So, to find optimum designs and strategies we need additional criteria and measures of data utility. Examples of such criteria and some related results

can be found in Agrawal et al. (2009), Kairouz et al. (2016b), Duchi et al. (2018) and Chai and Nayak (2018).

In this paper, we shall explore optimum RR strategies for linear unbiased estimation of π under squared error loss. Specifically, we shall consider only unbiased estimators that are linear in S, or equivalently linear in $\hat{\lambda} = S/n$. A linear estimator $\hat{\pi} = L\hat{\lambda}$ is unbiased, i.e., $E(L\hat{\lambda}) = \pi$ or $LP\pi = \pi$ for all π, if and only if $LP = I$, which can hold only if $r(P) = k$ (and $m \geq k$). Conversely, if $r(P_{m\times k}) = k$, there exists $L_{k\times m}$ such that $LP = I$. Thus, we must restrict our attention to RR designs P with $r(P) = k$. Adopting squared error loss, we define the risk function of a linear unbiased RR strategy (P, L) as

$$\mathbf{R}(P, L; \pi) = nE_{P,\pi}\left[\|L\hat{\lambda} - \pi\|^2\right] = n[tr(\mathbf{V}_{P,\pi}(L\hat{\lambda}))] = tr(LD_\lambda L') - \sum_{i=1}^{k} \pi_i^2, \tag{2.1}$$

where $\lambda = P\pi$, for a vector $v = (v_1, \ldots, v_k)'$, D_v denotes the diagonal matrix with diagonal elements $v_1, \ldots, v_k$, the expectation is with respect to both sampling and randomization. In (2.1), the multiplier n normalizes the risk for sample size.

Note that the conclusions of Theorem 2.1 hold also under the added restriction $r(P) = k$. If $r(P) = k$ and P is inadmissible, it follows easily that there exists $A \in \mathcal{C}_\gamma^a$ such that $r(A) = k$ and $P = CA$ for some TPM C, i.e., there exists a more informative design A with $r(A) = k$. Thus, we shall restrict our attention to $\mathcal{C}_\gamma^1$, the class of all admissible procedures P with $r(P) = k$. To be precise, $\mathcal{C}_\gamma^1$ consists of all $P_{m\times k}, m \geq k$, satisfying the conditions

C1: $r(P) = k$.

C2: No two rows of P are proportional to each other.

C3: $\eta_i(P) = \gamma$ for $i = 1, \ldots, m$.

C4: Each row of P contains two distinct values.

Note that C3 implies that all elements of P must be positive. A natural goal is to find $P \in \mathcal{C}_\gamma^1$ and an L such that the risk in (2.1) is minimum. First, consider minimizing (2.1) with respect to

L, for given P. If P is square and nonsingular, then $P^{-1}\hat{\lambda}$ is the unique linear unbiased estimator of π (see Chaudhuri and Mukerjee, 1988), hence the optimal L is P^{-1}. For any $P_{m\times k} \in \mathcal{C}_\gamma^1$ with $m > k$, unbiased L is not unique and the following result gives locally optimal linear unbiased estimators.

Proposition 2.1. *For any given $P \in \mathcal{C}_\gamma^1$ and π,*

$$\boldsymbol{R}(P, L; \pi) \geq tr(P' D_\lambda^{-1} P)^{-1} - \sum_{i=1}^{k} \pi_i^2 \tag{2.2}$$

for all L such that $LP = I$, and the lower bound is attained when

$$L = (P' D_\lambda^{-1} P)^{-1} P' D_\lambda^{-1} = L_*, \ \textit{say.} \tag{2.3}$$

Proof. For given $P \in \mathcal{C}_\gamma^1$, take any L such that $LP = I$. In view of (2.1), it suffices to show that $tr(LD_\lambda L') \geq tr(P' D_\lambda^{-1} P)^{-1}$. Let $U = D_\lambda^{-1/2} P$ and $U^- = L D_\lambda^{1/2}$. Then, $U^- U = I$ and $U'U = P' D_\lambda^{-1} P$, and thus

$$\begin{aligned} LD_\lambda L' = U^-(U^-)' &= \Big(U^- - (U'U)^{-1}U'\Big)\Big(U^- - (U'U)^{-1}U'\Big)' + (U'U)^{-1} \\ &= \Big(U^- - (U'U)^{-1}U'\Big)\Big(U^- - (U'U)^{-1}U'\Big)' + (P' D_\lambda^{-1} P)^{-1}. \end{aligned} \tag{2.4}$$

Now, (2.4) shows that $LD_\lambda L' - (P' D_\lambda^{-1} P)^{-1}$ is non-negative definite and thus $tr(LD_\lambda L') \geq tr(P' D_\lambda^{-1} P)^{-1}$. Moreover, the equality holds if and only if $U^- - (U'U)^{-1}U' = 0 \Leftrightarrow L = (P' D_\lambda^{-1} P)^{-1} P' D_\lambda^{-1}$. □

Remark 2.1. *Clearly, L_* depends on P and π, but for notational simplicity we do not write that explicitly. The proof of Proposition 2.1 shows that $L_*\hat{\lambda}$ is locally best also under D- and E-optimality criteria. Also, (2.2) holds more generally for any P (not limited to $\mathcal{C}_\gamma^1$) and π such that $r(P) = k$ and all elements of $P\pi$ are positive.*

The optimum L in (2.3) depends on π, unless P is square and non-singular. So, a uniformly minimum risk estimator among all linear estimators does not exist. This also shows that an RR strategy (P, L) with uniformly minimum risk does not exist. As an alternative, we shall use minimaxity to find an optimality RR strategy. Specifically, we shall try to find a strategy (P_0, L_0) such that $P_0 \in \mathcal{C}_\gamma^1$, $L_0 P_0 = I$ and

$$\inf_{P \in \mathcal{C}_\gamma^1} \inf_L \sup_\pi \mathbf{R}(P, L; \pi) = \sup_\pi \mathbf{R}(P_0, L_0; \pi). \tag{2.5}$$

In (2.5), for brevity, we do not show the requirement $LP = I$ explicitly. The left side of (2.5) is the minimax value. Duchi et al. (2018) considered a similar approach and derived some asymptotic results for a general class of loss functions. In particular, they obtained minimax rates of convergence for several estimation problems and loss function. In contrast, we shall derive exact minimax procedures, but under linearity, unbiasedness and squared error loss.

3. Derivation of Minimax Methods

To find a minimax strategy (P, L), we shall first find (P_0, L_0) that minimizes (2.1) at $\pi = (1/k, \ldots, 1/k) = \pi_u$, say, i.e.,

$$\inf_{P \in \mathcal{C}_\gamma^1} \inf_L \mathbf{R}(P, L; \pi_u) = \mathbf{R}(P_0, L_0; \pi_u). \tag{3.1}$$

Then, we shall prove that the solution (P_0, L_0) satisfies (2.5). In a sense, the degenerate distribution at π_u is least favorable.

In view of Proposition 2.1, solving (3.1) reduces to finding $P \in \mathcal{C}_\gamma^1$ such that $tr(P' D_\lambda^{-1} P)^{-1}$ is minimum, where $\lambda = P\pi_u$. Note that $P' D_\lambda^{-1} P$ is a symmetric positive definite matrix, and let

$\alpha_1 \leq \alpha_2 \leq \cdots \leq \alpha_k$ denote its eigenvalues. Then,

$$tr(P'D_\lambda^{-1}P)^{-1} = \sum_{i=1}^{k} \frac{1}{\alpha_i}, \tag{3.2}$$

which suggests that we should try to make α_i's large as we search for an optimum P, viz. a minimizer of (3.2). However, α_i's cannot be arbitrarily large, as they must satisfy certain restrictions. Note that here $\lambda = P\pi_u = k^{-1}P\mathbf{1}_k$, where $\mathbf{1}_k$ denotes the vector of dimension k whose all components are 1. So, $P'D_\lambda^{-1}P\mathbf{1}_k = kP'D_{(P\pi_u)}^{-1}P\pi_u = kP'\mathbf{1}_m = k\mathbf{1}_k$. Thus, for $\pi = \pi_u$, $(P'D_\lambda^{-1}P)/k$ is a doubly stochastic matrix and hence the dominant eigenvalue of $P'D_\lambda^{-1}P$ is $\alpha_k = k$ (Lax, 2007, p.241). Moreover, $\sum_{i=1}^{k} \alpha_i = tr(P'D_\lambda^{-1}P)$ has a tight upper bound, as the following lemma shows. Recall that conditions C3 and C4 imply that for any $P \in \mathcal{C}_\gamma^1$, each row of P contains two distinct values and the ratio of the largest to smallest values is γ (> 1). Subsequently, we shall refer to the smaller (larger) of the two values as the small (large) value.

