RESEARCH REPORT SERIES
(Disclosure Avoidance #2016-03)

A Concise Theory of Randomized Response Techniques for Privacy and Confidentiality Protection

Tapan K. Nayak, Samson A. Adeshiyan and Cheng Zhang

Center for Disclosure Avoidance Research
U.S. Census Bureau
Washington DC 20233

Report Issued: July 28, 2016

A Concise Theory of Randomized Response Techniques for Privacy and Confidentiality Protection

Tapan K. Nayak, * Samson A. Adeshiyan† and Cheng Zhang‡§

Abstract

A variety of randomized response (RR) procedures for privacy and confidentiality protection have been proposed, studied and compared in the literature. We describe statistically relevant attributes of RR mechanisms and use those to duly organize and unify existing estimation theories for diverse RR methods. Any RR procedure can be characterized by its transition probability matrix P, as it determines all statistical properties of the procedure. In RR surveys, P is fixed, but in post-randomization for confidentiality protection, P may depend on the original data. This affects statistical inferences significantly. We also review some optimality results in the comparison of RR surveys of a binary variable, based on both privacy protection and statistical efficiency.

Key words and Phrases: Categorical variable; post-randomization; random transformation; sampling design; transition probability matrix; unbiased estimation; variance inflation.

*Center for Disclosure Avoidance Research, U.S. Census Bureau, Washington, DC 20233 and Department of Statistics, George Washington University, Washington, DC 20052.

†U.S. Energy Information Administration, Washington, DC 20585.

‡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. The analysis and conclusions contained in this paper are those of the authors and do not represent the official position of the U.S. Energy Information Administration (EIA) or the U.S. Department of Energy (DOE).

1. Introduction

Concerns about privacy and confidentiality often yield substantial nonresponse and false responses, especially in surveys that ask direct questions about sensitive attributes such as criminal history, tax evasion, drug abuse, gambling and abortion. To protect respondents' privacy and increase truthful respondent participation, Warner (1965) proposed the first randomized response (RR) procedure for a binary characteristic. Consider a population where each person belongs either to a sensitive group A or to its complement A^c, where the true proportion (π) of the population that belongs to A is unknown and needs to be estimated from survey data. In Warner's (1965) method, each respondent is instructed to select one of the two questions Q_1: Do you belong to A? and Q_2: Do you belong to A^c? in a prescribed random manner (e.g., by drawing a card from a shuffled deck or using a spinner) and truthfully answer the question by "Yes" or "No" without disclosing the question, to protect his/her privacy. The probability (p) of selecting Q_1 is the method's design parameter and the survey organization needs to choose its value and devise an experiment to implement it. Basically, Warner's procedure converts each true "Yes" (or No) to a false "No" (or Yes) with probability $1 - p$.

Subsequent to Warner's (1965) pioneering paper, many other RR methods for both categorical and quantitative variables have been proposed and investigated; see Chaudhuri and Mukerjee (1988), Chaudhuri (2010), Chaudhuri and Christofides (2013) and other articles in this book for excellent exposition and references. All RR methods modify the true responses in some stochastic fashion, which necessitates developing theory for (i) making inferences from randomized data and (ii) assessing how much privacy and confidentiality protection an RR method provides. The main goal of this article is to discuss these two aspects of RR methods for categorical data in a general and unified manner. We shall concentrate on important logical elements of RR theory, leaving aside related mathematical derivations that are readily available in the literature. Also, we shall not discuss RR procedures for quantitative variables, which mostly employ noise addition or multiplication (see Fuller, 1993; Evans et al., 1998; Brand, 2002 and Nayak et al., 2011), as those are patently different from methods for categorical variables.

Privacy and confidentiality are often used synonymously, but incorrectly. Privacy is an

individual's right to control access to his information and privacy protection means hiding a respondent's true values from everyone, including the interviewer. Thus, Warner's method was designed to give privacy protection. Protecting confidentiality means keeping information about each respondent confidential. Often respondents give their information to survey organizations trusting that their data will be used by researchers and policy makers only to learn about the population as a whole and not about any individual or survey unit. Privacy arises at data collection stage whereas confidentiality emerges after the data have been collected. Confidentiality protection often prevents data agencies from releasing original data as that might enable others to gain much information about specific units in the survey. So, agencies typically release perturbed or masked versions of original data. Randomization of the original values is one technique for protecting confidentiality. However, we should note that in RR surveys the randomization mechanism is chosen before data collection and hence it cannot depend on the data, but for confidentiality protection, the randomization process may be chosen based on the data. This is a significant difference and it requires us to treat privacy and confidentiality differently. Nayak et al. (2015) discuss additional differences between privacy and confidentiality protection.

