Unification of Randomized Response Designs and Certain Aspects of Post-Randomization for Statistical Disclosure Control

by Samson A. Adeshiyan

B.Sc. (Hons), August 1991, University of Ibadan, Ibadan, Nigeria

M.Sc., February 1995, Shanghai University of Science and Technology, P. R. China

A Dissertation submitted to

The Faculty of

Columbian College of Arts and Sciences

of The George Washington University

in partial fulfillment of the requirements

for the degree of Doctor of Philosophy

May 15, 2011

Dissertation Directed by

Tapan K. Nayak

Professor of Statistics

The Columbian College of Arts and Science of The George Washington University certifies that Samson Ayoade Adeshiyan has passed the Final Examination for the degree of Doctor of Philosophy as of March 25, 2011. This is the final and approved form of the dissertation.

Unification of Randomized Response Designs and Certain Aspects of Post-Randomization for Statistical Disclosure Control

Samson A. Adeshiyan

Dissertation Research Committee:

Tapan K. Nayak, Professor of Statistics, Dissertation Director

Reza Modarres, Professor of Statistics, Committee Member

Sudip Bose, Associate Professor of Statistics, Committee Member

Dedication

To my dad who got me started

Acknowledgements

None of this work would have happened without Professor Tapan K. Nayak, so my foremost and profound gratitude goes to him. Over the course of my years as a student at GWU, he has won many hats. I first met Professor Nayak while I was a prospective student and he was department chair. He guided me through registering for my first class as a nondegree student and, subsequently, the admission process into the Statistics Ph.D. program. He was also the Ph.D. program director/academic advisor during most of my years at GWU. He taught me in class and ultimately was my dissertation advisor. Throughout my dissertation phase, he provided invaluable guidance and insightful comments that helped me clarify my ideas. His indefatigable approach demonstrated by his remarkable patience, ability to provide the necessary push and encouragement made this all come to fruition.

I thank the readers of my dissertation, Professors Reza Modarres and Sudip Bose. Their insightful comments, right from the proposal phase through the defense examination, were incorporated into this dissertation. I also thank my other examiners, Professor Michael Larsen and Dr. Promod Chandhok for their valuable comments and suggestions for extending this work. I am grateful to Professor Zhaohai Li for presiding over my final examination.

I thank the GWU Statistics Department for providing a viable graduate program even for students with full-time jobs. Thanks also go to my fellow students for their support. In particular, I acknowledge Dr. Mark Bauder, my study partner during our pre-candidacy phase. I also thank Dr. Min Qin for LaTex help.

I gratefully acknowledge the institutional and financial support that I received from the U.S. Census Bureau throughout my graduate studies at GWU. I am also grateful to

my supervisors and colleagues at work, most of whom were fully supportive and afforded me many accomodations. Special thanks to Mr. George Train for being a good friend over the years and for reading the latter sections of this dissertation.

I thank my parents for their support and prayers from the beginning. As I worked on this dissertation, they have been a constant source of encouragement especially at times when there appeared to be no end in sight.

Finally, I thank my wife, Olubukola, for her unreserved support. She and our son, Ayodeji, motivated and inspired me to finish this work.

Abstract

Unification of Randomized Response Designs and Certain Aspects of Post-Randomization for Statistical Disclosure Control

This dissertation deals with two closely related topics - randomized response (RR) surveys and post-randomization - that are concerned with survey respondents' privacy protection and confidentiality. First, we present a common framework for discussing various RR surveys of dichotomous populations with polychotomous responses. The unified approach addresses both respondents' privacy and statistical efficiency and is helpful for fair comparison of various procedures. We describe a general technique for constructing unbiased estimators of the proportion (π) of the population that belongs to a sensitive or stigmatized group based on arbitrary RR procedures, from unbiased estimators based on an open or direct survey with the same sampling design. The technique works well for any sampling design *p(s)* and also for variance estimation. We develop an approach for comparing RR procedures, taking both respondents' protection and statistical efficiency into account. For any given RR design with three or more response categories, we can find RR procedures with a binary response variable which provide the same respondents' protection and at least as much statistical information. This result suggests that RR surveys of dichotomous populations should use only binary response variables.

In many situations there may be more than two natural population categories, so we also investigate RR surveys for polychotomous populations, with k categories of which at least one is sensitive or stigmatized. We extend the theory and framework for RR surveys of dichotomous populations to RR surveys of polychotomous populations, including estimation in finite population settings. We also discuss comparison of polychotomous

RR designs where only one category is sensitive.

The second topic is post-randomization (PRAM), which is a statistical disclosure control technique for categorical variables. The PRAM stochastically transforms each record in a microdata set using pre-selected probabilities. We demonstrate that any PRAM procedure can be regarded as a PRAMing of the cross-classification of all the variables in the data set. We discuss some connections to RR surveys and note that the estimators developed for RR surveys are applicable for estimation from PRAMed data. We focus on a special case of PRAM, known as invariant PRAM and introduce the notion of a strongly invariant PRAM. The invariant PRAM is attractive in that in the *strong* situation, the PRAMed data can be analyzed without adjustment for post-randomization. We review methods for constructing invariant PRAM matrices, clarify certain misconceptions about invariant PRAM, and discuss estimation from an invariantly PRAMed microdata set. Finally, we examine the effectiveness of PRAM for limiting statistical disclosure.

Contents

Chapter 1

Introduction

This dissertation is about two closely connected topics that deal with data protection and confidentiality: The first topic is randomized response (RR) designs where we present a unified framework for the analysis of and comparison of RR surveys of binary and polychotomous characteristics. The second topic is post-randomization (PRAM), which is a statistical disclosure control technique. The following sections of this chapter introduce RR designs and the PRAM, and provide an outline of the dissertation.

1.1 Randomized Response Designs

In most surveys, the individuals selected in the sample are asked to answer direct questions relating to the survey variables. However, for questions on sensitive or stigmatizing characteristics such as criminal history, tax evasion, drug abuse, gambling and abortion, many respondents are unwilling to give honest answers, if at all they respond in the first place. The refusals and false answers lead to biased and unreliable estimates. The main reason for lack of respondents' cooperation is the lack of privacy.

To increase truthful respondent participation, Warner (1965) proposed the first randomized response (RR) procedure for a binary characteristic, which allowed respondents to respond truthfully to questions without revealing their personal information in the course of the survey. RR generates misclassified categorical variables with known mis-

classified probabilities. It is this misclassification that protects the privacy of the personal information of each individual respondent. Of the different types of surveys used in practice, such as mail, telephone, internet, etc., RR methods work only for personal visits which are otherwise known as face-to-face interviews.

Consider a dichotomous population where each person belongs either to a sensitive group A or to its complement A^c. The objective is to estimate the true proportion (π) of the population that belongs to group A. In Warner's (1965) method, each interviewee first selects one of the two questions:

Q_1: Do you belong to A?

Q_2: Do you belong to A^c?

with respective probabilities p and $(1-p)$, by performing a random experiment, unobserved by the interviewer. One example of a random experiment that implements RR surveys involves supplying each respondent with a deck of cards. The cards bear questions regarding membership in group A or A^c (i.e., Q_1 and Q_2) in proportions p and $(1-p)$ respectively. The respondent is asked to draw a card after shuffling the deck and reply to the question inscribed on the selected card. The respondent then truthfully replies "Yes" or "No" to the selected question without disclosing the question and thereby protecting his/her privacy. The process of selecting the question is unobserved by the interviewer. Also, the respondent does not disclose the question to which his/her answer applies. So, even though the interviewer gets a truthful "Yes" or "No" response from each respondent, the randomization procedure ensures that personal information regarding whether the respondent belongs to group A or A^c cannot be retrieved by the interviewer. With this understanding of the privacy protection provided by RR designs, the respondents are expected to respond truthfully to an RR survey. The probability p is chosen by the interviewer in designing the RR survey. The choice of p plays a dual role in determining the extent of privacy protection and the efficiency of estimates of π

obtained from the RR design.

There are two main types of variations to RR designs in the literature:

1. Change in the randomization process

2. Change in the question set

An example of a change in randomization process can be found in Mangat and Singh (1990) where they proposed the following two-stage randomization procedure: At the first stage, a binary experiment with $P(S) = T$ and $P(F) = 1 - T$ is performed. If the outcome is S then the respondent answers Q_1; otherwise, the respondent goes to the second stage and follows the Warner's method as described above.

An example of a change in question set is found in Greenberg et al. (1969) which discussed a related procedure, called Simmons' unrelated question method, in which the question Q_2 in Warner's method is replaced by an unrelated nonsensitive question:

Q_3: Do you belong to B?

An example of an unrelated question is: were you born in the month of June? The probabilities of "Yes" (Y) and "No" (N) responses to Q_3 may be known, in which case, a method of moments estimator ($\hat{\pi}_U$) of π can be derived easily. Greenberg et al. (1969) compared the variances of the estimates from their method to that of Warner. They showed that their design was more efficient when the probability (p) of asking the direct question Q_1 is the same in the two methods.

Some of the comparisons of various methods that have been done in past papers have not been fair, as noted by Leysieffer and Warner (1976) and Fligner et al. (1977), because competing procedures with a common value of p offer different degrees of privacy to the respondents. In particular, due to the variations in various RR designs, although two designs may have common parameter p, they may afford different degrees of privacy protection. For fair comparison, the two procedures should be required to

offer equal respondents' protection. Some unfair comparisons stemmed from considering two procedures with a common randomization parameter, but with disparate impact on respondents' protection, and then comparing variances of the estimators proposed under the two procedures. We believe, misleading comparisons could be avoided by discussing various RR procedures within a common framework.

Some RR procedures proposed in the literature use polychotomous responses (e.g., Leysieffer and Warner, 1976; Kuk, 1990; Christofides, 2003). Most of these procedures were compared to, say, Warner's original procedure without finding a way to hold the respondent's protection constant. This often led to unfair comparisons.

In this research, we developed theory and framework for other RR designs beyond those of dichotomous populations with binary response variables introduced above. We adopt an incremental approach by first examining RR surveys of dichotomous populations with polychotomous response variables. Overall, we present a unified framework for analysis and comparison of randomized response surveys of binary characteristics and clarified certain issues relating to statistical estimation and comparison of RR surveys. Next, we examined a general framework for analyzing RR surveys of polychotomous populations with polychotomous response variables. For both binary and polychotomous RR designs, we began our development under an infinite population setting and then extend our results to situations in finite population settings. Most of the results obtained in this part of the dissertation research have been presented in Nayak and Adeshiyan (2009).