Lemma 3.1. *For given $\gamma > 1$ and $k \geq 2$, let*

$$f(x) = \frac{k^2(x\gamma^2 + k - x)}{(x\gamma + k - x)^2}, \quad x \geq 0, \tag{3.3}$$

and

$$q = \begin{cases} \lfloor \frac{k}{1+\gamma} \rfloor, & \text{if } f(\lfloor \frac{k}{1+\gamma} \rfloor) \geq f(\lceil \frac{k}{1+\gamma} \rceil) \text{ and } \lfloor \frac{k}{1+\gamma} \rfloor \geq 1 \\ \lceil \frac{k}{1+\gamma} \rceil, & \text{otherwise.} \end{cases} \tag{3.4}$$

Then, for all $P \in \mathcal{C}_\gamma^1$,

$$tr(P'D_\lambda^{-1}P) \leq f(q), \tag{3.5}$$

and the equality holds if each row of P contains exactly q large values.

Proof. Take any $P_{m\times k} \in \mathcal{C}_\gamma^1$ and let $P\pi_u = \lambda = (\lambda_1, \ldots, \lambda_m)'$. Then,

$$tr(P'D_\lambda^{-1}P) = tr(D_\lambda^{-1}PP') = \sum_{i=1}^{m} \frac{1}{\lambda_i} \sum_{j=1}^{k} p_{ij}^2. \tag{3.6}$$

Recalling that each row of P contains two distinct values, for $i = 1, \ldots, m$, let s_i denote the 'small' value in the ith row of P, and so the 'large' value is γs_i. Also, suppose that the ith row contains q_i large values and $(k - q_i)$ small values, with $1 \le q_i \le k-1$. Note that $\lambda_i = (1/k)\sum_{j=1}^{k} p_{ij}$, as π_u is uniform. This implies that $q_i\gamma s_i + (k-q_i)s_i = k\lambda_i$ or $s_i = (\lambda_i k)/(q_i\gamma + k - q_i)$. Thus,

$$\sum_{j=1}^{k} p_{ij}^2 = q_i(\gamma s_i)^2 + (k-q_i)s_i^2 = \lambda_i^2 \frac{k^2(q_i\gamma^2 + k - q_i)}{(q_i\gamma + k - q_i)^2} = \lambda_i^2 f(q_i). \tag{3.7}$$

Combining (3.6) and (3.7), we get

$$tr(P'D_\lambda^{-1}P) = \sum_{i=1}^{m} \lambda_i f(q_i).$$

Taking derivative, it can be seen that as x increases, $f(x)$ first increases and then decreases, reaching its maximum at $x = k/(1+\gamma)$. Then, for $x \in \{1, \ldots, k-1\}$, it can be seen that $f(x)$ is maximized at q, as defined in (3.4). So,

$$\sum_{i=1}^{m} \lambda_i f(q_i) \le (\sum_{i=1}^{m} \lambda_i) f(q) = f(q),$$

which establishes (3.5). Clearly, the upper bound is attained if all rows of P contain exactly q large values. □

Remark 3.1. *If $f(q) \neq f(q+1)$, the upper bound is attained if and only if $q_i = q$, irrespective of the small value s_i in each row.*

Now, minimizing (3.2) reduces to minimizing $\sum_{i=1}^{k-1}(1/\alpha_i)$, subject to $\sum_{i=1}^{k-1}\alpha_i \le f(q) - k$, and

$\alpha_i > 0, i = 1, \ldots, k-1$, as $\alpha_k = k$. It can be seen easily that $\sum_{i=1}^{k-1}(1/\alpha_i)$ is a strictly Schur-convex function on $\Delta = \{(\alpha_1, \ldots, \alpha_{k-1}) : \sum_{i=1}^{k-1} \alpha_i = f(q) - k, \alpha_i > 0, i = 1, \ldots k-1\}$. So, $\sum_{i=1}^{k-1}(1/\alpha_i)$ is minimized over Δ if and only if $\alpha_i = [f(q) - k]/(k-1), i = 1, \ldots, k-1$ (Marshall et al., 2011). Now, the following conclusion can be reached readily.

Lemma 3.2. *A lower bound for* $tr(P'D_\lambda^{-1}P)^{-1}$ *is* $\dfrac{(k-1)^2}{f(q)-k} + \dfrac{1}{k}$, *and it is attained if and only if the eigenvalues of* $P'D_\lambda^{-1}P$ *are*

$$\alpha_k = k \quad \textit{and} \quad \alpha_i = \frac{f(q)-k}{k-1} \textit{ for } 1 \le i \le k-1.$$

As $P'D_\lambda^{-1}P\mathbf{1}_k = k\mathbf{1}_k$ (observed earlier), the eigenvector of $P'D_\lambda^{-1}P$ corresponding to the eigenvalue $k \ \ (= \alpha_k)$ is $\mathbf{1}_k$. Using this and the spectral decomposition of $P'D_\lambda^{-1}P$ we get the following alternative perspective of Lemma 3.2.

Lemma 3.3. *The lower bound in Lemma 3.2 is attained if and only if*

$$P'D_\lambda^{-1}P = a_q I + b_q \boldsymbol{1}_k \boldsymbol{1}_k', \tag{3.8}$$

where $a_q = [f(q) - k]/(k-1)$ *and* $b_q = 1 - a_q/k$.

Next, we need to explore existence of $P \in \mathcal{C}_\gamma^1$ satisfying (3.8) and find one, if it exists. For simplicity, consider the situation where $f(q) \neq f(q+1)$. Then, recall that to attain the lower bound in Lemma 3.2, each row of P must contain exactly q large values and $(k-q)$ small values. The positions for q large values can be chosen in $\binom{k}{q}$ ways. It is reasonable to explore RR designs which utilize all possible arrangements of large (and small) values. Wang et al. (2016) and Ye and Barg (2018) studied the following class of RR designs, requiring additionally all small values to be equal.

Definition 3.1. *For any integer* t *with* $1 \le t \le k-1$, *an RR design* $P \in \mathcal{C}_\gamma^1$ *is called a* t*-subset*

design if (i) P has $\binom{k}{t}$ rows, (ii) P contain exactly 2 distinct values; a large value and a small value and (iii) each row contains exactly t large and $(k-t)$ small values.

We shall denote a t-subset design by P_t and let $m_t = \binom{k}{t}$. Since proportional rows are not allowed, for each t, P_t is unique up to row permutation. We shall see in the sequel that our minimax RR design is P_q, i.e., P_t with $t = q$, with q as defined in (3.4). To review some basic properties of P_t, denote its small value by s_t; so its large value is γs_t. Clearly, P_t has m_t rows. It can be seen that each column of P_t contains exactly $\binom{k-1}{t-1} = (\frac{t}{k})m_t$ large values and $\binom{k-1}{t} = (\frac{k-t}{k})m_t$ small values. From these, we can find that

$$s_t = \frac{k}{m_t(t\gamma + k - t)}$$

and that the sum of each row is k/m_t.