Both data analysis and the protection of privacy or confidentiality depend on the randomization process and the sampling mechanism. Thus, we need to organize the theory by different types of sampling and RR mechanism. We shall consider two types of sampling: multinomial sampling and general probability sampling from a finite population. While RR methods may differ in many ways, only certain features are statistically relevant and affect data analysis or privacy and confidentiality protection. It is important to recognize those features for developing statistical theory and proper comparison of competing methods. The rest of this article is organized as follows. In Section 2, we examine RR experiments from a statistical perspective to discern their essential features and design parameters. This helps us to put different RR procedures in a common framework and grasp their substantive differences, if any. In Section 3, we discuss statistical estimation when the randomization experiment is specified prior to data collection. In Section 4, we review some recent results on estimation for invariant post-randomization, which is a particular type of data dependent randomization technique that is

useful for confidentiality protection. In Section 5, we discuss privacy protection and certain optimality results. Section 6 contains some closing remarks.

2. Vital Attributes of Randomization Experiments

In this section, we examine randomization experiments for categorical variables to bring out their statistically essential features. Let X be a categorical variable with categories labeled $c_1, \ldots, c_k$, and let $\pi_i = P(X = c_i), i = 1, \ldots, k$, and $\Pi = (\pi_1, \ldots, \pi_k)'$. Also, when we randomize several categorical variables, X will represent their cross-classification and Π will represent all joint probabilities. Typically, Π is unknown and we wish to make inferences about Π and functions of it. The main purpose of RR surveys is to collect useful information while protecting respondent's privacy. Fundamentally, all RR methods ask each respondent to perform a given random experiment and use its outcome, the respondent's true category and pre-specified rules to compute an output and report it. Let Z denote the output variable. Evidently, the input space (or domain) of the random transformation from X to Z is $\{c_1, \ldots, c_k\}$. Consider an RR process with a finite output space (or range), containing m elements, labeled $d_1, \ldots, d_m$. Then, the randomization process produces different outputs, for given input, with specific probabilities. Thus, let $p_{ij} = P(Z = d_i | X = c_j), i = 1, \ldots, m, j = 1, \ldots, k$. To define m uniquely, we should require each $d_i (i = 1, \ldots, m)$ to occur with positive probability, i.e., $p_{ij} > 0$ for some j. Then, the transition probability matrix of order $(m \times k)$ of the RR procedure is $P = ((p_{ij}))$ and $\sum_i p_{ij} = 1$ for $j = 1, \ldots, k$.

Any randomization process is characterized by its output space and the transition probability matrix P (just as a random variable is fully described by its probability distribution). So, all effects of an RR procedure on the distribution of Z and privacy protection are determined only through its P. One important (and under-appreciated) implication of this is that for statistical investigations we should consider only P, leaving aside ancillary features of the experiment, e.g., whether it uses a deck of cards or a spinner. However, as Leysieffer and Warner (1976), Fligner et al. (1977), Nayak (1994) and others have noted, some papers have improperly compared different methods using other features that are not really comparable across those experiments.

For example, holding the probability of asking Q_1 the same in Warner's and Simmons' (see below) methods do not make them comparable in terms of privacy protection.

Next, all RR procedures can be divided into two groups depending on whether P is fully known or not. For an example, we consider a widely known variation of Warner's procedure, called Simmons' unrelated question method (Greenberg et al., 1969), which uses an unrelated nonsensitive question Q_3: Do you belong to B? in place of Q_2. For the transformation in Warner's and Simmons' methods, the domain is $\{A, A^c\}$ and the range is $\{Yes, No\}$. However, in Warner's method, the transition probabilities are known, but in Simmons' method they are known only if $P(B)$, the probability of answering "yes" to Q_3, is known. If P is unknown, generally multiple samples from different settings (and correspondingly specialized theory) are needed for estimating Π. In contrast, for known P, one can estimate Π from one sample and fairly easily. For protecting privacy and confidentiality, we do not see a need for using an RR method with unknown P. Thus, we shall only consider RR procedures with known P.

Let $\lambda_i = P(Z = d_i), i = 1, \ldots, m$, and $\lambda = (\lambda_1, \ldots, \lambda_m)'$. Then, under multinomial sampling, i.e., random sampling from an infinite population or simple random sampling with replacement if the population is finite, $\lambda_i = \sum_{j=1}^{k} p_{ij}\pi_j$. In matrix notation, we have

$$\lambda = P\Pi. \tag{2.1}$$

We can use RR data to make inferences about λ, which in turn can be used to make inferences about Π, using (2.1). In particular, if $m = k$ and P is nonsingular, an estimator $\hat{\lambda}$ of λ yields the estimator

$$\hat{\Pi} = P^{-1}\hat{\lambda}. \tag{2.2}$$

of Π. However, if $m \neq k$, generally (2.1) will yield no solution or multiple solutions for Π, causing difficulties in estimating Π from RR data. For $k = 2$, Nayak and Adeshiyan (2009) showed that for any RR procedure with $m > 3$, there exists one RR procedure with $m = 2$ that is better in terms of privacy protection and statistical efficiency. Thus, in the following we shall only consider RR procedures for which $k = m$ and P is a known nonsingular matrix.