1.2 Post-Randomization

Statistical agencies around the world apply Statistical Disclosure Control (SDC) techniques before releasing public data products in order to protect the confidentiality of respondents to their various surveys and censuses. In particular, federal agencies in

the United States are bound by federal laws to protect the data that they collect from individuals and businesses. The US Census Bureau, for instance, is bound by Title 13 of the United States Code. There are other federal laws including the Confidential Information Protection and Statistical Efficiency Act and the Privacy Act that reinforce protections of data. As a result, federal agencies use SDC techniques ranging from simply stripping off identifying information from the data to more sophisticated data masking techniques such as adding noise, grouping, cell suppression, data swapping, multiple imputation/generation of synthetic data, etc. See Doyle et al. (2001) and Willenborg and De Waal (2001) for more detailed discussions of various disclosure control techniques. We are particularly interested in examining the post-randomization method (PRAM) for controlling statistical disclosure.

The PRAM method for controlling statistical disclosure, introduced by Kooiman et al. (1997) and further discussed by by Gouweleeuw et al. (1998), is concerned with protecting respondents' privacy while releasing microdata (for data already collected) for public use. PRAM stochastically transforms the values of categorical variables in a microdata using a known transition probability matrix. This deliberate misclassification of original response variables introduces uncertainty about the true category of any participant in the survey. On the other hand, since transition (or misclassification) probabilities are known and will be provided to an external legitimate user, valid statistical analyses can still be performed with suitable adjustment of standard methods.

As noted by Van den Hout and Van der Heijden (2002), mathematically, the PRAM is equivalent to a randomized response (RR) design procedure. Both are concerned with protection of respondents' privacy and statistical efficiency. One difference is that in RR surveys, the responder randomizes the response at data gathering stage; whereas, in PRAM, randomization is carried out by the surveyor after the data are collected. Thus, theoretical results developed for RR surveys can be used beneficially in SDC.

This suggests a link between polychotomous RR and PRAM. In fact Warner (1971) suggested that the idea of RR can also be used to protect data that has already been collected. Similar to the argument made for RR above, with this understanding of the privacy protection provided by SDC, the respondents' participation in surveys or censuses is expected to improve.

A variant of PRAM, known as an invariant PRAM, have been proposed to allow data users to perform statistical analysis without the need to perform any adjustments to standard methods. Invariant PRAM is attractive because it may not require a statistical agency to provide misclassification probabilities to an external user of the data.

Several European statistical agencies have investigated the use of PRAM as a SDC method. Notably, PRAM was used in the 2001 Individual Samples of Anonymised Records (SAR) drawn from the UK Census by the Office of National Statistics (ONS) in the UK (Gross, et al 2004). The authors discuss practical implementation of an invariant PRAM with special emphasis on how to apply PRAM while maintaining edit consistencies. Typically at the ONS, the main disclosure control method used is recoding; however, at the point where further recoding causes a large decrease in the information released for little decrease in disclosure risk, they were faced with choice of either removing the remaining high risk records, or alter one or more of their characteristics, in order to protect the records. Thus, ONS developed a method based on the PRAM (Bycroft and Merrett, 2005).

A lot of developmental work on PRAM has been carried out by Statistics Netherlands, but we could not find any documentation where PRAM has been implemented for production in the Netherlands statistical agency. Several papers, however, provide evaluation studies on how well PRAM performs on real survey data. One of such papers evaluates an application of PRAM on the Dutch National Travel Survey (Gouweleeuw, et al, 1998). Statistics Finland has also studied the use of PRAM for the Finnish Lon-

gitudinal Employer-Employee Data (Konnu 2005) while Shlomo and De Waal (2008) describe a PRAM study using data drawn from the 1995 Israel Census sample.

In RR, knowledge of each respondent's identity is not an issue; whereas in PRAM, the respondents' identities are typically protected. We shall elaborate on this difference in a formal manner. Also, aside from identity disclosure, we shall formally discuss statistical disclosure issues. We intend to give practical suggestions and guidelines about when and how to choose PRAM procedures that provide adequate privacy protection and public data releases on which we can perform meaningful statistical analyses. However the choice of the misclassification probabilities to reach a reasonable compromise between respondents' privacy protection and data utility is quite challenging and deserves further investigation.

1.3 Outline of Subsequent Chapters

In Chapter 2, we begin by reviewing in more detail the two early RR procedures mentioned in Section 1.1, and we discuss past unfair comparisons of the two procedures. We also review Nayak's (1994) work on the unification of binary response RR surveys of dichotomous populations. Subsequently, we develop a unified framework for RR surveys of dichotomous populations with polychotomous response variables and clarify certain issues relating to statistical estimation and comparison of RR surveys. We also develop some results relating to unbiased estimation of π and standard errors of estimators under any sampling design $p(s)$.

In Chapter 3, we discuss RR designs for polychotomous populations. Here we extend the theory and framework for RR surveys of dichotomous populations to RR surveys of polychotomous populations. Several of these RR methods have been proposed and investigated in the literature. We shall start with k population categories and k response categories under an infinite population setting as presented in Chaudhuri and Mukerjee

(1988), where they give an overview of relevant works by Bourke and Dalenius (1976), and Liu and Chow (1976). Next, we shall extend our results to situations in finite population settings. We also discuss comparisons of a class of polychotomous RR designs where only one category is sensitive. In particular, we derive conditions for which one can construct better polychotomous RR designs of a specified structure.

In Chapter 4, we start by describing several variations of the PRAM procedure. We demonstrate that, conceptually, any PRAM procedure can be regarded as a PRAMing of a derived variable which is the cross-classification of all the variables. We discuss the connection of PRAM and RR and note that the estimators developed for RR of polychotomous populations in Chapter 3 can be used for PRAM. Next, we discuss a special case of PRAM known as invariant PRAM and introduce the notion of a strongly invariant PRAM. We review methods for constructing invariant PRAM matrices, and we clarify certain perceptions of invariant PRAM that are not fully justified. We also discuss estimation from an invariantly PRAMed data. Finally, we examine the effectiveness of PRAM for limiting statistical disclosure.

Chapter 5 concludes this dissertation. There, we summarize our findings and discuss some questions and possible future research topics. We hope to continue our research and make further contributions to both RR surveys and PRAM.

Chapter 2

Randomized Response Designs for Binary Characteristics

2.1 Introduction

In this chapter we present a unified theory of randomized response (RR) designs for binary characteristics with polychotomous responses variables. We also discuss unbiased estimation under general sampling designs and compare RR surveys, paying attention to both respondents' protection and statistical efficiency. In the next section we review two of the earliest RR procedures, the Warner and Simmons' Unrelated Question methods, respectively. In Section 2.3, we discuss previous unfair comparisons that have been made between the two methods. We review Nayak's (1994) general RR design for binary characteristic with binary response in section 2.4.

In section 2.5, we lay out a unified framework for binary characteristics with polychotomous responses and express privacy measures and some basic statistical entities in terms of the randomization parameters. We hope the proposed framework will be helpful for thinking in a principled way about privacy and statistical efficiency. Most papers present estimators for simple random sample with replacement (SRSWR), but many surveys employ unequal probability sampling, e.g., stratified and multi-stage sampling. In Section 2.6, we discuss unbiased estimation of π and variances of estimators under a general sampling design $p(s)$. Following Padmawar and Vijayan (2000) and Chaudhuri

(2001, 2004), we present a technique for modifying a linear unbiased estimator under an open or direct survey to obtain an unbiased estimator for an RR survey. We investigate RR setups where the randomization process could vary from respondent to respondent. This general view of the application of RR may be useful in stratified SRS designs, in which case, the RR process can vary by strata. We also uncover and discuss an arbitrariness inherent in that approach. In Section 2.7, we compare RR surveys taking both respondents' protection and statistical information into account. We find that use of polychotomous responses is not really helpful for sampling dichotomous populations. Specifically, given any RR procedure with a polychotomous response variable, we can devise a better RR procedure using a dichotomous response variable. Finally in Section 2.8, we discuss admissibility results for binary response RR designs.

2.2 Two Early Procedures

Following the setup in the previous chapter, we begin with the first RR procedure for a binary characteristic, as proposed by Warner (1965). Consider a dichotomous population where each person belongs either to a sensitive group A or to its complement A^c. The objective is to estimate the true proportion (π) of the population that belongs to group A. In Warner's (1965) method, each interviewee first selects one of the two questions:

Q_1: Do you belong to A?

Q_2: Do you belong to A^c?

with respective probabilities p and $(1-p)$, by performing a random experiment, unobserved by the interviewer. The respondent then truthfully replies "Yes" or "No" to the selected question without disclosing the question and thereby protecting his/her privacy. The probabilities p and $(1-p)$ are known and are embedded in the randomization mechanism. Here, p is a design parameter, and not to confuse it with design parameters of other procedures, we shall denote it by p_W.

Consider a simple random sample with replacement (SRSWR). Then, the probability of the "Yes" response is:

$$P_W(\text{Yes}) = \lambda_W = \pi p_W + (1-\pi)(1-p_W) = (1-p_W) + (2p_W - 1)\pi. \tag{2.1}$$

Let n denote the sample size and X denote the number of "Yes" responses. Then, $X \sim b(n, \lambda_W)$. For $p_W \neq 0.5$, Warner (1965) proposed the following method of moments estimator of π :

$$\hat{\pi}_W = \frac{\hat{\lambda}_W - (1-p_W)}{2p_W - 1},$$

where $\hat{\lambda}_W = X/n$. The estimator $\hat{\pi}_W$ is obtained by first writing π in terms of λ_W, using (2.1), and then replacing λ_W with $\hat{\lambda}_W$.

The variance of $\hat{\pi}_W$ is

$$Var(\hat{\pi}_W) = \frac{\lambda_W(1-\lambda_W)}{n(2p_W-1)^2} = \frac{\pi(1-\pi)}{n} + \frac{p_W(1-p_W)}{n(2p_W-1)^2}. \tag{2.2}$$

The last two terms of (2.2) represent, respectively, the variance of the minimum variance unbiased estimator of π from an open (or direct) survey and the additional variance due to randomization. Both the variance of $\hat{\pi}_W$ and the degree of respondents' privacy depend on the value of p_W.

Greenberg et al. (1969) discussed a related procedure, called Simmons' unrelated question method, in which the question Q_2 in Warner's method is replaced by an unrelated nonsensitive question:

Q_3: Do you belong to B?

An example of an unrelated question is: were you born in the month of June? Here, let p_U denote the probability of asking the direct question Q_1, and hence the probability of asking the unrelated question Q_3 is $1 - p_U$. The probabilities of "Yes" (Y) and "No" (N) responses to Q_3 may be known to be, say, β and $(1-\beta)$ respectively. Then, the

probability of responding "Yes" overall is given by:

$$P_{U}(\text{Yes}) = \lambda_{U} = \pi p_{U} + (1 - p_{U})\beta. \tag{2.3}$$

Thus, a method of moments estimator ($\hat{\pi}_{U}$) of π is given as

$$\hat{\pi}_{U} = \frac{\hat{\lambda}_{U} - (1 - p_{U})\beta}{p_{U}}.$$

Greenberg et al. (1969) derived that

$$Var(\hat{\pi}_{U}) = \frac{\lambda_{U}(1 - \lambda_{U})}{n{p_{U}}^{2}},$$

which increases as p_{U} decreases.