Remark 3.2. *A t-subset design can be constructed as follows. Consider all $\binom{k}{t}$ subsets of size t of $\mathcal{S}_X = \{c_1, \ldots, c_k\}$, the sample space of X. Call the subsets $d_1, \ldots, d_{m_t}$, where $m_t = \binom{k}{t}$. Thus, each d_i contains a subset of the t categories in $\mathcal{S}_X$. Then, let $p_{ij} = \gamma s_t$ if $c_j \in d_i$, otherwise $p_{ij} = s_t$. Fugure 2 illustrates this for $k = 4$, $\gamma = 2$ and $t = 1, 2, 3$, where the columns represent c_1, c_2, c_3 and c_4, respectively, and the subsets are shown to the right of each TPM.*

$$\frac{1}{5}\begin{bmatrix} 2 & 1 & 1 & 1 \\ 1 & 2 & 1 & 1 \\ 1 & 1 & 2 & 1 \\ 1 & 1 & 1 & 2 \end{bmatrix} \begin{matrix} \{c_1\} \\ \{c_2\} \\ \{c_3\} \\ \{c_4\} \end{matrix} \qquad t = 1$$

$$\frac{1}{6}\begin{bmatrix} 2 & 2 & 1 & 1 \\ 2 & 1 & 2 & 1 \\ 2 & 1 & 1 & 2 \\ 1 & 2 & 2 & 1 \\ 1 & 2 & 1 & 2 \\ 1 & 1 & 2 & 2 \end{bmatrix} \begin{matrix} \{c_1, c_2\} \\ \{c_1, c_3\} \\ \{c_1, c_4\} \\ \{c_2, c_3\} \\ \{c_2, c_4\} \\ \{c_3, c_4\} \end{matrix} \qquad t = 2$$

$$\frac{1}{7}\begin{bmatrix} 2 & 2 & 2 & 1 \\ 2 & 2 & 1 & 2 \\ 2 & 1 & 2 & 2 \\ 1 & 2 & 2 & 2 \end{bmatrix} \begin{matrix} \{c_1, c_2, c_3\} \\ \{c_1, c_2, c_4\} \\ \{c_1, c_3, c_4\} \\ \{c_2, c_3, c_4\} \end{matrix} \qquad t = 3$$

Figure 2: t-subset designs

Next, we present some additional properties of t-subset designs, for $1 \leq t \leq k-1$. As each

row of P_t adds to k/m_t, it follows that

$$P_t\pi_u = P_t\Big(\frac{1}{k}\mathbf{1}_k\Big) = \frac{1}{m_t}\mathbf{1}_{m_t} \quad \Longrightarrow D_{(P_t\pi_u)} = \frac{1}{m_t}I. \tag{3.9}$$

By (3.9), we get $P_t' D^{-1}_{(P_t\pi_u)} P_t = m_t P_t' P_t$. As noted earlier, each column of P_t has $\binom{k-1}{t-1}$ values that are γs_t and the rest are s_t. Any two columns of P_t have the large value (γs_t) in $\binom{k-2}{t-2}$ common rows, the small value s_t in $\binom{k-2}{t}$ common rows and the remaining rows contain one large and one small value. Using these and routine algebra we can verify that

$$P_t' D^{-1}_{(P_t\pi_u)} P_t = m_t P_t' P_t = a_t I + b_t \mathbf{1}_k \mathbf{1}_k', \tag{3.10}$$

where

$$a_t = \frac{f(t) - k}{k-1} \quad \text{and} \quad b_t = 1 - \frac{a_t}{k}. \tag{3.11}$$

By (3.10), Lemma 3.3 and previous observations we obtain:

Theorem 3.1. *For given k and γ, let q and P_q be as defined earlier and let L_q denote the optimal L_* in (2.3) for $P = P_q$ and $\pi = \pi_u$. Then,*

$$\inf_{P \in \mathcal{C}^1_\gamma} \inf_L \boldsymbol{R}(P, L; \pi_u) = \boldsymbol{R}(P_q, L_q; \pi_u) = \frac{(k-1)^2}{f(q) - k}. \tag{3.12}$$

This result tells us that (P_q, L_q) is a locally (at $\pi = \pi_u$) optimal RR strategy at privacy level γ. To investigate its properties more generally and to solve the minimax problem of (2.5), we next describe some additional properties of t-subset designs. For a given P_t, the locally (at π_u)

optimum L, to be denoted L_t, can be simplified by using (3.9) and (3.10) in (2.3). Specifically,

$$\begin{aligned} L_t &= m_t(a_t I + b_t \mathbf{1}_k \mathbf{1}_k')^{-1} P_t' \\ &= m_t(a_t^{-1} I - d_t \mathbf{1}_k \mathbf{1}_k') P_t' \\ &= a_t^{-1}(m_t P_t' - b_t \mathbf{1}_k \mathbf{1}_{m_t}'), \end{aligned} \tag{3.13}$$

where $d_t = b_t/\{a_t(a_t + kb_t)\} = b_t/(ka_t)$ and the last $=$ follows from $\mathbf{1}_k' P_t' = (k/m_t)\mathbf{1}_{m_t}'$.

Lemma 3.4. *For any t-subset design P_t and L_t as in (3.13), $tr(L_t D_\lambda L_t')$ is a constant, independent of π, where $\lambda = P_t \pi$.*

Proof. Let $((f_{ij})) = F_{m_t \times m_t} = L_t' L_t$. Then, f_{ii} is the squared length of the ith row of L_t'. Using (3.13) and considering the structure of P_t, we see that in each row of L_t', exactly t values are $a_t^{-1}(m_t \gamma s_t - b_t)$ and $(k - t)$ are $a_t^{-1}(m_t s_t - b_t)$. So, all rows are have the same length and consequently, $f_{11} = \cdots = f_{m_t m_t} = f_0, say$. Now,

$$tr(L_t D_\lambda L_t') = tr(D_\lambda L_t' L_t) = \sum_{i=1}^{m_t} \lambda_i f_0 = f_0 \sum_{i=1}^{m_t} \lambda_i = f_0,$$

which is independent of π, as the lemma asserts. □

The next two theorems give a minimax estimator for given P_t and a minimax strategy satisfying (2.5).

Theorem 3.2. *For any given t-subset design P_t, a linear unbiased minimax estimator of π is $L_t \hat{\lambda}$, where L_t is as given by (3.13).*

Proof. First, we note that

$$\begin{aligned}\sup_{\pi}\mathbf{R}(P_t, L_t;\pi) &= \sup_{\pi}\left[tr(L_t D_\lambda L_t') - \sum \pi_i^2\right] \\ &= tr(L_t D_\lambda L_t') - \inf_{\pi}\sum \pi_i^2 \\ &= \mathbf{R}(P_t, L_t;\pi_u), \end{aligned} \tag{3.14}$$

as $tr(L_t D_\lambda L_t')$ is independent of π by Lemma 3.4 and $\sum \pi_i^2$ is minimum when $\pi = \pi_u$. Consider any L such that $LP_t = I$. Then, by (3.14) and Proposition 2.1,

$$\sup_{\pi}\mathbf{R}(P_t, L_t;\pi) = \mathbf{R}(P_t, L_t;\pi_u) \leq \mathbf{R}(P_t, L;\pi_u) \leq \sup_{\pi}\mathbf{R}(P_t, L;\pi),$$

which proves the theorem. □

Theorem 3.3. *For a given privacy level γ, a minimax strategy that solves (2.5) is (P_q, L_q) and*

$$\inf_{P\in\mathcal{C}_\gamma^1}\inf_{L}\sup_{\pi}\boldsymbol{R}(P, L;\pi) = \sup_{\pi}\boldsymbol{R}(P_q, L_q;\pi) = \frac{(k-1)^2}{f(q)-k}, \tag{3.15}$$

where $f(.), q, P_q$ and L_q are as defined earlier.