Next, note that in general, the labels $d_1, \ldots, d_m$ for possible outputs of an RR procedure are arbitrary. In particular, any permutation of the labels does not alter the procedure or its

properties, but changes the transition probability matrix. This implies that two RR procedures with transition probability matrices P_1 and P_2 are statistically equivalent if P_1 can be obtained by permuting the rows of P_2. Obviously, P_1 and P_2 can be equivalent only if they have common dimensions, i.e., the ranges of the two procedures contain the same number of elements. As an example, two Warner's methods, where Q_1 is asked with probabilities p and $(1-p)$, respectively, are equivalent, as one can be obtained from the other by interchanging the two responses "Yes" and "No". For $k = m = 2$, i.e., when the original and perturbed variables are binary, Nayak (1994) imposed $p_{11} > p_{12}$ and $p_{22} > p_{21}$ to characterize each RR procedure uniquely by its transition probabilities. It will be interesting to develop similar constraints to specify P uniquely for general k and m. However, in the following, we shall only consider RR procedures with output space $\{c_1, \ldots, c_k\}$. This is helpful in two ways. First, fixing the range to $\{c_1, \ldots, c_k\}$ makes all transition probability matrices directly comparable. In particular, P_1 and P_2 would be equivalent if and only if $P_1 = P_2$. Second, for data users, released data are easier (and less confusing) to understand when they are reported using the categories of original variables.

As we noted in Section 1, in the context of confidentiality protection, P can be chosen based on the original data. Whether P depends on the data or not is an important consideration in RR theory, because if P depends on the data, then it is a random matrix and this should be properly accounted for in statistical inferences. We shall elaborate this further in Section 4.

Logically, to design an RR survey, we should first choose a transition probability matrix P, considering both privacy protection and statistical efficiency, and then devise an experiment to implement it. First, any P can be implemented using k experiments, where the jth experiment is used only by respondents with $X = c_j$ and it outputs $c_1, \ldots, c_k$ with probabilities $p_{1j}, \ldots, p_{kj}$. This can be accomplished, for example, by using k decks of cards, one for each category of X. This is similar to Kuk's (1990) RR procedure. In this approach, each respondent needs to select and use a randomization mechanism (e.g., one of k decks of cards), unobserved by the interviewer, based on his true category of X. Nonetheless, respondents may be suspicious of such methods as a respondent's experiment selection reveals his true category of X.

Next, we discuss another view of RR procedures, following Nayak (1994), and show that it is

also possible to devise a common experiment, independent of true X category, for implementing any given P. Randomized response is often considered as a random transformation of the true response. Mathematically, a transformation is a function and a random transformation can be viewed as probabilistically selecting one function from a set of functions and then applying it to the true response to generate a randomized response. For $k = 2$, there are four possible functions from $\{c_1, c_2\}$ to $\{c_1, c_2\}$, given by $g_1 : c_1 \to c_1, c_2 \to c_2$ (i.e., $g(c_1) = c_1$ and $g(c_2) = c_2$); $g_2 : c_1 \to c_2, c_2 \to c_1$; $g_3 : c_1 \to c_1, c_2 \to c_1$; $g_4 : c_1 \to c_2, c_2 \to c_2$ and any RR procedure essentially uses $g_1, \ldots, g_4$ with some probabilities $\alpha_1, \ldots, \alpha_4$. Two procedures can be differentiated by their values of $\alpha_1, \ldots, \alpha_4$. Note that g_1 and g_2 correspond to using Q_1 and Q_2, and g_3 and g_4 represent *forced A* and *forced* A^c responses (labeling A and A^c by c_1 and c_2). Thus, for example, Warner's method uses only g_1 and g_2 with probabilities p and $(1-p)$, and the triangular method of Tan et al. (2009) used only g_1 and g_3 (with given probabilities). As Quatember (2009) noted, many existing methods use different combinations of Q_1, Q_2, Q_3 and *forced A* and *forced* A^c responses. However, using Q_3 is equivalent to using g_3 and g_4 with probabilities $P(B)$ and $1 - P(B)$, respectively.

More generally, for X with k categories, there are k^k functions from $\{c_1, \ldots, c_k\}$ to $\{c_1, \ldots, c_k\}$, which we shall denote by $g_1, \ldots, g_{k^k}$. Let $\alpha = (\alpha_1, \ldots, \alpha_{k^k})$ be a probability distribution on the set of these functions. An RR procedure may be viewed as selecting one function according to α and applying it to the true response. Here, the randomization process is the same for all respondents. For given α, the transition probabilities are:

$$p_{ij} = P(Z = c_i | X = c_j) = \sum_{l=1}^{k^k} \alpha_l P(Z = c_i | X = c_j, g_l), \quad i, j = 1, \ldots, k, \tag{2.3}$$

where $P(Z = c_i | X = c_j, g_l)$ is the conditional probability of $Z = c_i$ given that $X = c_j$ and g_l is selected by the randomization process. Since each g_l is a (deterministic) function, this conditional probability is either 0 or 1. Specifically, $P(Z = c_i | X = c_j, g_l) = 1$ if $g_l(c_j) = c_i$; otherwise $P(Z = c_i | X = c_j, g_l) = 0$. Obviously, any α induces a unique transition probability matrix P. Conversely, given any transition probability matrix P, we can find α to satisfy (2.3). Note that there are $k(k-1)$ equations in (2.3), in view of the fact that $\sum_i p_{ij} = 1$ for $j = 1, \ldots, k$, and $(k^k - 1)$ unknowns (in α). Thus, the number of unknowns is much larger than the number of

equations, and for given P, (2.3) will have multiple solutions. We can use any of those solutions to implement P by an experiment that is common to all respondents.