2.3 Unfair Comparisons

Greenberg et al. (1969) also illustrated numerically that $Var(\hat{\pi}_{U})$ is smaller than $Var(\hat{\pi}_{W})$ when the probability of asking the direct question Q_1 is the same in the two methods, that is, $p_{U} = p_{W}$. They considered the case where $p_{U} = p_{W} = 0.8$, $\beta = 0.1$ and $n = 1000$. For $\pi = 0.05$, $Var(\hat{\pi}_{W}) = 0.000492$ and $Var(\hat{\pi}_{U}) = 0.000088$; and for $\pi = 0.2$, $Var(\hat{\pi}_{W}) = 0.000604$ and $Var(\hat{\pi}_{U}) = 0.000231$.

As noted by Leysieffer and Warner (1976) and Fligner et al. (1977), that comparison is not fair because the two procedures, with a common value of the probability of asking the direct question Q_1, offer different degrees of privacy to the respondents. Since membership in the sensitive group A is considered a privacy disclosure issue; whereas, membership in A^{c} is not, the larger the conditional probability of belonging to A, given a certain response, the greater the privacy disclosure caused by giving that response. The conditional probabilities of belonging to A under Warner's design can be expressed as

$$P_{W}(A|Yes) = \frac{\pi p_{W}}{1 - p_{W} + (2p_{W} - 1)\pi} \quad \text{and} \quad P_{W}(A|No) = \frac{\pi(1 - p_{W})}{p_{W} - (2p_{W} - 1)\pi}.$$

For Simmons' procedure, the conditional probabilities of belonging to A are

$$P_{U}(A|Yes) = \frac{\pi[p_{U} + (1-p_{U})\beta]}{p_{U}\pi + (1-p_{U})\beta} \quad \text{and} \quad P(A|No) = \frac{\pi(1-p_{U})(1-\beta)}{(1-\pi)p_{U} + (1-p_{U})(1-\beta)}.$$

Using the same numeric example above, we can see that for $\pi = 0.05$, and given a "No" response, the two procedures are about equally protective since $P_{W}(A|No) = 0.01$ and $P_{U}(A|No) = 0.01$. However, $P_{W}(A|Yes) = 0.22$ and $P_{U}(A|Yes) = 0.68$ which suggests that, for this example, Warner's procedure is more protective of respondents in group A given a "Yes" response. Following Lanke (1976), a single measure of degree of privacy is $max\{P(A|Yes), P(A|No)\}$. Thus, we see that Warner's procedure, which has the smaller value of Lanke's measure, is more protective than Simmons' procedure.

For fair comparison, the two procedures should be required to offer equal respondents' protection. Under the assumption that $P(A|Yes) > P(A|No)$, so that $max\{P(A|Yes), P(A|No)\} = P(A|Yes)$, Lanke (1976) derived conditions for Warner's and Simmons' methods to be equally protective for a given π, that is, $P_{W}(A|Yes) = P_{U}(A|Yes)$. He showed that for every p_{U} and β there is a unique value of $p_{W} = \frac{1}{2} + \frac{p_{U}}{2p_{U}+4(1-p_{U})\beta}$ for which $P_{W}(A|Yes) = P_{U}(A|Yes)$.

Several other RR methods have been proposed and investigated in the literature, e.g., Kuk (1990), Mangat and Singh (1990), Mangat (1994) and Kim and Warde (2004); see Chaudhuri and Mukerjee (1988) for a detailed discussion of many of these procedures. However, some of the efficiency comparisons, e.g., Greenberg et al. (1969), Mangat and Singh (1990) and Mangat (1994), are flawed as they do not hold respondents' protection at the same level.

In summary, various RR methods can be found in the literature, but each RR procedure has usually been discussed using features (parameters) that are specific to its randomization mechanism. Often the randomization mechanisms of two procedures share a common element, but with different effect on privacy and efficiency of the two

procedures. For example, the question Q_1 is common to Warner's and Simmons' procedures, but the probability of asking Q_1 affects respondents' protection and statistical efficiency differently for the two procedures; also see Fligner et al. (1977) for additional discussion and numerical illustrations. The unfair comparison of Warner's and Simmons' methods, discussed above, stemmed from considering two procedures with a common randomization parameter, but with disparate impact on respondents' protection, and then comparing variances of the estimators proposed under the two procedures as discussed above. This is a fairly common mistake in RR designs literature, e.g., Greenberg et al. (1969), Mangat and Singh (1990) and Mangat (1994). We believe, misleading comparisons could be avoided by discussing various RR procedures within a common framework. A general framework is also important for identifying and placing the substantive logical issues at the forefront.

2.4 General Binary Response Design

For binary response RR surveys of dichotomous populations, Nayak (1994) proposed a unified framework, which we now briefly discuss. Let Y be an indicator of the sensitive characteristic, viz., $Y = 1$ if the respondent belongs to the sensitive group A and $Y = 0$ otherwise. Let $Z = 0$ and $Z = 1$ label the two response categories. For example, in Warner's and Simmons' procedures, the "Yes" and "No" responses may be recorded as $Z = 1$ and $Z = 0$, respectively. Let, a and b denote $P(Z = 1|Y = 1)$ and $P(Z = 1|Y = 0)$, respectively. Then, the posterior probabilities of $Y = 1$, which determine the level of respondents' privacy, are:

$$P(Y = 1|Z = 1) = \frac{a\pi}{a\pi + b(1-\pi)}$$
$$P(Y = 1|Z = 0) = \frac{(1-a)\pi}{(1-a)\pi + (1-b)(1-\pi)}.$$

Note that these two probabilities depend on the randomization only through a and b.

Let n denote the sample size and X denote the number of respondents reporting $Z = 1$. Then, under the common assumptions of SRSWR and truthful answering, $X \sim b(n, \theta)$, where $\theta = a\pi + b(1 - \pi)$. Since $\frac{X}{n}$ is the uniformly minimum variance unbiased estimator (UMVUE) of θ, if $a \neq b$, the UMVUE of π, based on X, is given by

$$\hat{\pi} = (\frac{X}{n} - b)/(a - b) \tag{2.4}$$

and its variance is

$$\begin{aligned} V(\hat{\pi}) &= V(\frac{X}{n} - b)/(a - b)^2 \\ &= \frac{V(X)}{[n(a - b)]^2}. \end{aligned}$$

Thus, $V(\hat{\pi}) = \theta(1 - \theta)/[n(a - b)^2]$.

If a and b are known and are determined only by the randomization mechanism, as is the case for most binary response RR procedures, all statistical properties, including protection of privacy and accuracy of statistical inferences, depend on the randomization step only through the values of a and b. So, such procedures can be characterized by a and b, taking them as the RR design parameters. Thus, a unified approach ensues from discussing various binary response RR procedures in terms of their design parameters a and b.

Remark 2.1. Any one-to-one transformation of (a, b) can also be used as the RR design parameters for developing a unified framework. In particular, Nayak (1994) used $P(Yes|A)$ and $P(No|A^c)$ as the RR design parameters, which correspond to our a and $1 - b$ if $Z = 1$ and $Z = 0$ represent the "Yes" and "No" responses, respectively. Leysieffer and Warner (1976) expressed respondents' protection and $Var(\hat{\pi})$ in terms of $u = a/b$ and $v = (1 - b)/(1 - a)$. As the transformation $\{a, b\} \to \{u, v\}$ is one-to-one, a unified

framework can also be developed in terms of u and v.

Remark 2.2. As it was noted in Nayak (1994), the interchanging of the two responses "Yes" and "No" (or equivalently $Z = 1$ and $Z = 0$) does not alter any statistical property of a procedure, which implies that for any $0 \leq a, b \leq 1$, the two RR procedures with RR design parameters (a, b) and $(1 - a, 1 - b)$, respectively, are equivalent. For unique representation, we may impose the restriction $a > b$ and take $\{(a, b) : 0 \leq a, b \leq 1, a > b\}$ as the RR design space. In this framework, Nayak (1994) showed that respondents' protection and statistical efficiency do not necessarily move in opposite directions and an RR design (a, b) is admissible if and only if $a = 1$. We shall present a stronger result in Section 2.8 below.

Remark 2.3. The general framework presented above covers all binary response RR procedure for which the randomization probabilities a and b are known. It does not cover Simmons' two sample procedure, where the probability of the "Yes" answer to Q_3 is unknown and the sampled individuals are divided into two groups to receive the questions Q_1 and Q_3 with different but known probabilities (Greenberg et al., 1969).

2.5 A Unified Framework

Some RR procedures proposed in the literature use polychotomous responses (e.g., Leysieffer and Warner, 1976; Kuk, 1990; Christofides, 2003). In Chow's procedure, discussed in Leysieffer and Warner (1976), each respondent selects k balls at random, without replacement and unobserved by the interviewer from an urn containing L red and M blue balls, where L and M are known and $k \leq \min\{L, M\}$. The respondent then reports the number of red balls if he/she belongs to the sensitive group A; otherwise he/she reports the number of blue balls. Christofides (2003) proposed a similar, albeit more general, procedure where each person in the sample is provided with a device which produces the integers $1, \cdots, k$ with known probabilities $p_1, \cdots, p_k$, respectively. Each respondent

uses the device, in the absence of the interviewer, to produce one integer J and then reports the value of $(k+1-J)$ if he/she belongs to A; otherwise, he/she reports the value of J. In Kuk's (1990) repeated trials design, each respondent is given two decks of cards. Both decks comprise of cards of two colors, say red and blue, but with different proportions. A respondent selects k cards at random and with replacement from deck 1(2) if he/she belongs to $A(A^c)$ and reports the number of red cards selected. In all of these procedures, the response variable Z is integer valued and the probabilities $\{P(Z=z|A)\}$ and $\{P(Z=z|A^c)\}$ are known and specified by the randomization device.

Here we shall present a unified framework for RR surveys of dichotomous populations using polychotomous response variables. As before, let $Y=1$ if the respondent belongs to the sensitive group A and $Y=0$ if the respondent belongs to A^c and let $\pi=P(Y=1)$ be the unknown parameter of interest. We shall denote the response variable by Z, the number of response categories by $k(k\geq 2)$ and the possible responses by $c_1,\cdots,c_k$, satisfying $c_i\neq c_j$ for $i\neq j$. Further, let $\alpha_i=P(Z=c_i|Y=1), \beta_i=P(Z=c_i|Y=0)$ for $i=1,\cdots,k$, $\boldsymbol{\alpha}=(\alpha_1,...,\alpha_k)$ and $\boldsymbol{\beta}=(\beta_1,...,\beta_k)$. For uniqueness of k, we shall require that $\min(\alpha_i,\beta_i)>0, i=1,\cdots,k$. We shall call a RR survey with k response categories a $(2\to k)$ RR survey. The binary response surveys, as considered in Nayak (1994), correspond to $k=2$. We shall consider all $(2\to k)$ RR surveys with known $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, noting that for protecting respondent's privacy it is not necessary to use a randomization device for which $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are unknown.