Proof. By (3.14) and Theorem 3.1 we get

$$\begin{aligned}\sup_{\pi}\mathbf{R}(P_q, L_q;\pi) &= \mathbf{R}(P_q, L_q;\pi_u) \\ &= \inf_{P\in\mathcal{C}_\gamma^1}\inf_{L}\mathbf{R}(P, L;\pi_u) \\ &\leq \inf_{P\in\mathcal{C}_\gamma^1}\inf_{L}\sup_{\pi}\mathbf{R}(P, L;\pi).\end{aligned}$$

Now, (3.15) follows readily from (3.12). □

Clearly, the minimax risk in (3.15) is a function of k and γ. From (3.4), $q \approx k/(1+\gamma)$, and

hence it can be seen that

$$\text{minimax risk} = \frac{(k-1)^2}{f(q)-k} \approx 4\frac{(k-1)^2}{k}\frac{\gamma}{(\gamma-1)^2}.$$

Figure 3 exhibits the dependence of the minimax risk on k and γ. Approximately, the risk is proportional to k and inversely proprtional to γ.

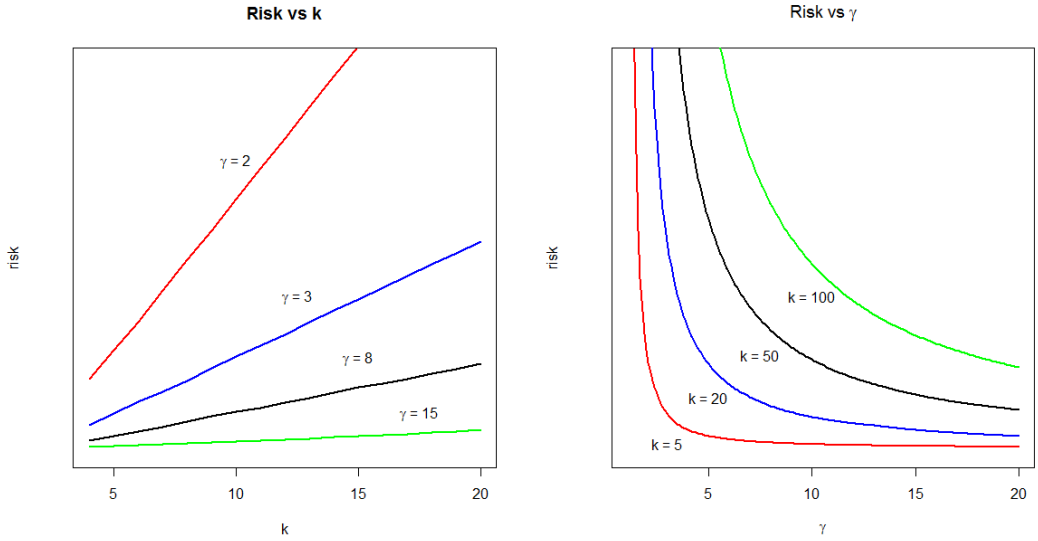

Figure 3: The dependence of minimax risk on k and γ

We should briefly discuss how our results differ from those of Wang et al. (2016) and Ye and Barg (2018). Ye and Barg (2018) also introduced t-subset designs and arrived at P_q and a method of moments estimator of π as an optimal strategy, but by considering minimax rate of convergence, similar to Duchi et al. (2018). Wang et al. (2016) considered finding an RR design P, subject to (1.2) or equivalently (1.5), that maximizes the mutual information between true and randomized responses under $\pi = \pi_u$ and proved that P_q is an optimal design. They also proposed the same estimator of π as in Ye and Barg (2018). We shall prove in the next section that the method of moments estimator coincides with our minimax estimator.

	γ					
k	1.1	1.5	2	5	10	20
4	2 6	2 6	1 4	1 4	1 4	1 4
6	3 20	2 15	2 15	1 6	1 6	1 6
10	5 252	4 210	3 120	2 45	1 10	1 10
20	10 184,756	8 125,970	7 77,520	3 1140	2 190	1 20

Table 1: values of q and m_q for some combinations of k and γ

4. Further Properties of t-subset and RAPPOR designs

4.1. Practical Aspects of t-subset Designs

The mathematical solution of the minimax problem derived in the preceding section is a bit abstract and hard to use. To construct and implement the minimax design P_q and calculate the estimator $L_q\hat{\lambda}$, following Section 3 results literally, we need to calculate q, define $d_1, \ldots, d_{m_q}$, compute P_q, apply randomization, obtain $\hat{\lambda}$ and finally calculate $L_q\hat{\lambda}$. That can be overly burdensome and time consuming because m_q may easily be very large. Table 1 gives values of q and m_q for several combinations of k and γ. In each cell, the top number is the value of q and the bottom number is m_q. As an example, for $k = 20$ and $\gamma = 2$, we have $q = 7$ and $m_q = 77,520$, and so we shall need to create 77,520 response categories, randomize each true value among 77,520 categories (with very small probabilities) etc.

Wang et al. (2016) and Ye and Barg (2018) described an alternative and simpler method for using t-subset designs. In the following, we review that approach and present some new results. Using indicator vectors, both original and perturbed data may be presented conveniently as $n \times k$ matrices. Recording each true category with a row vector $X = (X_1, \ldots, X_k)$ whose ith component is 1 if the true category is c_i and 0 otherwise, the unperturbed data from n units yields an $n \times k$ data matrix $\mathcal{D}_*$. Perturbed data can also be organized as a matrix using the following scheme.

Take any t-subset design P_t with parity γ. Recall that P_t has $m_t = \binom{k}{t}$ rows and each row has t large values (γs_t) and $(k-t)$ small values (s_t). The output variable has m_t categories, which we labeled earlier $d_1, \ldots, d_{m_t}$ (arbitrarily) and attached those to the rows of P_t. The alternative scheme represents the output categories with the k dimensional row vectors that are obtained by replacing the large values by 1 and small values by 0 in P_t. Specifically, for $i = 1, \ldots, m_t$, the response corresponding to the ith row of P_t is recorded as $(z_{i1}, \ldots, z_{ik})$, where z_{ij} is 1 if $p_{ij} = \gamma s_t$ and zero otherwise. This data representation scheme is the reverse of the construction method noted in Remark 3.2. Note that $\sum_j z_{ij} = t$ for all i and the possible responses are the indicator vectors for all subsets of $\{c_1, \ldots, c_k\}$ of size t. Denoting the randomized response with a vector $Z = (Z_1, \ldots, Z_k)$ and using one row for each respondent, the data from using P_t can be given as a matrix $\mathcal{D}$ of order $n \times k$.

Wang et al. (2016) and Ye and Barg (2018) gave the following algorithm for implementing P_t and generating a data matrix $\mathcal{D}$. For a true response $(x_1, \ldots, x_k)$ the algorithm generates a randomized response $(z_1, \ldots, z_k)$ as follows. Recall that only one of $x_1, \ldots, x_k$ is 1 and the rest are 0. Suppose $x_j = 1$. Then, first using a suitable binary experiment set $z_j = 1$ with probability $p = tm_t\gamma s_t/k$, else set $z_j = 0$. Next, if $z_j = 1$, randomly select $(t-1)$ of the remaining $(k-1)$ components of z and set those to 1. For $z_j = 0$, assign 1 to t of the remaining components of z, selected at random. In either case, all other components of z are 0. Then, it can be verified that

$$P(Z = (z_1, \ldots, z_k) | x_j = 1) = \begin{cases} \gamma s_t, & \text{if } z_j = 1 \\ s_t, & \text{if } z_j \neq 1 \end{cases}$$

and hence the algorithm implements P_t. Actually, the algorithm can be motivated and justified by noting that each column of P_t contains $\binom{k-1}{t-1}$ large values and $\binom{k-1}{t}$ small values and we have the following:

$$P(Z_i = 1 | X_i = 1) = \binom{k-1}{t-1}\gamma s_t = \frac{t}{k} m_t \gamma s_t,$$

$$P(Z=(z_1,\ldots,z_k)|Z_i=1,X_i=1)=\begin{cases}1/\binom{k-1}{t-1}, & \text{if } z_i=1\\ 0, & \text{otherwise,}\end{cases}$$

$$P(Z=(z_1,\ldots,z_k)|Z_i=0,X_i=1)=\begin{cases}1/\binom{k-1}{t}, & \text{if } z_i=0\\ 0, & \text{otherwise.}\end{cases}$$