3. Statistical Estimation for Fixed P

Most authors have considered estimating Π under multinomial sampling when P is known and fixed. Let n denote the sample size, T_i and S_i denote the sample frequencies of $X = c_i$ and $Z = c_i$, respectively, $T = (T_1, \ldots, T_k)'$ and $S = (S_1, \ldots, S_k)'$. Then, $T \sim Mult(n, \Pi)$, $S \sim Mult(n, \lambda)$ and $\hat{\lambda} = S/n$ is the MLE (and UMVUE) of λ, which in turn yields

$$\hat{\Pi} = P^{-1}\hat{\lambda} = P^{-1}(S/n), \tag{3.1}$$

as an unbiased estimator of Π with

$$Var(\hat{\Pi}) = \frac{1}{n}(D_\Pi - \Pi\Pi') + \frac{1}{n}[P^{-1}D_\lambda(P^{-1})' - D_\Pi], \tag{3.2}$$

where D_Π is a diagonal matrix with diagonal elements being $\pi_1, \ldots, \pi_k$ and D_λ is defined similarly (see Chaudhuri and Mukerjee, 1988, p. 43). The first term on the right side of (3.2) is the sampling variance and the last term is the additional variance due to randomization. Most authors estimate (3.2) by replacing Π by $\hat{\Pi}$ and give interval estimates only for large n when $\hat{\Pi}$ is approximately normally distributed. However for small and moderate n, both $\hat{\Pi}$ and interval estimates can take values outside the interval $[0, 1]$. Frey and Perez (2012) nicely discuss this aspect for binary X and present a novel method for constructing exact confidence intervals.

While research on RR methods started long ago and in survey context, inferences under general sampling designs have been discussed only recently, initiated by Padmawar and Vijayan (2000). Their basic ideas have been used and further developed by Chaudhuri (2001, 2004), Nayak and Adeshiyan (2009) and others. In the following, we summarize some results from Adeshiyan (2011). Consider a finite population of N units, labeled $i = 1, \ldots, N$, and suppose a sample s, which is a subset of $\{1, \ldots, N\}$, is selected using a non-informative sampling design $p(s)$. As before, the survey variable X is categorical with k categories and we observe only a randomized version (Z) of it, where the transition probability matrix P is known, fixed and nonsingular.

For unit i ($i = 1, \ldots, N$), we introduce two k-dimensional indicator vectors I_i and J_i. Specifically, if the true category of unit i is c_j, then the jth component of I_i is 1 and all other components are 0. We define J_i similarly, to record the (randomized) response category of unit i, if it is sampled. Thus, the data can be represented as $\{(i, J_i); i \in s\}$. In this setup, consider estimating $\tau = \sum_{i=1}^{N} I_i$, which is equivalent to estimating $\Pi = (\sum_{i=1}^{N} I_i)/N$. If the true categories, i.e., $I_i, i \in s$ were observed, Π can be estimated unbiasedly using the Horvitz-Thompson estimator. More generally, suppose $\hat{\tau} = \sum_{i \in s} w_{si} I_i$ is a homogeneous linear (design) unbiased estimator of τ based on original data, i.e.,

$$\sum_{s} (\sum_{i \in s} w_{si} I_i) p(s) = \sum_{i=1}^{N} I_i \quad \text{for all } I_1, \ldots, I_N.$$

Then, from the fact $E(J_i|I_i) = I_i$, where the expectation is with respect to the RR process, it follows that $\hat{\tau}^* = \sum_{i \in s} w_{si} \left(P^{-1} J_i\right)$ is a design unbiased estimator of τ based on RR survey data. Thus, any unbiased estimator for an open survey can be easily modified to obtain an unbiased estimator when RR is used.

Adeshiyan (2011) showed that the variance of $\hat{\tau}^*$ can be expressed as

$$V(\hat{\tau}^*) = E_p[\sum_{i \in s} w_{si}^2 \left\{P^{-1} D_{PI_i} (P')^{-1}\right\} - D_{I_i}] + V_p(\hat{\tau}),$$

where E_p denotes expectation with respect to the sampling design,

$$V_p(\hat{\tau}) = \sum_{i=1}^{N} b_i D_{I_i} + \sum_{i,j=1 i \neq j}^{N} b_{ij} I_i I_j',$$

$b_i = \sum_{s \ni i} w_{si}^2 p(s) - 1$ and $b_{ij} = \sum_{s \ni i,j} w_{si} w_{sj} p(s) - 1$. An unbiased estimator of $V(\hat{\tau}^*)$ is also presented in Adeshiyan (2011).