Let θ_i denote $P(Z=c_i)$, i.e., $\theta_i=\alpha_i\pi+\beta_i(1-\pi)$. Then, the posterior probabilities which determine the level of respondents' privacy are:

$$P(A|Z=c_i)=\frac{\alpha_i\pi}{\alpha_i\pi+\beta_i(1-\pi)}=\frac{\pi}{\pi+(\beta_i/\alpha_i)(1-\pi)},\quad i=1,\cdots,k. \tag{2.5}$$

For $i=1,\cdots,k$, let X_i denote the observed frequency of the response c_i. Then, under

SRSWR, $(X_1, \cdots, X_k) \sim mult(n; \theta_1, \cdots, \theta_k)$ with probability mass function

$$f_\pi(x_1, \cdots, x_k) = P(X_1 = x_1, \cdots, X_k = x_k) = \frac{n!}{x_1! \cdots x_k!} \theta_1^{x_1} \cdots \theta_k^{x_k}.$$

The log likelihood is given by

$$l(\theta) = constant + \sum x_i ln[\alpha_i \pi + \beta_i (1 - \pi)]$$

Hence, the maximum likelihood estimate (MLE) of π, based on $(X_1, \cdots, X_k)$, is the solution of

$$\frac{\partial}{\partial \pi} \ln f_\pi(x_1, \cdots, x_k) = 0 \quad \text{or} \quad \sum_{i=1}^{k} \frac{x_i(\alpha_i - \beta_i)}{\alpha_i \pi + \beta_i (1 - \pi)} = 0,$$

provided that the solution is in $[0, 1]$. In the case of sampling from an infinite population, the Fisher information in a single response is

$$i(\pi) = \sum_{i=1}^{k} \frac{(\alpha_i - \beta_i)^2}{\alpha_i \pi + \beta_i (1 - \pi)}$$

and the asymptotic distribution of the MLE $(\hat{\pi}_{ML})$ is normal:

$$\sqrt{n}(\hat{\pi}_{ML} - \pi) \xrightarrow{L} N(0, i^{-1}(\pi)) \quad \text{as} \quad n \to \infty,$$

which can be used to construct large sample confidence intervals for π.

Note that the posterior probabilities in (2.5), the distribution of $(X_1, \cdots, X_k)$ and the Fisher's information depend on the randomization mechanism only through $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. Thus, all $(2 \to k)$ RR procedures can be characterized by the values of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. This implies that for designing a $(2 \to k)$ RR survey we should first determine the values of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ and then devise a mechanism for implementing them. For a unified approach, we suggest to take $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ as the RR design parameters and discuss and examine all $(2 \to k)$ RR procedures through them.

Remark 2.4. The ordering of the k response categories should have no bearing on the substantive properties of a $(2 \to k)$ RR design. A $(2 \to k)$ RR design essentially remains unchanged under any permutation of the response categories and corresponding permutations of the components of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. Thus, for any $(\boldsymbol{\alpha}, \boldsymbol{\beta}) = (\alpha_1, \cdots, \alpha_k, \beta_1, \cdots, \beta_k)$ and any permutation $(i_1, \cdots i_k)$ of $(1, \cdots, k)$, the two $(2 \to k)$ RR designs with randomization probabilities $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ and $(\alpha_{i_1}, \cdots, \alpha_{i_k}, \beta_{i_1}, \cdots, \beta_{i_k})$, respectively, are equivalent. This implies that any $(2 \to k)$ RR design can be characterized by many different sets of values of the RR design parameters $(\boldsymbol{\alpha}, \boldsymbol{\beta})$, generated by permutations of the response categories. So, for unique characterization of a $(2 \to k)$ RR design by $(\boldsymbol{\alpha}, \boldsymbol{\beta})$, we need a convention for ordering the response categories. One possibility is to order the categories first by the values of $\{\alpha_i\}$ and then by the β values, i.e., require that $\alpha_1 \geq \cdots \geq \alpha_k$ and if $\alpha_j = \cdots = \alpha_{j+m}$ for any j and m, then $\beta_j \geq \cdots \geq \beta_{j+m}$. Noting that the posterior probability $P(A|Z = c_j)$ in (2.5) depends on $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ only through the ratio (α_i/β_i) and $P(A|Z = c_j)$ is an increasing function of (α_i/β_i), we believe a more meaningful approach would be to order the categories first in decreasing order of magnitude of (α_i/β_i), i.e., in decreasing order of the probability of being classified in the sensitive group A, and then by decreasing order of magnitude of $\{\alpha_i\}$. Thus, to make $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ unique we suggest to require $(\alpha_1/\beta_1) \geq \cdots \geq (\alpha_k/\beta_k)$ and if $(\alpha_j/\beta_j) = \cdots = (\alpha_{j+m}/\beta_{j+m})$ for any j and m, then $\alpha_j \geq \cdots \geq \alpha_{j+m}$.

2.6 Estimation of π

While most authors discussed statistical analyses of RR data assuming random sampling from an infinite population or SRSWR, practical surveys often involve complex survey designs and many variables, only a few of which may be sensitive. Thus, it is important to derive estimators based on RR data and unequal probability sampling. For a quantitative response variable, Padmawar and Vijayan (2000) discussed linear unbiased

estimation of a finite population total based on RR data obtained under a general sampling design. Analogously, Chaudhuri (2001, 2004) presented linear unbiased estimators of a population proportion (π) based on certain RR procedure including the Christofides' (2003) procedure. Specifically, they showed how a linear unbiased estimator based on an open survey can be modified to obtain an unbiased estimator under an RR survey. They also discussed unbiased estimation of the variance of the estimators from RR survey data. In this section, we first develop similar results for a general $(2 \to k)$ RR procedure and then discuss an arbitrariness of the approach.

Consider a finite population of N units, labeled $i = 1, \cdots, N$, and let Y_i denote the value of Y (an indicator of the sensitive variable) for unit i. Let Z_i denote the response of unit i and suppose the sample is selected using a non-informative sampling design $p(s)$. So, the data can be represented as $\{(i, Z_i); i \in s\}$, where s is a subset of $\{1, \cdots, N\}$, and our goal is to estimate $\pi = (\sum_{i=1}^{N} Y_i)/N$. Note that while Y_i are fixed, Z_i are random variables, and estimation of π is equivalent to estimation of the population total $T(Y) = \sum_{i=1}^{N} Y_i$.

Suppose

$$e(s, y) = w_{s0} + \sum_{i \in s} w_{si} Y_i \tag{2.6}$$

is a linear unbiased estimator of $T(Y)$, i.e.,

$$E_p[e(s, y)] = \sum_{s} e(s, y) p(s) = \sum_{i=1}^{N} Y_i \quad \text{for all } Y_1, \cdots, Y_N$$

that is,

$$\sum_{i=1}^{N} Y_i = \sum_{s} (w_{s0} + \sum_{i \in s} w_{si} Y_i) p(s)$$

$$= \sum_s w_{s0} p(s) + \sum_s \sum_{i \in s} w_{si} Y_i p(s)$$
$$= \sum_s w_{s0} p(s) + \sum_{i=1}^{N} Y_i \sum_{s \ni i} w_{si} p(s).$$

Thus,

$$\sum_s w_{s0} p(s) = 0 \quad \text{and} \quad \sum_{s \ni i} w_{si} p(s) = 1, \quad i = 1, \cdots, N.$$

To extend Chaudhuri's (2001, 2004) results, in this section we shall require $c_1, \cdots, c_k$ to be real numbers. Then, since $P(Z = c_j | Y = 1) = \alpha_j$ and $P(Z = c_j | Y = 0) = \beta_j$, we have

$$E[Z|Y = 1] = \sum_{j=1}^{k} \alpha_j c_j \quad \text{and} \quad E[Z|Y = 0] = \sum_{j=1}^{k} \beta_j c_j$$

or

$$E[Z|Y] = \sum_{j=1}^{k} \beta_j c_j + [\sum_{j=1}^{k} (\alpha_j - \beta_j) c_j] Y = d_1 + d_2 Y, \quad \text{say,}$$

where $d_1 = \sum_{j=1}^{k} \beta_j c_j$ and $d_2 = \sum_{j=1}^{k} (\alpha_j - \beta_j) c_j$. So, if $d_2 \neq 0$, i.e., $(\boldsymbol{\alpha} - \boldsymbol{\beta})$ is not orthogonal to $\boldsymbol{c} = (c_1, \cdots, c_k)$, letting $U = (Z - d_1)/d_2$, it follows that $E_R(U) = E(U|Y) = Y$, where E_R denotes expectation with respect to the randomization mechanism. Let

$$e^*(s, z) = w^*_{s0} + \sum_{i \in s} w^*_{si} Z_i, \tag{2.7}$$

where $w^*_{s0} = w_{s0} - (d_1/d_2) \sum_{i \in s} w_{si}$ and $w^*_{si} = w_{si}/d_2$. Then,

$$E[e^*(s, z)] = E_p E_R [w^*_{s0} + \sum_{i \in s} w^*_{si} Z_i]$$
$$= E_p E_R [w_{s0} - (d_1/d_2) \sum_{i \in s} w_{si} + \sum_{i \in s} (w_{si}/d_2) Z_i]$$

$$
\begin{aligned}
&= E_p E_R[w_{s0} + \sum_{i \in s} (w_{si}/d_2)(Z_i - d_1)] \\
&= E_p E_R[w_{s0} + \sum_{i \in s} w_{si} U_i] \\
&= E_p[w_{s0} + \sum_{i \in s} w_{si} Y_i] \\
&= E_p[e(s, y)] = T(Y).
\end{aligned}
$$

Thus, we have the following:

Theorem 2.1. *For any given sampling design $p(s)$, if $e(s, y)$ in (2.6) is a linear design unbiased estimator of the population total $T(Y) = \sum_{i=1}^{N} Y_i$ based on the open survey, then the estimator $e^*(s, z)$ in (2.7) is a linear design unbiased estimator of $T(Y)$ based on the RR survey with RR design $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ and sampling design $p(s)$.*

Conversely, from any given linear unbiased estimator for an RR survey we can derive a linear unbiased estimator for an open survey. Specifically, suppose $e^*(s, z) = b_{s0} + \sum_{i \in s} b_{si} Z_i$ is an unbiased estimator of $T(Y)$ based on an RR survey, i.e.,

$$
Ee^*(s, z) = T(Y) = E_p E_R[b_{s0} + \sum_{i \in s} b_{si} Z_i] = E_p[b^*_{s0} + \sum_{i \in s} b^*_{si} Y_i], \tag{2.8}
$$

where $b^*_{s0} = b_{s0} + d_1 \sum_{i \in s} b_{si}$ and $b^*_{si} = d_2 b_{si}$. Then, (2.8) shows that $e(s, y) = b^*_{s0} + \sum_{i \in s} b^*_{si} Y_i$ is a linear unbiased estimator of $T(Y)$ based on the corresponding open survey. Thus, for a given sampling design $p(s)$, there is a one-to-one relationship between the two classes of all linear unbiased estimators of π based on an open survey and an RR survey, respectively.