Both Wang et al. (2016) and Ye and Barg (2018) used method of moments for estimating π from the data matrix $\mathcal{D}$. Let $V'=(V_1,\ldots,V_k)$ denote the vector of column sums of $\mathcal{D}$. For $j=1,\ldots,k$, let n_j denote the original frequency of c_j. Then, it follows that

$$E(V_j|\mathcal{D}_*)=n_jp+\big(n-n_j\big)\Big[p\frac{t-1}{k-1}+(1-p)\frac{t}{k-1}\Big]$$

and unconditionally,

$$E(\frac{V_j}{n})=p\pi_j+\big(1-\pi_j\big)\Big[p\frac{t-1}{k-1}+(1-p)\frac{t}{k-1}\Big], \tag{4.1}$$

which is a linear function of π_j. Recall that $p=tm_t\gamma s_t/k$, $m_t=\binom{k}{t}$ and $s_t=k/[\binom{k}{t}(t\gamma+k-t)]$. So, $p=(t\gamma)/(t\gamma+k-t)$. Using this in (4.1) and standard algebra, one obtains the following method of moments estimator of π_j (for $j=1,\ldots,k$):

$$\tilde{\pi}_j=\frac{(k-1)(t\gamma+k-t)}{t(\gamma-1)(k-t)}\Big(\frac{V_j}{n}\Big)+\frac{1}{k}\Big[\frac{(1-k)(t\gamma+k-t)}{(\gamma-1)(k-t)}+1\Big]. \tag{4.2}$$

Note that the method of moments estimator $\tilde{\pi}$ of π requires only the column totals of the data matrix $\mathcal{D}$ and is very easy to calculate. It can also verified that $\tilde{\pi}$ is an unbiased estimator of π. Another interesting property of $\tilde{\pi}$ is the following:

Proposition 4.1. *The method of moments estimator $\tilde{\pi}$ is also a minimax linear unbiased estimator of π under the t-subset design P_t.*

Proof. We shall simplify our minimax estimator $\hat{\pi} = L_t\hat{\lambda}$ to prove this result. Using (3.13), we get

$$L_t\hat{\lambda} = a_t^{-1}m_tP_t'\hat{\lambda} - (a_t^{-1}b_t)\mathbf{1}_k = a_t^{-1}m_tP_t'\hat{\lambda} - (a_t^{-1} - k^{-1})\mathbf{1}_k.$$

So,

$$\hat{\pi}_j = a_t^{-1}m_t\sum_{i=1}^{m_t} p_{ij}\hat{\lambda}_i - (a_t^{-1} - k^{-1}). \tag{4.3}$$

Represent the response categories using indicator vectors $d_i = (z_{i1}, \ldots, z_{ik}), i = 1, \ldots, m_t$, as discussed above. Recall that $z_{ij} = 1$ if $p_{ij} = \gamma s_t$ and $z_{ij} = 0$ if $p_{ij} = s_t$. Let $B_j = \{d_i : z_{ij} = 1\}$. Then,

$$\sum_{i=1}^{m} P_{ij}\hat{\lambda}_i = \gamma s_t \sum_{i\in B_j}\hat{\lambda}_i + s_t\sum_{i\notin B_j}\hat{\lambda}_i = \gamma s_t\big(\frac{V_j}{n}\big) + s_t\Big(1 - \big(\frac{V_j}{n}\big)\Big)$$

and (4.3) reduces to

$$\hat{\pi}_j = a_t^{-1}(\gamma - 1)m_ts_t\Big(\frac{V_j}{n}\Big) + a_t^{-1}(m_ts_t - 1) + k^{-1}. \tag{4.4}$$

Next, we can verify the following identities:

$$m_ts_t = \frac{k}{t\gamma + k - t}, \quad m_ts_t - 1 = \frac{t(1-\gamma)}{t\gamma + k - t} \quad \text{and} \quad a_t = \frac{kt(\gamma-1)^2(k-t)}{(k-1)(t\gamma + k - t)^2}.$$

Using these and routine algebra (4.4) can be reduced to (4.2). □

In view of the preceding result and (4.2), the minimax estimator π under P_t is $\hat{\pi} = c_1(\frac{V}{n}) + c_2$, where c_1 and c_2 are evident from (4.2). Note that V is the sum of n independent realizations of the response vector $Z = (Z_1, \ldots, Z_k)$. So, the variance-covariance matrix of $\hat{\pi}$ is $V(\hat{\pi}) = (c_1^2/n)\Sigma$, where $\Sigma = ((\sigma_{ij})) = V(Z)$. Since each Z_i's are binary variables, $\sigma_{jj} = P(Z_j = 1)[1 - P(Z_j = 1)]$ and $\sigma_{ij} = P(Z_i = 1, Z_j = 1) - P(Z_i = 1)P(Z_j = 1)$ for $i \neq j$. Moreover, the right side of (4.1) is

$P(Z_j = 1)$ and simplifying it further we get

$$P(Z_j = 1) = \Big[\frac{t(\gamma m_t s_t - 1)}{k-1}\Big]\pi_j + \frac{t(k - \gamma m_t s_t)}{k(k-1)}$$
$$= \frac{t}{(k-1)(t\gamma + k - t)}\Big[(\gamma - 1)(k-t)\pi_j + \{t(\gamma - 1) + k - t\}\Big].$$

For $t = 1$, $P(Z_i = 1, Z_j = 1) = 0$. For $t \geq 3$, using simpler algorithm for implementing P_t, discussed above, and letting $p = (t\gamma m_t s_t)/k$, we get

$$\begin{aligned}
P(Z_i = 1, Z_j = 1) =& \sum_{r=1}^{k} \pi_r P(Z_i = 1, Z_j = 1 | X_r = 1) \\
=& (\pi_i + \pi_j) p \Big[\binom{k-2}{t-2} \div \binom{k-1}{t-1}\Big] + (1 - \pi_i - \pi_j)\Big[p\Big\{\binom{k-3}{t-3} \div \binom{k-1}{t-1}\Big\} \\
& + (1-p)\Big\{\binom{k-3}{t-2} \div \binom{k-1}{t}\Big\}\Big] \\
=& (\pi_i + \pi_j) p \frac{t-1}{k-1} + (1 - \pi_i - \pi_j)\Big[p\frac{(t-1)(t-2)}{(k-1)(k-2)} \\
& + (1-p)\frac{t(t-1)}{(k-1)(k-2)}\Big] \\
=& \frac{t(t-1)}{(k-1)(k-2)(t\gamma + k - t)}\Big[(k-t)(\gamma-1)(\pi_i + \pi_j) + (t\gamma - 2\gamma + k - t)\Big].
\end{aligned} \tag{4.5}$$

Actually, (4.5) holds for all $1 \leq t \leq k-1$. For $t = 2$, the above derivation remains valid if $\binom{k-3}{t-3}$ is interpreted as 0.

4.2. Mixture of t-subset Designs

This section is motivated by the RAPPOR (randomized aggregatable privacy-preserving ordinal response) algorithm proposed by Erlingsson et al. (2014). It is an RR procedure and has been further discussed by Kairouz et al. (2016a), Fanti et al. (2016), Wang et al. (2017), Ye and Barg (2018) and others. Quite importantly, Google and Apple use RAPPOR for privacy protection.

Basic RAPPOR is directly relevant to our context and it works as follows. As in Section 4.1, it represents the true category with an indicator vector $X = (X_1, \ldots, X_k)$. Then, it produces a perturbed output $Z = (Z_1, \ldots, Z_k)$ by changing each component of X independently with probability $p = 1/(\sqrt{\gamma}+1)$. So, the output space has 2^k elements. As we explain next, RAPPOR is a mixture of t-subset designs, with $t = 0, 1, \ldots, k$.