4. Estimation Under Invariant Post-randomization

Statistical agencies often release only a perturbed or masked version of original data to protect the confidentiality of respondent level information. Doyle et al. (2001) and Willenborg and De Waal (2001) described a variety of methods, such as grouping, top coding, swapping, multiple imputation and noise addition, for creating perturbed data. Warner (1971) indicated that RR

techniques may be used for confidentiality protection. Gouweleeuw et al. (1998) developed the idea as the Post-randomization Method (PRAM) for perturbing categorical data. As before, consider a categorical variable X, which may also represent the cross-classification of several variables. Just as in RR surveys, PRAM converts the true categories to randomized responses (categories) using a known transition probability matrix P, also called the PRAM matrix. The randomization process is applied to each observation in the data set, independently of all other records. Also, this process is carried out by the survey organization rather than the respondents. For further discussion of PRAM and additional references, we refer the reader to Gouweleeuw et al. (1998), Van den Hout and Van der Heijden (2002), Van den Hout and Elamir (2006), Cruyff et al. (2008), Shlomo and Skinner (2010), Nayak and Adeshiyan (2015) and Nayak et al. (2015).

For fixed P, PRAM is mathematically equivalent to RR surveys and hence all inferential methods developed for RR surveys are applicable to data perturbed by PRAM. However, data agencies rarely release the values of the parameters used in their perturbation procedures, such as P for PRAM. Thus, data users usually would not know the PRAM matrix P and hence would not be able to use the results for known P, as in the preceding section. In response to this obstacle, statistical agencies try to use data perturbation procedures for which standard inferential methods for original data remain valid, at least approximately, for masked data. This motivated Gouweleeuw et al. (1998) to define a PRAM to be an invariant PRAM if P satisfies

$$PT = T \quad \text{or equivalently} \quad P\hat{\Pi}_0 = \hat{\Pi}_0, \tag{4.1}$$

where $\hat{\Pi}_0 = T/n$. Since $E[S|T] = PT$, under multinomial sampling, (4.1) implies that $\hat{\Pi}_* = S/n$, the relative frequency vector based on perturbed data, is an unbiased estimator of Π. Also, $\hat{\Pi}_*$ is always a probability vector and it can be calculated without using P or its inverse.

To apply invariant PRAM, we need to select and use one P satisfying (4.1). We should note that generally (4.1) has many solutions for P and the solution space of (4.1) is a non-empty convex set, which also contains the identity matrix. Gouweleeuw et al. (1998) and Nayak and Adeshiyan (2015) give some methods for solving (4.1). One important point to note is that solutions of (4.1) depend on original data through T. Thus, in invariant PRAM, P is a random

matrix. Consequently, mathematical results for fixed P may not hold for invariant PRAM. In the following we summarize some properties of $\hat{\Pi}_*$, from Nayak and Adeshiyan (2015).

Let $P = [P_1 : \ldots : P_k]$, i.e., denote the ith column of P by P_i, and rewrite (4.1) as

$$\sum_{i=1}^{k} T_i P_i = T. \tag{4.2}$$

Let F_{ij} denote the number of units whose category changed from c_i to c_j due to PRAM, and let $F_i = (F_{i1}, \ldots, F_{ik})'$. Then, $S = \sum_i^k F_i$, and given T and P, $F_1, \ldots, F_k$ are independently distributed with $F_i \sim Mult(T_i, P_i), i = 1, \ldots, k$. Now,

$$E(\hat{\Pi}_*|T, P) = \frac{1}{n}\sum_{i=1}^{k} E[F_i|T, P] = \frac{1}{n}\sum_{i=1}^{k} T_i P_i = \frac{T}{n}, \tag{4.3}$$

by (4.2). From (4.3), it follows easily that under multinomial sampling $E(\hat{\Pi}_*) = \Pi$, i.e., $\hat{\Pi}_*$ is an unbiased estimator of Π. It can also be seen that

$$V(\hat{\Pi}_*|T, P) = \frac{1}{n^2}\sum_{i=1}^{k} T_i[D_{P_i} - P_i P_i'] = \frac{1}{n}[D_{\hat{\Pi}_0} - \sum_{i=1}^{k}(\frac{T_i}{n})P_i P_i'], \tag{4.4}$$

and

$$\begin{aligned} V(\hat{\Pi}_*) &= V[E(\hat{\Pi}_*|T, P)] + E[V(\hat{\Pi}_*|T, P)] \\ &= V(\hat{\Pi}_0) + \frac{1}{n}[D_\Pi - E\{\sum_{i=1}^{k}(\frac{T_i}{n})P_i P_i'\}]. \end{aligned} \tag{4.5}$$