2.6.1 Variance Estimation

We shall now focus on the variance of the estimator $e^*(s,z)$, defined in (2.7). First note that

$$V[Z|Y=1] = \sum_{j=1}^{k} \alpha_j c_j^2 - (\sum_{j=1}^{k} \alpha_j c_j)^2 = v_1 \quad \text{and}$$
$$V[Z|Y=0] = \sum_{j=1}^{k} \beta_j c_j^2 - (\sum_{j=1}^{k} \beta_j c_j)^2 = v_0$$

and hence we can write

$$V[Z|Y] = v_0 + (v_1 - v_0)Y. \tag{2.9}$$

Using (2.9), the variance of $e^*(s,z)$ can be written as

$$\begin{aligned} V(e^*(s,z)) &= E_p V_R(e^*(s,z)|s,Y) + V_p E_R(e^*(s,z)|s,Y) \\ &= E_p[\sum_{i\in s} \frac{w_{si}^2}{d_2^2}(v_0 + (v_1 - v_0)Y_i)] + V_p(e(s,y))]. \end{aligned} \tag{2.10}$$

The first term in (2.10) is the extra variation due to randomization. Noting that in our application $Y_i^2 = Y_i$,

$$\begin{aligned} V_p(e(s,y)) &= E_p\{e(s,y)\}^2 - \{E_p(e(s,y))\}^2 \\ &= E_p\{w_{s0} + \sum_{i\epsilon s} w_{si}Y_i\}^2 - \{E_p[w_{s0} + \sum_{i\epsilon s} w_{si}Y_i]\}^2 \\ &= E_p\{w_{s0}^2 + 2\sum_{i\epsilon s} w_{s0}w_{si}Y_i + (\sum_{i\epsilon s} w_{si}^2 Y_i + \sum_{\substack{i,j\epsilon s \\ i\neq j}} w_{si}w_{sj}Y_iY_j)\} \\ &\quad - \{E_p[w_{s0} + \sum_{i\epsilon s} w_{si}Y_i]\}^2 \end{aligned}$$

$$
\begin{aligned}
&= \sum_s w_{s0}^2 p(s) + \sum_{i=1}^N Y_i \sum_{s\ni i} (w_{si}^2 + 2w_{s0}w_{si})p(s) \\
&+ \sum_{\substack{i,j=1 \\ i\neq j}}^N Y_i Y_j \sum_{s\ni i,j} w_{si}w_{sj}p(s) - (\sum_{i=1}^N Y_i)^2 \\
&= \sum_s w_{s0}^2 p(s) + \sum_{i=1}^N Y_i \sum_{s\ni i} (w_{si}^2 + 2w_{s0}w_{si})p(s) \\
&+ \sum_{\substack{i,j=1 \\ i\neq j}}^N Y_i Y_j \sum_{s\ni i,j} w_{si}w_{sj}p(s) - (\sum_{i=1}^N Y_i + \sum_{\substack{i,j=1 \\ i\neq j}}^N Y_i Y_j) \\
&= g_0 + \sum_{i=1}^N g_i Y_i + \sum_{\substack{i,j=1 \\ i\neq j}}^N g_{ij} Y_i Y_j, \quad \text{say,}
\end{aligned}
$$

where,

$$
g_0 = \sum_s w_{s0}^2 p(s), \quad g_i = \sum_{s\ni i} (w_{si}^2 + 2w_{s0}w_{si})p(s) - 1, \quad \text{and} \quad g_{ij} = \sum_{s\ni i,j} w_{si}w_{sj}p(s) - 1.
$$

Now we discuss how an unbiased estimator of $V(e^*(s,z)$ can be obtained from an unbiased estimator of $V_p(e(s,y)$ based on an open survey. Let d_{si} and d_{sij} be such that

$$
\sum_{s\ni i} d_{si}p(s) = g_i \quad \text{and} \quad \sum_{s\ni i,j} d_{sij}p(s) = g_{ij},
$$

so that

$$
t(s,y) = g_0 + \sum_{i\in s} d_{si}Y_i + \sum_{\substack{i,j\in s \\ i\neq j}} d_{sij}Y_iY_j \tag{2.11}
$$

is an unbiased estimator of $V_p(e(s,y))$ based on the open survey data. A specific unbiased estimator of $V_p(e(s,y))$ is obtained by using $d_{si} = g_i/\pi_i$ and $d_{sij} = g_{ij}/\pi_{ij}$, where $\pi_i = \sum_{s\ni i} p(s)$ and $\pi_{ij} = \sum_{s\ni i,j} p(s)$. From (2.11) and (2.10) the following result can now be established using the fact that $E_R(U|Y) = Y$.

Theorem 2.2. *An unbiased estimator of $V(e^*(s,z))$, based on RR survey data, is given by*

$$t^*(s,z) = g_0 + \sum_{i\in s} d_{si}U_i + \sum_{\substack{i,j\in s \\ i\neq j}} d_{sij}U_iU_j + \sum_{i\in s} \frac{w_{si}^2}{d_2^2}(v_0 + (v_1 - v_0)U_i).$$

Remark 2.5. So far we have assumed that the randomization probabilities $\{\alpha_j, \beta_j\}$ are the same for all population units. However, in some situations, especially in stratified sampling, as discussed in Kim and Warde (2004) and Christofides (2005), the randomization probabilities may vary over the populations units. We note that the above discussed technique for deriving unbiased estimators based on a RR survey from unbiased estimators based on an open survey also works for varying randomization probabilities. Suppose the randomization probabilities for unit i are $(\boldsymbol{\alpha}_i, \boldsymbol{\beta}_i) = (\alpha_{i1}, \cdots, \alpha_{ik}, \beta_{i1}, \cdots, \beta_{ik}), i = 1, \cdots, N$. Then letting $d_{i1} = \sum_{j=1}^k \beta_{ij}c_j, d_{i2} = \sum_{j=1}^k (\alpha_{ij} - \beta_{ij})c_j$ and $U_i = (Z_i - d_{i1})/d_{i2}$, it can be seen that $e^*(s,z) = w_{s0} + \sum_{i\in s} w_{si}U_i$ is an unbiased estimator of $T(Y)$. Furthermore, letting

$$v_{i0} = V[Z_i|Y_i = 0] = \sum_{j=1}^k \beta_{ij}c_j^2 - (\sum_{j=1}^k \beta_{ij}c_j)^2 \quad \text{and}$$
$$v_{i1} = V[Z_i|Y_i = 1] = \sum_{j=1}^k \alpha_{ij}c_j^2 - (\sum_{j=1}^k \alpha_{ij}c_j)^2,$$

we can verify that the variance of $e^*(s,z)$ is

$$V(e^*(s,z)) = E_p[\sum_{i\in s} \frac{w_{si}^2}{d_{i2}^2}(v_{i0} + (v_{i1} - v_{i0})Y_i)] + V_p(e(s,y)). \tag{2.12}$$

and an unbiased estimator of (2.12) is

$$t^*(s,z) = g_0 + \sum_{i\in s} d_{si}U_i + \sum_{\substack{i,j\in s \\ i\neq j}} d_{sij}U_iU_j + \sum_{i\in s} \frac{w_{si}^2}{d_{i2}^2}(v_{i0} + (v_{i1} - v_{i0})U_i).$$

We may also mention that following Chaudhuri (2001, 2004), one can express $V(e^*(s,z))$

in other forms and thence obtain other unbiased estimators of it.

2.6.2 An Arbitrariness of $e^*(s,z)$.

We note that the construction of $e^*(s,z)$ from a given open survey estimator $e(s,y)$, discussed above, depends on the numerical values $c_1,\cdots,c_k$ used to label the response categories of the RR procedure. For a given sampling design $p(s)$ and a given unbiased estimator for the open survey, the technique yields many unbiased estimators for an RR survey (with $k \geq 3$), by associating different sets of numbers to the response categories. To put this in another way, let $c_i^* = \psi(c_i), i = 1,\cdots,k$ be a transformation of $\{c_1,\cdots,c_k\}$. Then it can be seen that $E[\psi(Z)|Y] = d_1^* + d_2^* Y$, where $d_1^* = \sum_{j=1}^k \beta_j \psi(c_j)$ and $d_2^* = \sum_{j=1}^k (\alpha_j - \beta_j)\psi(c_j)$ and if $d_2^* \neq 0$, letting $U^* = (\psi(Z) - d_1^*)/d_2^*$, it follows that $E_R(U^*|Y) = Y$, and

$$e^{**}(s,z) = w_{s0} + \sum_{i\in s} w_{si} U_i^* = w_{s0}^{**} + \sum_{i\in s} w_{si}^{**} \psi(Z_i),$$

where $w_{s0}^{**} = w_{s0} - (d_1^*/d_2^*)\sum_{i\in s} w_{si}$ and $w_{si}^{**} = w_{si}/d_2^*$, is an unbiased estimator of $T(Y)$ based on the RR survey. Thus, from a given $e(s,y)$, we can construct many unbiased estimators of $T(Y)$ based on the RR survey, by employing different transformations $\psi(.)$.