Consider RAPPOR perturbation and let $T = \sum_{j=1}^{k} Z_j$ denote the number of 1's in a randomized response $(Z_1, \ldots, Z_k)$. Then, for any $0 \le t \le k$ and $1 \le j \le k$,

$$\begin{aligned} P(T = t|X_j = 1) &= P(T = t, Z_j = 1|X_j = 1) + P(T = t, Z_j = 0|X_j = 1) \\ &= \binom{k-1}{t-1} p^{t-1}(1-p)^{k-t+1} + \binom{k-1}{t} p^{t+1}(1-p)^{k-t-1}. \end{aligned} \tag{4.6}$$

Since (4.6) is independent of j, it is also the unconditional probability $P(T = t)$, which we shall denote by p_t. Also,

$$P(Z = z, T = t|X_j = 1) = \begin{cases} p^{t-1}(1-p)^{k-t+1}, & \text{if } z_j = 1, \sum z_i = t, \\ p^{t+1}(1-p)^{k-t-1}, & \text{if } z_j = 0, \sum z_i = t. \end{cases}$$

Recall that $p = 1/(\sqrt{\gamma}+1)$ and so $\gamma = [(1-p)/p]^2$. Using this and the above, conditionally on $T = t$ and X we have

$$P(Z = z|T = t, X_j = 1) = \begin{cases} 1/[\binom{k-1}{t-1} + \binom{k-1}{t}\gamma^{-1}] = \gamma s_t, & \text{if } z_j = 1, \sum z_i = t, \\ 1/[\binom{k-1}{t-1}\gamma + \binom{k-1}{t-1}] = s_t, & \text{if } z_j = 0, \sum z_i = t, \end{cases}$$

which are the transition probabilities of the t-subset design with parity γ.

From the preceding observations it follows that RAPPOR perturbation is equivalent to a two step procedure: Perturb each true response by first drawing a value t from $\{0, 1, \ldots, k\}$ with probabilities $p_0, p_1, \ldots, p_k$ and then applying the t-subset design with parity γ. Thus, RAPPOR

is a mixture of t-subset designs and its TPM is $P = \left[p_0 P_0' \mid p_1 P_1' \mid \ldots \mid p_k P_k'\right]'$, where P_t is the TPM of the t-subset design, for $t = 0, 1, \ldots, k$.

Note that Theorem 2.1 implies that the basic RAPPOR design is inadmissible, as the two rows of its TPM corresponding to $t = 0$ and $t = k$ have parity 1 (i.e., each row contains a common value). The two associated outputs, i.e., $Z = (0, 0, \ldots, 0)$ and $Z = (1, 1, \ldots, 1)$, give no information about the true category and hence about π. Effectively, RAPPOR throws away the units that yield those two responses. This wastage is minimal for large k, where both p_0 and p_k are small. But, for small k, the loss can be substantial. We can remove those two rows and normalize the TPM to get an admissible design.

Motivated by the preceding discussion, we shall next explore properties of mixtures of t-subset designs, with $t = 1, \ldots, k-1$. The TPM of such a design is a partitioned matrix

$$P_M = \left[w_1 P_1' \mid w_2 P_2' \mid \ldots \mid w_{k-1} P_{k-1}'\right]', \tag{4.7}$$

where $w_j \geq 0$ are the mixing probabilities and $\sum w_j = 1$. Naturally, if $w_j = 0$ for some j, the corresponding rows should be omitted. We may conveniently view P_M as a two-step procedure: first select a value t from $\{1, \ldots, k-1\}$ with probabilities $w_1, \ldots, w_{k-1}$ and then apply the t-subset design. Note that $D_\lambda = diag(P_M \pi)$ is a block diagonal matrix

$$D_\lambda = diag(D_\lambda^{(1)}, D_\lambda^{(2)}, \ldots, D_\lambda^{(k-1)}),$$

where $D_\lambda^{(t)} = diag(w_t P_t \pi)$ for $t = 1, \ldots k-1$.

We can derive the minimax linear unbiased estimator of π under P_M, using arguments similar to those used in Section 3. The uniform distribution $\pi = \pi_u$ turns out to be a least favorable distribution in this case too. When $\pi = \pi_u$, $D_\lambda^{-1} = diag(\frac{m_1}{w_1} I_{m_1}, \frac{m_2}{w_2} I_{m_2}, \ldots, \frac{m_{k-1}}{w_{k-1}} I_{m_{k-1}})$ with

$m_t = \binom{k}{t}$, and by Proposition 2.1, the locally best unbiased estimator of π is $\hat{\pi} = L_M \hat{\lambda}$, where

$$L_M = \Big(\sum_{t=1}^{k-1} w_t P_t' (D_\lambda^{(t)})^{-1} w_t P_t \Big)^{-1} P_M' D_\lambda^{-1} \tag{4.8}$$

$$= \Big((\sum_{t=1}^{k-1} w_t a_t) I_k + (\sum_{t=1}^{k-1} w_t b_t) \mathbf{1}_k \mathbf{1}_k' \Big)^{-1} P_M' D_\lambda^{-1} \tag{4.9}$$

$$= \Big(a_* I_k + b_* \mathbf{1}_k \mathbf{1}_k' \Big)^{-1} P_M' D_\lambda^{-1}$$

$$= \Big(a_*^{-1} I_k - \frac{b_*}{k a_*} \mathbf{1}_k \mathbf{1}_k' \Big) P_M' D_\lambda^{-1}, \tag{4.10}$$

$a_* = \sum_{t=1}^{k-1} w_t a_t$ and $b_* = \sum_{t=1}^{k-1} w_t b_t$ (and a_t and b_t are as defined in (3.11)). Moreover, $P_M' D_\lambda^{-1} = [m_1 P_1' \ | \ m_2 P_2' \ | \ \ldots \ | m_{k-1} P_{k-1}']$, and so

$$L_M = [L_1^* \ | \ L_2^* \ | \ \ldots \ | L_{k-1}^*], \text{ with } L_t^* = a_*^{-1} (m_t P_t' - b_* \mathbf{1}_k \mathbf{1}_{m_t}'). \tag{4.11}$$

Note that L_t^* and L_t (in (3.13)) have the same structure, with different constants. So, letting $\lambda = P_M \pi$, it can be seen as in Lemma 3.4 that $tr(L_t^* D_\lambda^{(t)} L_t^{'*})$ is independent of π for all t. Now, using $L_M D_\lambda L_M' = \sum_{t=1}^{k-1} (L_t^* D_\lambda^{(t)} L_t^{'*})$, we can prove the following:

Lemma 4.1. *Consider any mixture of t-subset designs, P_M as in (4.7), and the corresponding L_M in (4.11), and let $\lambda = P_M \pi$. Then, $tr(L_M D_\lambda L_M')$ is a constant, independent of π.*

This lemma leads to the following result, whose proof is similar to that of Theorem 3.2 and hence omitted.