Notice the difference between the last terms of (3.2) and (4.5), the variance inflations due to PRAM with fixed P and invariant PRAM. The last expectation in (4.5) also involves the conditional distribution of P given T, which is determined by the data agency's algorithm (or method) for choosing a solution of (4.1) for P. Typically, data users would not know P or the agency's full process for selecting P and hence would not be able to evaluate or estimate $V(\hat{\Pi}_*)$. These observations show that under invariant PRAM, data users will be able to calculate an unbiased estimate of Π (using $\hat{\Pi}_*$), but not its sampling variance (and hence reliability). However, as Nayak and Adeshiyan (2015) describe, the data agency can calculate (4.4) and also estimates of (4.5). They also proved that

$$V_{max}(\hat{\Pi}_*) = (2 - \frac{1}{n})[\frac{D_\Pi - \Pi\Pi'}{n}] \tag{4.6}$$

is a tight upper bound of $V(\hat{\Pi}_*)$, in the sense that $[V_{max}(\hat{\Pi}_*) - V(\hat{\Pi}_*)]$ is nonnegative definite for all invariant PRAM and there exists an invariant PRAM for which $V(\hat{\Pi}_*) = V_{max}(\hat{\Pi}_*)$.

An obvious lower bound for $V(\hat{\Pi}_*)$ is $V(\hat{\Pi}_0) = [D_\Pi - \Pi\Pi]/n$. These upper and lower bounds can be estimated by replacing Π with $\hat{\Pi}_*$. Estimating linear combinations of Π is a common problem. A natural (and unbiased) estimator of $a'\Pi$ is $a'\hat{\Pi}_*$. The upper and lower bounds of $V(\hat{\Pi}_*)$ immediately yields bounds for $V(a'\hat{\Pi}_*) = a'[V(\hat{\Pi}_*)]a$. We conclude this section by mentioning that Nayak and Adeshiyan (2015) also present some results on estimation under invariant PRAM and general probability sampling.

5. Assessing Privacy and Confidentiality Protection

The main reason for randomizing true responses, although it reduces data quality, is privacy and confidentiality protection. Thus, we must assess how well a method protects privacy or confidentiality. However, this is a very difficult task, because privacy and confidentiality are complex, multifaceted concepts. Disclosure of private or confidential information can occur in many different ways and forms (see Willenborg and De Waal, 2001), and it is difficult to develop criteria for measuring privacy and confidentiality protection (or lack of it) that are both credible and widely applicable. Lambert (1993) discusses several intrinsic challenges in measuring risk and consequences of disclosure. While many researchers have investigated this topic, precise and substantive results have been obtained only for some specific situations.

In RR survey context, most work on measuring and comparing privacy protection has focused on binary survey variables with one category being stigmatizing, as Warner (1965) considered. For reviewing the results, it will be convenient to use Warner's (1965) terminology (but not limited to his experiment). Suppose X is a binary variable with two categories A and A^c, with $\pi = P(A)$, and only A is sensitive or stigmatizing. Suppose the randomized response also has two categories, denoted Y and $\bar{Y}$ (for "Yes" and "No", but they can also mean other responses). Here, a transition probability matrix P is of order 2×2 and has two free elements, as each column adds to 1. Thus, each procedure can be characterized by the two probabilities embedded in it, viz., $\gamma = P(Y|A)$ (i.e., the probability that a respondent answers 'Yes' given

that his/her true category is A) and $\delta = P(Y|A^c)$. The pair (γ, δ) defines the *design* of an RR procedure. However, considering the effects of interchanging Y and $\bar{Y}$ (as discussed in Section 2), it is seen that two procedures with designs (γ, δ) and (δ, γ) are equivalent. Following Nayak (1994), the *design space* can be specified appropriately as $\mathcal{D} = \{(\gamma, \delta) : 0 \leq \delta < \gamma \leq 1\}$, where we associate Y with A and $\bar{Y}$ with A^c (which seems natural).

Many authors, including Leysieffer and Warner (1976), Lanke (1976) and Fligner et al. (1977), Nayak (1994) and Guerriero and Sandri (2007), have discussed measuring privacy protection in the preceding context. Generally, all measures involve the two posterior probabilities

$$P(A|Y) = \frac{\gamma\pi}{\gamma\pi + \delta(1-\pi)} \quad \text{and} \quad P(A|\bar{Y}) = \frac{(1-\gamma)\pi}{(1-\gamma)\pi + (1-\delta)(1-\pi)} \tag{5.1}$$

and suggest that protection increases as these two probabilities decrease. It can be seen that the estimator $\hat{\pi}$ of π based on (3.1) has variance $V(\hat{\pi}) = [\theta(1-\theta)]/[n(\gamma-\delta)^2]$, where $\theta = \delta + (\gamma - \delta)\pi$. Noting that we desire $P(A|Y), P(A|\bar{Y})$ and $V(\hat{\pi})$ to be small and all three quantities depend on unknown π, Nayak (1994) presented the following.