In view of the above discussion, we may choose $c_1,\cdots,c_k$ to minimize the variance in (2.10). Since the second term of (2.10) does not depend on $c_1,\cdots,c_k$ we need to consider only the first term. First, we note that $(Z - d_1)/d_2$ and $e^*(s,y)$ are invariant under location and scale transformations of $c_1,\cdots,c_k$, i.e., under $c_i \to \gamma c_i + \delta, i = 1,\cdots,k$, for all γ, δ. This follows from the fact that

$$(Z - d_1)/d_2 \to [\gamma Z + \delta - \sum_{j=1}^k \beta_j(\gamma c_j + \delta)] / \sum_{j=1}^k (\alpha_j - \beta_j)(\gamma c_j + \delta) \tag{2.13}$$

$$
\begin{aligned}
&= [\gamma Z + \delta - \gamma \sum_{j=1}^{k} \beta_j c_j - \delta \sum_{j=1}^{k} \beta_j] / [\gamma \sum_{j=1}^{k} (\alpha_j - \beta_j) c_j + \delta \sum_{j=1}^{k} (\alpha_j - \beta_j)] \\
&= [\gamma Z + \delta - \gamma \sum_{j=1}^{k} \beta_j c_j - \delta] / [\gamma \sum_{j=1}^{k} (\alpha_j - \beta_j) c_j + \delta . 0] \\
&= [\gamma Z - \gamma \sum_{j=1}^{k} \beta_j c_j] / [\gamma \sum_{j=1}^{k} (\alpha_j - \beta_j) c_j] \\
&= \gamma [Z - \sum_{j=1}^{k} \beta_j c_j] / \gamma [\sum_{j=1}^{k} (\alpha_j - \beta_j) c_j] \\
&= [Z - \sum_{j=1}^{k} \beta_j c_j] / [\sum_{j=1}^{k} (\alpha_j - \beta_j) c_j] \\
&= (Z - d_1)/d_2,
\end{aligned}
$$

and, from (2.7),

$$
\begin{aligned}
e^*(s,z) &= w_{s0} - (d_1/d_2) \sum_{i \in s} w_{si} + \sum_{i \in s} (w_{si}/d_2) Z_i \\
&= w_{s0} + \sum_{i \in s} w_{si} \left(\frac{Z_i - d_1}{d_2} \right) \\
&\to w_{s0} + \sum_{i \in s} w_{si} \left(\frac{Z_i - d_1}{d_2} \right) = e^*(s,z)
\end{aligned}
$$

using (2.13). From this, it can also be seen that for $k = 2$, i.e., for a $(2 \to 2)$ RR design, $e^*(s,z)$ is unique, independent of the choice of c_1 and c_2. So the estimation methods discussed earlier is well defined for $(2 \to 2)$ RR designs. For $k \geq 3$, without loss of generality (in view of the above mentioned invariance under location and scale transformations), we impose the restrictions:

$$
\sum_{i=1}^{k} \alpha_i c_i = 1 \quad \text{and} \quad \sum_{i=1}^{k} \beta_i c_i = 0. \tag{2.14}
$$

Then, $d_1 = 0, d_2 = 1, v_0 = \sum_{i=1}^{k} c_i^2 \beta_i$ and $v_1 = \sum_{i=1}^{k} c_i^2 \alpha_i - 1$, and the first term of (2.10) reduces to

$$E_p[\sum_{i \in s} \frac{w_{si}^2}{d_2^2}(v_0 + (v_1 - v_0)Y_i)] = v_0 \sum_s \{\sum_{i \in s} w_{si}^2\} p(s) + (v_1 - v_0) \sum_{i=1}^{N} Y_i \sum_{s \ni i} w_{si}^2 p(s)$$
$$= A_1(\sum_{i=1}^{k} c_i^2 \alpha_i - 1) + A_2(\sum_{i=1}^{k} c_i^2 \beta_i), \tag{2.15}$$

where

$$A_1 = \sum_{i=1}^{N} Y_i \sum_{s \ni i} w_{si}^2 p(s) \quad \text{and} \quad A_2 = \sum_s \{\sum_{i \in s} w_{si}^2\} p(s) - A_1.$$

Now, using Lagrangian multipliers, it can be seen that (2.15) is minimized, subject to (2.14), by

$$c_i = \frac{D_2 \alpha_i - D_3 \beta_i}{(D_1 D_2 - D_3^2)(A_1 \alpha_i + A_2 \beta_i)}, \quad i = 1, \cdots, k, \tag{2.16}$$

where,

$$D_1 = \sum_{i=1}^{k} \frac{\alpha_i^2}{A_1 \alpha_i + A_2 \beta_i}, \quad D_2 = \sum_{i=1}^{k} \frac{\beta_i^2}{A_1 \alpha_i + A_2 \beta_i} \quad \text{and} \quad D_3 = \sum_{i=1}^{k} \frac{\alpha_i \beta_i}{A_1 \alpha_i + A_2 \beta_i}.$$

The optimum values in (2.16) depend on the sampling design $p(s)$, the estimator $e(s, y)$ for the open survey and also on $Y_1, \cdots, Y_N$, which are unknown. However, the $\{c_i\}$ in (2.16) depend on $Y_1, \cdots, Y_N$ only through A_1, which may be approximated by

$$A_1 \approx \pi \sum_{i=1}^{N} \sum_{s \ni i} w_{si}^2 p(s).$$

2.7 Comparison of RR Procedures

For comparing two RR procedures one should examine both statistical information and respondents' privacy offered by the two procedures. Several authors, including Leysieffer and Warner (1976), Lanke (1976) and Fligner et al. (1977), suggested to compare statistical efficiency, e.g., variances of the estimators, of competing procedures while

requiring them to offer the same degree of respondents' protection. We shall adopt this approach. The variance of an estimator depends not only on the RR design but also on the choice of the estimator. So for comparing statistical efficiency of two designs, it may be more appropriate to compare some measure of "statistical information" afforded by the two designs. In the following, we shall use Fisher information to compare statistical efficiencies of two RR designs.

We now suggest a criterion for controlling respondents' protection. We start our deliberation with the posterior probabilities in (2.5), which are the determinants of respondents' privacy. The response c_j alters the probability of the respondent's belonging to A by the factor $r_j = P(A|Z = c_j)/\pi$. The ratio r_j may be taken as a measure of respondents' hazard yielded from reporting the response c_j. Clearly, the respondents' hazards $r_1, \cdots, r_k$ corresponding to the responses $c_1, \cdots, c_k$ may be different and a response c_j is hazardous only if $r_j > 1$. Logically, a $(2 \to k)$ RR design is totally non-hazardous if all posterior probabilities equal the prior probability (π), i.e., $r_1 = \cdots = r_k = 1$. However, it can be seen that $r_1 = \cdots = r_k = 1$ if and only if $\alpha_j = \beta_j$ for $j = 1, \cdots, k$, in which case, $\theta_j = P(Z = c_j), j = 1, \cdots, k$, are independent of π and hence the data do not contain any information on π. Thus, to be statistically useful, the design cannot be totally non-hazardous and some r_j must be greater than one. Let $(\alpha_{i1}, \cdots, \alpha_{ik}, \beta_{i1}, \cdots, \beta_{ik}), i = 1, 2$ be two $(2 \to k)$ RR designs with respondents' hazards $(r_{i1}, \cdots, r_{ik}), i = 1, 2$. Strictly speaking, these two RR designs offer equal respondents' protection if and only if $r_{1j} = r_{2j}, j = 1, \cdots, k$. However, this can be satisfied if and only if $\alpha_{1j} = \alpha_{2j}$ and $\beta_{1j} = \beta_{2j}$ for $j = 1, \cdots, k$ (assuming that the design parameters are specified uniquely following a convention, as discussed in Remark 2.4), i.e., the two designs are the same. Thus, to proceed further we need to employ a weaker criterion for defining equal respondents' protection.

It seems sensible to work with the maximum respondents' hazard:

$MRH = \max\{P(A|Z = c_1)/\pi, \cdots, P(A|Z = c_k)/\pi\} = \max\{r_1, \cdots, r_k\}$. This MRH measure is similar to the "primary protection" measure of Fligner et al. (1977). Two RR designs will be considered to offer equal respondents' protection if they have the same MRH value. For controlling respondents' protection, it seems sensible to require the MRH value to be less than a pre-specified number. Several authors, e.g., Anderson (1976), Lanke (1976), Leysieffer and Warner (1976) and Fligner et al. (1977), essentially suggested this approach. However, the values of r_j and hence MRH depend on the unknown parameter π. Thus, in practice, we would need to put an upper bound on MRH for a specific value of π. Since $P(A|Z = c_j)/\pi$ is an increasing function of (α_j/β_j), putting an upper bound on MRH, for a specific value of π, is equivalent to putting an upper bound on $\max\{\alpha_1/\beta_1, \cdots, \alpha_k/\beta_k\}$. Thus, we shall take

$$R(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \max\{\frac{\alpha_1}{\beta_1}, \cdots, \frac{\alpha_k}{\beta_k}\} \tag{2.17}$$

as our measure of the degree of privacy afforded by the $(2 \to k)$ RR design $(\vec{\alpha}, \vec{\beta})$. It can be seen that two RR designs $(\vec{\alpha}_1, \vec{\beta}_1)$ and $(\vec{\alpha}_2, \vec{\beta}_2)$ have a common value of MRH if and only if they have same value for R, i.e., $R(\boldsymbol{\alpha}_1, \boldsymbol{\beta}_1) = R(\boldsymbol{\alpha}_2, \boldsymbol{\beta}_2)$. Noting that

$$\frac{\alpha_j}{\beta_j} = \frac{P(Z = c_j|A)}{P(Z = c_j|A^c)} = \frac{P(A|Z = c_j)/P(A)}{P(A^c|Z = c_j)/P(A^c)}, \quad j = 1, \cdots, k,$$

the ratios $\{\alpha_j/\beta_j\}$ may be regarded as Bayes factors. They were used by Leysieffer and Warner (1976) in their discussion of respondents' protection.

In summary, for comparing two procedures, we suggest to hold the privacy measure in (2.17) equal for the two procedures and then compare statistical efficiency, measured by Fisher information. Using this approach, we shall next show that for any $(2 \to k)$ design with $k \geq 3$, there exists a better $(2 \to 2)$ design. The result also helps us to identify the most efficient RR design at any given level of privacy protection.

Theorem 2.3. *Let D be any $(2 \to k)$ RR design with $k \geq 3$ and randomization parameters $(\boldsymbol{\alpha}, \boldsymbol{\beta})$. Then, there exists a $(2 \to 2)$ RR design D_0 which provides the same respondents' protection, measured by (2.17), as D and at least as much statistical information as D.*

Proof. For notational simplicity, without loss of generality suppose that $\alpha_1/\beta_1 = \max\{\alpha_j/\beta_j\}$. If $(\alpha_1/\beta_1) = 1$, the data do not contain any information on π. So, we shall only consider the case of $(\alpha_1/\beta_1) > 1$. Let $b_0 = \beta_1/\alpha_1 (< 1)$ and D_0 be the $(2 \to 2)$ RR design with response variable V and $P(V = 1|A) = 1$ and $P(V = 1|A^c) = b_0$. It is easy to verify that D_0 and D offer the same degree of respondents' protection, as measured by (2.17).