Theorem 4.1. *Under P_M in (4.7), a linear unbiased minimax estimator of π is $L_M \hat{\lambda}$, where L_M is as in (4.11).*

Next, we give a simpler view of the minimax estimator $L_M \hat{\lambda}$. For $t = 1, \ldots, k-1$, let $\hat{\lambda}^{(t)}$ denote the vector of relative frequencies of the response types that satisfy $\sum_j z_{ij} = t$, i.e., generated by a t-subset design (in the second step of our two-step view of P_M). The data $\mathcal{D}$ can be be represented

as $\mathcal{D}' = [\mathcal{D}'_1 \mid \mathcal{D}'_2 \mid \ldots \mid \mathcal{D}'_{k-1}]$, where $\mathcal{D}'_t$ contains all responses generated by the t-subset design. Let n_t denote the sample size of $\mathcal{D}'_t$. Then, using (4.11) we get

$$L_M\hat{\lambda} = \sum_{t=1}^{k-1} L_t^*\hat{\lambda}^{(t)} = \sum_{t=1}^{k-1} a_* m_t P_t'\hat{\lambda}^{(t)} - \sum_{t=1}^{k-1} \frac{n_t}{n}(a_*^{-1} - k^{-1})\mathbf{1}_k.$$

Now, using some results from the proof of Proposition 4.1, we get

$$\begin{aligned}\hat{\pi}_j &= \sum_{t=1}^{k-1}\Big[a_*^{-1} m_t s_t(\gamma-1)\frac{V_j^{(t)}}{n} + \frac{n_t}{n}a_*^{-1}(m_t s_t - 1) + \frac{n_t}{n}k^{-1}\Big] \\ &= a_*^{-1}\sum_{t=1}^{k-1}\Big[m_t s_t(\gamma-1)\frac{V_j^{(t)}}{n} + \frac{n_t}{n}(m_t s_t - 1)\Big] + k^{-1},\end{aligned}$$

where $V_j^{(t)}$ is the jth column sum of $\mathcal{D}_t$.

The minimax criterion compares maximum (over the parameter space) risks of competing procedures, and in general, a minimax procedure need not dominate (or be uniformly better) another procedure, i.e., have a uniformly smaller risk. Interestingly, the following theorem shows that the minimax estimator $L_q\hat{\lambda}$ based on the q-subset design dominates the minimax estimator $L_M\hat{\lambda}$ based on any mixture of t-subset designs.

Theorem 4.2. *Let P_M be a mixture of t-subset designs and suppose $P_M \neq P_q$. Then, the strategy (P_M, L_M) is dominated by the minimax strategy (P_q, L_q), i.e., $\boldsymbol{R}(P_q, L_q; \pi) \leq \boldsymbol{R}(P_M, L_M; \pi)$ for all π and the "=" holds if and only if $P_M = P_q$.*

Proof. By (2.1) and lemmas 3.4 and 4.1, the difference of the risk functions of the two strategies,

$$\mathbf{R}(P_M, L_M; \pi) - \mathbf{R}(P_q, L_q; \pi) = tr(L_M D_\lambda L_M') - tr(L_q D_\lambda L_q'),$$

is independent of π. So,

$$\mathbf{R}(P_M, L_M; \pi) - \mathbf{R}(P_q, L_q; \pi) = \mathbf{R}(P_M, L_M; \pi_u) - \mathbf{R}(P_q, L_q; \pi_u). \tag{4.12}$$

Now, the proof can be completed by noting that if $P_M \neq P_q$, then P_M contains rows that have more than q large values and hence by Theorem 3.1, (4.12) > 0. □

Remark 4.1. *For a mixture design* $P_M = \left[w_0 P_0' \mid w_1 P_1' \mid \ldots \mid w_k P_k'\right]'$ *that includes the two constant rows corresponding to* $t = 0$ *and* $t = k$*, as in RAPPOR design, the preceding results hold with simple changes. In particular, the sums in (4.8) and (4.9) will be over* $t = 0$ *to* k *and* L_M *in (4.11) will include* L_0^* *and* L_k^**. With these changes, theorems 4.1 and 4.2 hold true. Note from (3.3) and (3.11) that* $a_0 = a_k = 0$ *and* $b_0 = b_k = 1$ *and so in (4.10),* a_* *remains the same and* $b_* = (w_0 + w_k) + \sum_{t=1}^{k-1} w_t b_t$*.*

Next, we shall discuss some directions for improving upon the basic RAPPOR strategy. The empirical estimator currently being used with RAPPOR (see Erlingsson et al. (2014) and Ye and Berg (2018)) is

$$\tilde{\pi}_R = \Big(\frac{\sqrt{\gamma}+1}{\sqrt{\gamma}-1}\Big)\frac{V}{n} - \frac{1}{\sqrt{\gamma}-1}\mathbf{1}_k, \tag{4.13}$$

where V is the vector of column sums of the data matrix, as in Section 4.1. Specifically, the jth component (V_j) of V is the number of responses with $Z_j = 1$. Kairouz et al. (2016a) derived the risk of the empirical estimator as

$$\mathbf{R}(P_R, \tilde{\pi}_R; \pi) = \frac{k\sqrt{\gamma}}{(\sqrt{\gamma}-1)^2} + 1 - \sum_{i=1}^{k} \pi_i^2. \tag{4.14}$$

It can be seen that $\tilde{\pi}_R$ is different from the minimax estimator $\hat{\pi}_R = L_R\hat{\lambda}$ under RAPPOR design, where L_R is similar to (4.11), as noted in Remark 4.1.

Theorem 4.3. *For RAPPOR design,* $\boldsymbol{R}(P_R, L_R; \pi) < \boldsymbol{R}(P_R, \tilde{\pi}_R; \pi)$ *for all* π *and thus, the em-*

pirical estimator $\tilde{\pi}_R$ in (4.13) is dominated by the minimax linear unbiased estimator $L_R\hat{\lambda}$.

Proof. By Lemma 4.1 and (4.14), the difference of the two risks is

$$\mathbf{R}(P_R, \tilde{\pi}; \pi) - \mathbf{R}(P_R, L_R; \pi) = \frac{k\sqrt{\gamma}}{(\sqrt{\gamma}-1)^2} + 1 - tr(L_R D_\lambda L_R'),$$

which is independent of π. Now, using $\pi = \pi_u$, we get

$$\mathbf{R}(P_R, \tilde{\pi}_R; \pi) - \mathbf{R}(P_R, L_R; \pi) = \mathbf{R}(P_R, \tilde{\pi}_R; \pi_u) - \mathbf{R}(P_R, L_R; \pi_u) > 0$$

as in the proof of Theorem 4.2. □

The preceding result shows that the RAPPOR strategy $(P_R, \tilde{\pi}_R)$ can be improved by replacing the empirical estimator by the linear unbiased minimax estimator $\hat{\pi} = L_R\hat{\lambda}$. As we noted earlier, the basic RAPPOR design is inadmissible, as its TPM includes two constant rows. Deleting those two rows and normalizing the weights we get a modified RAPPOR design that is admissible. Thus, a better idea would be to use this modified design and the corresponding minimax estimator. However, this modified RAPPOR method is worse than the minimax strategy (P_q, L_q), by Theorem 4.2.

We compared the sample size adjusted risks, defined in (2.1), of (P_q, L_q) and (P_R, L_R) with that of $(P_R, \tilde{\pi}_R)$ for some γ and k. The results are presented in Figure 4. The solid curves represent the relative efficiency of $(P_R, \tilde{\pi}_R)$ compared to (P_q, L_q). They show the ratio of the risk of (P_q, L_q) to that of $(P_R, \tilde{\pi}_R)$. Similarly, the dashed curves compare the risks of (P_R, L_R) and $(P_R, \tilde{\pi}_R)$. Thus, they show possible efficiency gain from just replacing the RAPPOR's estimator by the minimax estimator under RAPPOR design, which can be obtained from (4.11). All relative efficiency curves are always less than 1, consistent with our theoretical results. We see that the minimax strategy is substantially better than the original RAPPOR strategy, especially for moderate to

large γ, i.e., in moderate to low privacy paradigm. However, the difference diminishes as k gets fairly large. Under RAPPOR design, the efficiency gain from using the corresponding minimax estimator is for small to moderate k, depending on γ.