Definition 5.1. *An RR design $D_1 = (\gamma_1, \delta_1)$ is said to be better than another design $D_2 = (\gamma_2, \delta_2)$, or D_1 dominates D_2, if $P_{D_1}(A|Y) \leq P_{D_2}(A|Y), P_{D_1}(A|\bar{Y}) \leq P_{D_2}(A|\bar{Y})$ and $V_{D_1}(\hat{\pi}) \leq V_{D_1}(\hat{\pi})$ for all $0 \leq \pi \leq 1$, with at least one strict inequality holding for some π.*

An RR design D is said to inadmissible if it is dominated by some other design D_*. Otherwise, D is said to be admissible.

Theorem 5.1. *An RR procedure with design (γ, δ) is admissible if and only if $\gamma = 1$.*

Nayak (1994) discussed some important implications of this result. It shows that any admissible procedure must require all respondents in A to respond Y. In particular, any procedure that uses Q_2 or Q_3 or 'forced no' cannot be admissible. In terms of the functions g_1, g_2, g_3, g_4 in Section 2, an admissible procedure can use only g_1 and g_3. We can also see easily that of the 15 RR procedures in Table 1 of Quatember (2009), only one is admissible.

Now suppose X is binary and only A is stigmatizing, but the RR variable Z has m categories, $d_1, \ldots, d_m$ (e.g., Leysieffer and Warner, 1976; Kuk, 1990; Christofides, 2003). Consider $\psi = \max\{P(A|Z = d_1), \ldots, P(A|Z = d_m)\}$ as a (summary) measure for respondent's risk (of being

classified in A). With this, Nayak and Adeshiyan (2009) proved that for any RR procedure $\mathcal{R}$ with $m \geq 3$ response categories, there exists an RR procedure $\mathcal{R}_*$ with 2 response categories such that $\mathcal{R}_*$ is at least as good as $\mathcal{R}$ in terms of both privacy protection and statistical efficiency.

If both A and A^c are sensitive or stigmatizing, measuring and controlling privacy protection requires a different approach. Intuitively, for any response, say Y, we would like both $P(A|Y)$ and $P(A^c|Y)$ to be small, but that is not sensible as $P(A|Y) + P(A^c|Y) = 1$. However, some papers, e.g., Quatember (2009) and Chaudhuri (2010), have discussed the topic. Similar issues arise in post-randomization for confidentiality protection, where true variables commonly have multiple categories of which many may be sensitive. Also, confidentiality protection goals can be quite different. Some methods for evaluating the efficacy of PRAM in disclosure control and ideas on how to choose the PRAM matrix have appeared in Gouweleeuw et al. (1998), Van den Hout and Elamir (2006), Shlomo and De Waal (2008) and Shlomo and Skinner (2010). We believe that additional theoretical and empirical research will help to develop PRAM as a valuable disclosure control method.

6. Discussion

Many RR techniques for privacy and confidentiality protection have appeared in the literature. We discussed some basic elements of RR mechanisms and stressed that for proper understanding, unification and fair comparison, RR procedures should be characterized and examined through their transition probability matrices. The main factors that affect inferential methods are the dimensions and rank of the transition probability matrix P, and whether P is (i) fully known or not and (ii) fixed or selected based on the data. However, RR procedures with known, square and nonsingular P are most convenient for statistical estimation and they should also be adequate for protecting respondents' privacy and confidentiality.

A common assumption in mathematical theory of RR methods is that survey participants respond truthfully, which is unrealistic. Also, respondents' perception of privacy protection may depend on secondary features, such as familiarity and complexity, of the RR experiment. Thus, respondents' participation and truthfulness may be different in two procedures with a common

P. In practice, one should assess respondents' willingness and concerns by conducting pilot studies, and utilize the findings in selecting an appropriate RR procedure. For devising an RR experiment, after choosing P, it may be helpful to think in terms of selecting a mathematical function probabilistically and then applying it to the respondent's true category, as discussed in Section 2. This approach offers much flexibility, as any P can be executed by many probability distributions on the set of all functions and one may use secondary criteria to choose one of those; see Nayak (1994) for some specific proposals and related results for a binary variable.

Acknowledgment. We thank Mr. Joseph Conklin and Professor Lucio Barabesi for carefully reading an earlier draft and giving helpful suggestions for improving the presentation.

References

[1] Adeshiyan, S. A., 2011. Unification of Randomized Response Designs and Certain Aspects of Post-Randomization for Statistical Disclosure Control. Doctoral dissertation, The George Washington University.

[2] Brand, R., 2002. Microdata protection through noise addition. In: J. Domingo-Ferrer (Ed.), Inference Control in Statistical Databases, Springer, Berlin, 97-116.

[3] Chaudhuri, A., 2001. Using randomized response from a complex survey to estimate a sensitive proportion in a dichotomous finite population. Journal of Statistical Planning and Inference 94 (1), 37-42.

[4] Chaudhuri, A., 2004. Christofides randomized response technique in complex sample surveys. Metrika 60 (3), 223-228.

[5] Chaudhuri, A., 2010. Randomized Response and Indirect Questioning Techniques in Surveys, CRC Press, Boca Raton.