Next, we shall show that D is equivalent to post-randomizing the data generated by D_0. Let $\boldsymbol{\gamma} = (\gamma_1, \cdots, \gamma_k) = (\boldsymbol{\beta} - b_0\boldsymbol{\alpha})/(1 - b_0)$, and randomly transform V to $c_1, \cdots, c_k$ according to the probabilities $P(c_j|V = 1) = \alpha_j$ and $P(c_j|V = 0) = \gamma_j$ for $j = 1, \cdots, k$, and denote the resulting variable by Z. From the fact that $\alpha_1/\beta_1 \geq \alpha_i/\beta_i, i = 1, \cdots, k$, it can be checked easily that $\gamma_i \geq 0, i = 1, \cdots, k$ and $\sum_{i=1}^{k} \gamma_i = 1$, i.e., $\vec{\gamma}$ is a probability vector. We may also note that the transformation of V to Z is performed without using the true category of the respondent or the true value of π. Now, it follows easily that for $i = 1, \cdots, k$,

$$P(Z = c_i|A) = P(Z = c_i|V = 1)P(V = 1|A) + P(Z = c_i|V = 0)P(V = 0|A) = \alpha_i \tag{2.18}$$

and

$$P(Z = c_i|A^c) = P(Z = c_i|V = 1)P(V = 1|A^c) + P(Z = c_i|V = 0)P(V = 0|A^c) = \beta_i. \tag{2.19}$$

So, generating data using D is equivalent to first generating data using D_0 and then

randomizing them using the known probabilities $\{\alpha_i\}$ and $\{\gamma_i\}$.

In effect, D adds "random noise" to the data generated by D_0, from which it is quite intuitive that D_0 is more informative than D. This can be established formally following Anderson (1977), as discussed below. Let $I_V(\pi), I_Z(\pi)$ and $I_{VZ}(\pi)$ denote Fisher's information on π contained in V, Z and (V, Z), respectively. Then from general properties of Fisher's information it follows that

$$I_{VZ}(\pi) = I_V(\pi) + I_{Z|V}(\pi) = I_Z(\pi) + I_{V|Z}(\pi), \tag{2.20}$$

where $I_{Z|V}(\pi)$ is the average conditional information in Z given V and $I_{V|Z}(\pi)$ is defined similarly. Since the conditional distribution of Z given V does not depend on π, $I_{Z|V}(\pi) = 0$ and (2.20) implies that

$$I_V(\pi) - I_Z(\pi) = I_{V|Z}(\pi) \geq 0,$$

which completes the proof of the theorem. □

The proof of Theorem 2.3 shows that D is as informative as D_0 only when $I_{V|Z}(\pi) = 0$, or equivalently, the conditional distribution of V given Z is independent of π. Note that

$$P_\pi(V = 1|Z = c_j) = \frac{\alpha_j\pi + \alpha_j(\beta_1/\alpha_1)(1-\pi)}{\alpha_j\pi + \beta_j(1-\pi)}$$

is independent of π if and only if $\alpha_j(\beta_1/\alpha_1) = \beta_j$, i.e., $\alpha_j/\beta_j = \alpha_1/\beta_1$. So, the conditional distribution of V given Z is independent of π if and only if $\alpha_1/\beta_1 = \cdots = \alpha_k/\beta_k$, i.e., $\alpha_j = \beta_j$ for $j = 1, \cdots, k$, in which case D is non-informative. Thus, for any informative design D, $I_{V|Z}(\pi) > 0$ and hence D_0 is more informative than D.

Our proof of Theorem 2.3 is constructive; we not only show existence of a better design D_0 but also provide a recipe for finding one. The main implication of Theorem

2.3 is that for surveying dichotomous populations one should use only binary response variables, i.e., use only $(2 \to 2)$ RR designs. Then, Nayak's (1994) admissibility result (see, Remark 2.2) suggests that one should use only $(2 \to 2)$ RR designs with $P(Yes|A) = 1$. The design D_0, constructed in the proof of Theorem 2.3, is an admissible design. From Nayak (1994) and our Theorem 2.3 it can be seen that the best RR design at a specified level $(r \geq 1)$ of the privacy measure $R(\boldsymbol{\alpha}, \boldsymbol{\beta})$ is the $(2 \to 2)$ design with $\alpha_1 = 1, \alpha_2 = 0, \beta_1 = 1/r$ and $\beta_2 = 1 - 1/r$.

Remark 2.6. Information domination of D_0 over D can also be seen using Blackwell's (1951) ideas for comparing statistical experiments. Equations (2.18) and (2.19) show that D_0 is sufficient for D, by Blackwell's definition of sufficiency. Then, from Blackwell (1953) it follows that for every loss function $L(\pi, \hat{\pi})$ and any estimator $\hat{\pi}_D$ based on D, there exists an estimator $\hat{\pi}_*$ based on D_0 such that $E[L(\pi, \hat{\pi}_*)] \leq E[L(\pi, \hat{\pi})]$ for all $0 \leq \pi \leq 1$.

Remark 2.7. We may note that some papers, e.g., Greenberg et al. (1969) and Mangat and Singh (1990), compared variances without holding respondents' protection equal. Similarly, Christofides (2003) compared the variances of Warner's estimator and his estimator, based on his $(2 \to k)$ procedure, not taking respondents' protection into account. In his design, $\beta_i = \alpha_{(k-i+1)}, i = 1, \cdots, k$, and he showed that for any Warner's design with given p, one can find suitable values of the parameters $\{\alpha_i\}$ of his design, with $k \geq 3$, such that the variance of his estimator is less than that of Warner's estimator. Thence he concluded that his RR technique improves upon Warner's procedure. As an illustrative example, he took the Warner's estimator with $p = 0.6$, in which case

$$Var(\hat{\pi}_W) = \frac{\pi(1-\pi)}{n} + \frac{6}{n}. \tag{2.21}$$

Then he showed that his estimators, with $k = 6$ and

$(\alpha_1, \cdots, \alpha_6) = (0.38, 0.02, 0.19, 0.1, 0.05, 0.26)$, has variance

$$Var(\hat{\pi}_C) = \frac{\pi(1-\pi)}{n} + \frac{3.76}{n},$$

which is clearly less than $Var(\hat{\pi}_W)$ in (2.21) for all $0 \leq \pi \leq 1$. However, we find Christofides' (2003) argument and his conclusion to be flawed. First, he did not take respondents' protection into account. Second, if only the variances are compared, the fact that for any Warner's estimator, with given p, *there exists* a Christofides' estimator with smaller variance does not validate the conclusion that Christofides' *procedure* is better than Warner's *procedure*. This is because it can also be seen that for any given Christofides' estimator, *there exists* a Warner's estimator, with suitable choice of p, with uniformly smaller variance; note that the last term of (2.2) can be made arbitrarily small because it approaches 0 as p tends to 1 (or 0). For example, the Warner's estimator with $p = .65$, in which case the last term of (2.2) is $2.528/n < 3.76/n$, is better (in terms of variance) than the Christofides' estimator considered in the illustrative example.

2.8 Admissibility Results for $(2 \to 2)$ RR designs

For comparing $(2 \to 2)$ RR designs, Nayak (1994) used the posterior probabilities $P(Y = 1|Yes)$ and $P(Y = 1|No)$ under the given design to determine the level of respondents' privacy afforded by a design, and the variance of the UMVUE of π as a measure of efficiency of the design. With this approach he showed that a design D with transition probability matrix $\begin{pmatrix} \alpha & \beta \\ 1-\alpha & 1-\beta \end{pmatrix}$, where $\alpha = P_D(Yes|Y = 1)$ and $\beta = P_D(Yes|Y = 0)$, is admissible if and only if $\alpha = 1$.

While posterior probabilities $P(Y = 1|Yes)$ and $P(Y = 1|No)$ seem appropriate for assessing respondents' privacy, measuring efficiency by the variance of the UMVUE is a specific approach. First, this approach is specific to the choice of a particular estimator.

Second, the approach is tied to variance as a measure of statistical efficiency. A broader approach may come from using Blackwell's (1951) sufficiency principle for comparing experiments. However, for simplicity we generalize the comparison by taking the Fisher information as an efficiency measure, where $I_D(\pi)$ denote the Fisher's information on π contained in one observation using D. Thus, our approach does not depend on the choice of the estimator.

Definition 2.1. An RR design D is said to be inadmissible if there exists another RR design D_0 such that $P_{D_0}(Y=1|Yes) \leq P_D(Y=1|Yes)$,
$P_{D_0}(Y=1|No) \leq P_D(Y=1|No)$ and $I_{D_0}(\pi) \geq I_D(\pi)$ for all $\pi \epsilon [0,1]$ and at least one strict inequality holds for some π.

Theorem 2.4. *Let D be a given design with $P_D(Yes|Y=1) = \alpha$ and $P_D(Yes|Y=0) = \beta$. If $\alpha < 1$, then D is inadmissible.*

Proof. Let D be a given design with transition probability matrix $T = \begin{pmatrix} \alpha & \beta \\ 1-\alpha & 1-\beta \end{pmatrix}$, where $\alpha < 1$. Since the design space is $\{(\alpha,\beta) : 0 \leq \alpha, \beta \leq 1, \alpha > \beta\}$, this implies $0 \leq \beta < \alpha < 1$. The conditional probabilities of $Y=1$ under D are:

$$\begin{aligned} P_D(Y=1|Yes) &= \frac{\alpha\pi}{\alpha\pi + \beta(1-\pi)} \\ &= \left[1 + \frac{\beta}{\alpha}\frac{(1-\pi)}{\pi}\right]^{-1} \qquad (2.22) \\ P_D(Y=1|No) &= \frac{(1-\alpha)\pi}{(1-\alpha)\pi + (1-\beta)(1-\pi)} \qquad (2.23) \\ &= \left[1 + \frac{(1-\beta)}{(1-\alpha)}\frac{(1-\pi)}{\pi}\right]^{-1}. \end{aligned}$$

In the following, we present a design D_0 such that $P_{D_0}(Y=1|Yes) \leq P_D(Y=1|Yes)$ and $P_{D_0}(Y=1|No) \leq P_D(Y=1|No)$. We shall also show that D is equivalent to a random transformation of the outcome of our D_0, which implies that D_0 is more informative than D.

Specifically, let D_0 be the design whose transition probability matrix is given by $S = \begin{pmatrix} 1 & \frac{\beta}{\alpha} \\ 0 & 1-\frac{\beta}{\alpha} \end{pmatrix}$. Then, the conditional probabilities of $Y = 1$ under D_0 are $P_{D_0}(Y=1|Yes) = \left[1+\frac{\beta}{\alpha}\frac{(1-\pi)}{\pi}\right]^{-1}$ and $P_{D_0}(Y=1|No) = 0$. It is easy to that $P_{D_0}(Y=1|Yes) = P_D(Y=1|Yes)$ and $P_{D_0}(Y=1|No) = 0 < P_D(Y=1|No)$ for all $0 \leq \pi \leq 1$. Thus, D_0 provides greater respondents' privacy protection than D.