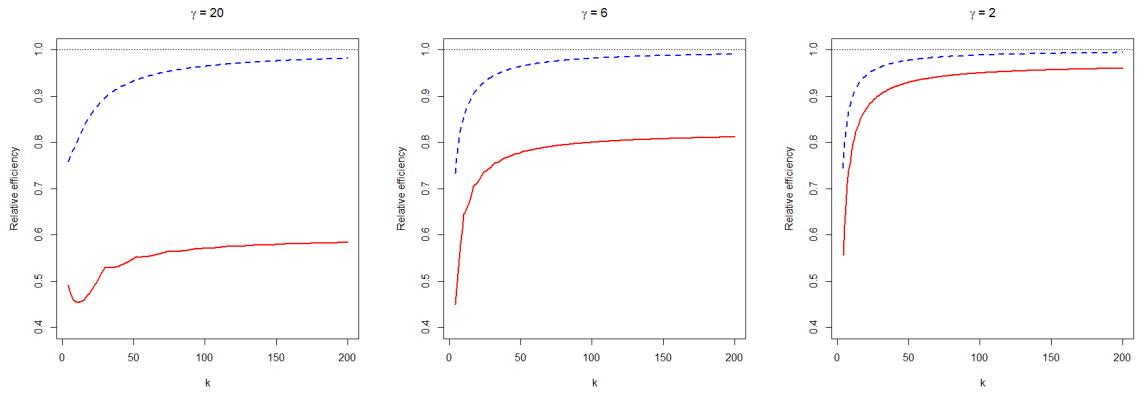

Figure 4: Relative efficiency comparison

5. Discussion

In this paper, we derived minimax RR designs and estimators for estimating multinational cell probabilities under squared error loss, unbiasedness and linearity. However, we believe that unbiased estimation is important, especially in presence of data perturbation. As we discussed, both response randomization and estimation from RR data in the optimal method are easy to carry out. Wang et al. (2016) and Ye and Barg (2018) also proposed the design and the estimator, but from other considerations. We presented the variance of the minimax estimator under a t-subset design, which can be estimated easily by replacing π_i's by their estimates. We also derived minimax estimators under mixtures of t-subset designs. That yields a better estimator for the RAPPOR design, which we hope will help to improve the method, especially in medium to low privacy domains. We should note our results are applicable to multiple categorical variables, by considering their cross-classification.

Duchi et al. (2018) considered sequentially interactive RR mechanisms, where the randomization probabilities for a response are allowed to depend on previous randomized outputs. We restricted our attention to only non-interactive mechanisms. Finding exact minimax methods among interactive mechanisms is an interesting problem, which we leave for future investigation. For a categorical variable, estimation of the cell probabilities is a basic problem, as those can be used to derive other inferences, but perhaps not optimally. So, optimal methods under other loss functions should be investigated.

References

[1] Aggarwal, C.C. and Yu, P.S. (Eds.) (2008). *Privacy-Preserving Data Mining: Models and Algorithms*, New York: Springer Science and Business Media.

[2] Agrawal, R. and Srikant, R. (2000). Privacy-preserving data mining. In *Proceedings of the ACM SIGMOD International Conference on Management of Data*, Dallas, Texas, pp. 439-450.

[3] Agrawal, S., Haritsa, J.R. and Prakash, B.A. (2009). FRAPP: a framework for high-accuracy privacy-preserving mining. *Data Mining and Knowledge Discovery*, **18**, 101-139.

[4] Anderson, H. (1976). Estimation of a proportion through randomized response. *International Statistical Review*. **44**, 213-217.

[5] Blackwell, D. (1951). Comparison of experiments. *In Proceedings of the second Berkeley Symposium on Mathematical Statistics and Probability* (Vol. 1, pp. 93-102).

[6] Blackwell, D. (1953). Equivalent comparisons of experiments. *The Annals of Mathematical Statistics* **24**, 265-272.

[7] Chai, J., and Nayak, T.K. (2018). A criterion for privacy protection in data collection and its attainment via randomized response procedures. *Electronic Journal of Statistics*, **12**, 4264-4287.

[8] Chaudhuri, A. (2011). *Randomized Response and Indirect Questioning Techniques in Surveys*. Boca Raton, FL, CRC Press.

[9] Chaudhuri, A. and Mukerjee, R. (1988). *Randomized Response: Theory and Techniques*. New York, Marcel Dekker.

[10] Chen, B-C., Kifer, D., LeFevre, K. and Machanavajjhala, A. (2009) Privacy-preserving data publishing. *Foundations and Trends in Databases*, **2**, 1-167.

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

[12] Erlingsson, U., Pihur, V., and Korolova, A. (2014). Rappor: Randomized aggregatable privacy-preserving ordinal response. *In Proceedings of the 2014 ACM SIGSAC conference on computer and communications security (pp. 1054-1067)*. ACM.

[13] Evfimievski, A., Gehrke, J. and Srikant, R. (2003). Limiting privacy breaches in privacy preserving data mining. *In Proceedings of the twenty-second ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems (pp. 211-222)*.ACM.

[14] Evfimievski, A., Srikant, R., Agrawal, R. and Gehrke, J. (2004). Privacy preserving mining of association rules. *Information Systems*, **29**, 343-364.

[15] Fanti, G., Pihur, V. and lfar Erlingsson, U. (2016). Building a RAPPOR with the unknown: Privacy-preserving learning of associations and data dictionaries. *Proceedings on Privacy Enhancing Technologies*, 3, 4161

[16] Fligner, M.A., Policello, G.E. and Singh, J. (1977). A comparison of two randomized response survey methods with consideration for the level of respondent protection. *Communications in Statistics-Theory and Methods*. **6**, 1511-1524.

[17] Fox, J.A. (2016). *Randomized Response and Related Methods: Surveying Sensitive Data*. Thousand Oaks, CA, Sage Publications.

[18] Fung, B.C.M., Wang, K., Chen, R. and Yu, P.S. (2010). Privacy-preserving data publishing: A survey of recent developments. *ACM Computing Surveys*, **42**, 14.

[19] Greenberg, B.G., Abul-Ela, A.-L.A., Simmons, W.R. and Horvitz, D.G. (1969). The unrelated question randomized response model: theoretical framework. *Journal of the American Statistical Association*. **64**, 520-539.

[20] Kairouz, P., Bonawitz, K., and Ramage, D. (2016a). Discrete Distribution Estimation under Local Privacy. In *International Conference on Machine Learning* (pp. 2436-2444).

[21] Kairouz, P., Oh, S., and Viswanath, P. (2016b). Extremal Mechanisms for Local Differential Privacy. *Journal of Machine Learning Research*, **17**, 1-51.

[22] Lax, Peter D. (2007). *Linear Algebra and Its Applications.* 2nd ed., Wiley-Interscience.

[23] Marshall, A.W., Olkin, I., and Arnold, B. (2011). Inequalities: Theory of Majorization and Its Applications. *New York.*

[24] Nayak, T.K. (1994). On randomized response surveys for estimating a proportion. *Communications in Statistics-Theory and Methods*, **23**, 3303-3321.

[25] Nayak, T.K., and Adeshiyan, S.A. (2009). A unified framework for analysis and comparison of randomized response surveys of binary characteristics. *Journal of Statistical Planning and Inference*, **139**, 2757-2766.

[26] Nayak, T.K., Zhang, C. and Adeshiyan, S.A. (2015). Emerging applications of randomized response concepts and some related issues. *Model Assisted Statistics and Applications*, **10**, 335-344.

[27] Rizvi, S.J. and Haritsa, J.R. (2002). Maintaining data privacy in association rule mining. *In Proceedings of the 28th international conference on Very Large Data Bases (pp. 682-693)*. VLDB Endowment.

[28] Wang, S., Huang, L., Wang, P., Nie, Y., Xu, H., Yang, W., Li, X-Y. and Qiao, C. (2016). Mutual Information Optimally Local Private Discrete Distribution Estimation. *arXiv preprint arXiv:1607.08025*

[29] Wang, T., Blocki, J, Li, N. and Jha, S. (2017). Locally deferentially private protocols for frequency estimation. *Proceedings of 26th USENIX Security Symposium*.

[30] Warner, S.L. (1965). Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association*, **60**, 63-69.

[31] Ye, M. and Barg, A. (2018). Optimal schemes for discrete distribution estimation under locally differential privacy. *IEEE Transactions on Information Theory*, **64**, 5662-5676.