[6] Chaudhuri, A., Christofides, T.C., 2013. Indirect Questioning in Sample Surveys, Springer, New York.

[7] Chaudhuri, A., Mukerjee, R., 1988. Randomized Response: Theory and Techniques, Marcel Dekker, New York.

[8] Christofides, T. C., 2003. A generalized randomized response technique. Metrika 57 (2), 195-200.

[9] Cruyff, M.J.L.F., Van Den Hout, A., Van Der Heijden, P.G.M., 2008. The analysis of randomized response sum score variables. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 70 (1), 21-30.

[10] Doyle, P., Lane, J., Theeuwes, J., Zayatz, L. (Eds.), 2001. Confidentiality, Disclosure and Data Access: Theory and Practical Applications for Statistical Agencies, Elsevier, Amsterdam.

[11] Evans, T., Zayatz, L., Slanta, J., 1998. Using noise for disclosure limitation of establishment tabular data. Journal of Official Statistics 14 (4), 537-551

[12] Fligner, M.A., Policello, G.E., 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 (15), 1511-1524.

[13] Frey, J., Prez, A., 2012. Exact binomial confidence intervals for randomized response. The American Statistician 66 (1), 8-15.

[14] Fuller, W.A., 1993. Masking procedures for microdata disclosure limitation. Journal of Official Statistics 9 (2), 383-406.

[15] Gouweleeuw, J.M., Kooiman, P., Willenborg, L.C.R.J., De Wolf, P.-P., 1998. Post randomisation for statistical disclosure control: theory and implementation. Journal of Official Statistics 14 (4), 463–478.

[16] Greenberg, B.G., Abul-Ela, A-L. A., Simmons, W.R., Horvitz, D.G., 1969. The unrelated question randomized response model: theoretical framework. Journal of the American Statistical Association 64 (326), 520-539.

[17] Guerriero, M., Sandri, M.F., 2007. A note on the comparison of some randomized response procedures. Journal of Statistical Planning and Inference 137 (7), 2184-2190

[18] Kuk, A.Y.C., 1990. Asking sensitive questions indirectly. Biometrika 77 (2), 436-438.

[19] Lambert, D., 1993. Measure of disclosure risk and harm. Journal of Official Statistics 9 (2), 313-331.

[20] Lanke, J., 1976. On the degree of protection in randomized interviews. International Statistical Review 44 (2), 197-203.

[21] Leysieffer, R.W., Warner, S.L., 1976. Respondent jeopardy and optimal designs in randomized response models. Journal of the American Statistical Association 71 (355), 649-656.

[22] Nayak, T.K., 1994. On randomized response surveys for estimating a proportion. Communications in Statistics - Theory and Methods 23 (11), 3303-3321.

[23] Nayak, T.K., 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 (8), 2757-2766.

[24] Nayak, T.K., Adeshiyan, S.A., 2015. On invariant post-randomization for statistical disclosure control. International Statistical Review (to appear), doi:10.1111/insr.12092.

[25] Nayak, T.K., Zhang, C., Adeshiyan, S.A., 2015. Emerging applications of randomized response concepts and some related issues. Model Assisted Statistics and Applications (to appear).

[26] Nayak, T.K., Sinha, B., Zayatz, L., 2011. Statistical properties of multiplicative noise masking for confidentiality protection. Journal of Official Statistics 27 (3), 527-544.

[27] Padmawar, V.R., Vijayan, K., 2000. Randomized response revisited. Journal of Statistical Planning and Inference 90 (2), 293-304.

[28] Quatember, A., 2009. A standardization of randomized response strategies. Survey Methodology 35 (2), 143-152.

[29] Shlomo, N., De Waal, T., 2008. Protection of micro-data subject to edit constraints against statistical disclosure. Journal of Official Statistics 24 (2), 229-253.

[30] Shlomo, N., Skinner, C., 2010. Assessing the protection provided by misclassification-based disclosure limitation methods for survey microdata. The Annals of Applied Statistics 4 (3), 1291-1310.

[31] Tan, M. T., Tian, G. L., Tang, M. L., 2009. Sample surveys with sensitive questions: a nonrandomized response approach. The American Statistician 63 (1).

[32] Van den Hout, A., Van der Heijden, P.G.M., 2002. Randomized response, statistical disclosure control and misclassification: a review. International Statistical Review 70 (2), 269-288.

[33] Van den Hout, A., Elamir, E.A.H., 2006. Statistical disclosure control using post randomisation: variants and measures for disclosure risk. Journal of Official Statistics 22 (4), 711-731.

[34] Warner, S.L., 1965. Randomized response: a survey technique for eliminating evasive answer bias. Journal of the American Statistical Association 60 (309), 63-69.

[35] Warner, S.L., 1971. The linear randomized response model. Journal of the American Statistical Association 66 (366), 884-888.

[36] Willenborg, L.C.R.J., De Waal, T., 2001. Elements of Statistical Disclosure Control, Springer, New York.