Next, we shall show that D is equivalent to a random transformation of the outcome of D_0. Thus, if we are given T and S as defined above, we shall show that there exist a transition probability matrix R such that $T = RS$. Obviously, since S is invertible by construction, $R = TS^{-1}$ satisfies the equation $T = RS$. However, we also need to verify that R is also a probability transformation matrix, that is, $\mathbf{1}'R = \mathbf{1}$, and individual entries of R are probabilities with values between zero and one. Here,

$$\begin{aligned} R = TS^{-1} &= \begin{pmatrix} \alpha & \beta \\ 1-\alpha & 1-\beta \end{pmatrix} \begin{pmatrix} 1 & \frac{\beta}{\alpha} \\ 0 & 1-\frac{\beta}{\alpha} \end{pmatrix}^{-1} \\ &= \frac{1}{(1-\frac{\beta}{\alpha})} \begin{pmatrix} \alpha & \beta \\ 1-\alpha & 1-\beta \end{pmatrix} \begin{pmatrix} 1-\frac{\beta}{\alpha} & -\frac{\beta}{\alpha} \\ 0 & 1 \end{pmatrix} \\ &= \frac{1}{(1-\frac{\beta}{\alpha})} \begin{pmatrix} \alpha(1-\frac{\beta}{\alpha}) & 0 \\ (1-\alpha)(1-\frac{\beta}{\alpha}) & (1-\beta)-\frac{\beta}{\alpha}(1-\alpha) \end{pmatrix} \\ &= \begin{pmatrix} \alpha & 0 \\ (1-\alpha) & 1 \end{pmatrix}. \end{aligned} \tag{2.24}$$

It is easy to verify from (2.24) that $\mathbf{1}'R = \mathbf{1}$. If we consider the individual entries of R, we note that $0 \leq \alpha \leq 1$ by construction. Thus, R is a transition probability matrix.

In effect, D adds "random noise" to the data generated by D_0, from which it is intuitive that D_0 is more informative than D. Let V be the response variable for D_0. Noting that $I_D(\pi) \equiv I_Z(\pi)$ and $I_{D_0}(\pi) \equiv I_V(\pi)$, let $I_V(\pi)$, $I_Z(\pi)$ and $I_{VZ}(\pi)$ denote Fisher's information on π contained in V, Z and (V, Z) respectively. Then following

the arguments used in the proof of Theorem 2.3, we get:

$$I_{D_0}(\pi) - I_D(\pi) = I_V(\pi) - I_Z(\pi) = I_{V|Z}(\pi) \geq 0,$$

which completes the proof. □

Therefore, a necessary condition for D to be admissible is that $P_D(Z = 1|Y = 1) = 1$. We shall now prove that $P_D(Z = 1|Y = 1) = 1$ is also a sufficient condition for the admissibility of D.

Theorem 2.5. *For a given D, if $P_D(Yes|Y = 1) = 1$ then D is admissible.*

Proof. Let D be a design with $P_D(Yes|Y = 1) = 1$. So, the transition probability matrix of D is $T = \begin{pmatrix} 1 & \beta \\ 0 & 1-\beta \end{pmatrix}$, for some $\beta < 1$. We shall show that there cannot exist any other design D^*, with transition probability matrix $\begin{pmatrix} a & b \\ 1-a & 1-b \end{pmatrix}$, such that all three following conditions hold: $P_{D^*}(Y = 1|Yes) \leq P_D(Y = 1|Yes)$, $P_{D^*}(Y = 1|No) \leq P_D(Y = 1|No)$ and $I_{D^*}(\pi) \geq I_D(\pi)$ for all $\pi \epsilon [0, 1]$ and at least one strict inequality holds for some π. We do this by showing that if D^* satisfies the first two conditions then it cannot satisfy the third condition. Let D^* satisfy the first two conditions. The inequality $P_{D^*}(Y = 1|No) \leq P_D(Y = 1|No) = 0$ for all π implies

$$\frac{(1-a)\pi}{(1-a)\pi + (1-b)(1-\pi)} \leq 0 \quad \text{for all} \quad 0 \leq \pi \leq 1,$$

which holds if and only if $a = P_{D^*}(Yes|Y = 1) = 1$. Then, $P_{D^*}(Y = 1|Yes) \leq P_D(Y = 1|Yes)$ implies

$$\left[1 + b\frac{(1-\pi)}{\pi}\right]^{-1} \leq \left[1 + \beta\frac{(1-\pi)}{\pi}\right]^{-1} \quad \text{for all} \quad 0 \leq \pi \leq 1,$$

or $b \geq \beta$. Thus, D^* must have a transition probability matrix of the form $S = \begin{pmatrix} 1 & b \\ 0 & 1-b \end{pmatrix}$ where $b \geq \beta$. To examine the third condition, we first note that

$$\begin{aligned} I_D(\pi) &= \frac{(1-\beta)^2}{\pi + \beta(1-\pi)} + \frac{(1-\beta)^2}{(1-\beta)(1-\pi)} \\ &= (1-\beta)^2 \left[\frac{1}{\pi + \beta(1-\pi)} + \frac{1}{(1-\beta)(1-\pi)} \right] \\ &= \frac{1-\beta}{(1-\pi)[\pi + \beta(1-\pi)]}. \end{aligned} \tag{2.25}$$

It is easy to see that (2.25) is a decreasing function of β. This implies $I_{D^*}(\pi) \leq I_D(\pi)$, because S and T have the same structure but with $b \geq \beta$. Thus, D cannot be dominated by another design D^* and consequently D is admissible. □

Chapter 3

Randomized Response Designs for Polychotomous Characteristics

3.1 Introduction

In the previous chapter, we dealt with $(2 \to 2)$ RR designs involving a dichotomous population with one sensitive or stigmatizing category. However in many situations, there may be more than two natural population categories of which at least one is sensitive. In this chapter, we shall use examples to illustrate how RR designs for polychotomous characteristics arise and examine how the previous results for $(2 \to 2)$ RR designs for dichotomous populations extend to $(k \to k)$ RR designs for polychotomous populations. Chaudhuri and Mukerjee (1988) gave an overview several methods that have been proposed and investigated to handle $(k \to k)$ RR designs. Two of the methods, as described in Abdul-Ela *et al.* (1967) and Greenberg *et al.* (1969), are extensions of the Warner's and Simmons' unrelated question methods respectively. These methods utilize dichotomous response categories to analyze RR designs for polychotomous characteristics, and they require the surveyor to select multiple samples for unbiased estimation of the proportions of the polychotomous characteristic. In this chapter, we adopt using polychotomous response categories as suggested by Bourke and Dalenius (1976). They argued that, using polychotomous response, a single sample is sufficient for estimation the distribution of the polychotomous characteristic.

Example 3.1. In a study of the prevalence of tax evasion, consider the problem of estimating the true population proportions π_1 and π_2 (and $\pi_3 = 1 - \pi_1 - \pi_2$) in three mutually exclusive and exhaustive categories consisting of

1. Never filed

2. Filed before the April 15 deadline

3. Sought an extension and filed before the October deadline

Category 1 is clearly sensitive while the other two categories can be regarded as innocuous. Since asking such sensitive questions directly could lead to respondents giving false answers, it makes sense to adopt an RR survey approach. To achieve this, an interviewer first selects a sample and then uses a randomization device such the one described by Liu and Chow (1976). The device consists of a jar containing, say, red and white balls. Each of the white balls is labeled $i = 1, \ldots, k$; In our Example 3.1, $k = 3$. The proportion of the red balls is p, while that of the white balls labeled j is p_j. The proportions are chosen such that $p + \sum_{j=1}^{k} p_j = 1$. The jar's neck is transparent so that only one ball can stand when the jar is turned upside down with a cork holding the balls up. Using this randomization device, the respondents are requested to shake the jar thoroughly, turn the jar upside down; thereby, allowing a ball to appear in the neck. If this ball is red, the respondent is requested report their true category. Otherwise, if the ball is white, the respondent will simply report the labeled number on the ball. This randomization experiment is unobserved by the surveyor, and the probability of getting the randomized response i is given by

$$\lambda_i = p\pi_i + p_i.$$

Another way in which a $(k \to k)$ RR design may arise is when starting with several independent $(2 \to 2)$ RR surveys of several binary variables, one wants to estimate the

joint probabilities arising from the cross classification of those variables. We illustrate this concept in the following example.

Example 3.2. Consider a population with two sensitive characteristics, such as drug use and gambling. Let the subset of the population who use drugs be classified as group A, while those who gamble be classified as group D. There are four population categories $\{A \cap D, A \cap D^c, A^c \cap D, A^c \cap D^c\}$ and we are interested in estimating all or some of the joint probabilities. Here the respondents are given a questionnaire where they are asked to answer one question from each of the following two pairs of questions :

R_1: Do you belong to group A? or R_1': Do you belong to group A^c?

R_2: Do you belong to group D? or R_2': Do you belong to group D^c?

Suppose the survey questions on these two variables are randomized independently via two separate $(2 \to 2)$ RR design experiments. We have four response categories given by $\{YY, YN, NY, NN\}$, and we can estimate the marginal probabilities, π_A and π_D by applying methods described in chapter 2. Using the independence property of the two separate RR design experiments, we can also estimate joint probabilities such as the proportion of the population that belongs to both groups A and D, π_{AD}. To do that, we need to analyze the two responses jointly by considering a $(4 \to 4)$ RR design. We can construct the *4* x *4* transition probability matrix P in terms of two *2* x *2* transition probability matrices as follows: Let the probability of answering R_1 be δ so that the probability of answering R_1' is $(1 - \delta)$. Likewise, let the probability of answering R_2 be γ so that the probability of answering R_2' is $(1-\gamma)$. Then, the transition probability matrices for the two separate RR designs are $P_1 = \begin{pmatrix} \delta & 1-\delta \\ 1-\delta & \delta \end{pmatrix}$ and $P_2 = \begin{pmatrix} \gamma & 1-\gamma \\ 1-\gamma & \gamma \end{pmatrix}$ respectively. Due to the independence of the two Warner RR designs, the transition probability matrix of the resulting $(4 \to 4)$ RR design is given by the Kronecker product

$$P_1 \otimes P_2 = \begin{pmatrix} \delta\gamma & \delta(1-\gamma) & (1-\delta)\gamma & (1-\delta)(1-\gamma) \\ \delta(1-\gamma) & \delta\gamma & (1-\delta)(1-\gamma) & (1-\delta)\gamma \\ (1-\delta)\gamma & (1-\delta)(1-\gamma) & \delta\gamma & \delta(1-\gamma) \\ (1-\delta)(1-\gamma) & (1-\delta)\gamma & \delta(1-\gamma) & \delta\gamma \end{pmatrix}.$$

In the next section, we present a general framework for polychotomous RR designs. In Section 3.3, we discuss estimation under an infinite population setting and extend our results to situations in finite population settings. In Section 3.4, we discuss comparisons of a class of polychotomous RR designs where only one category is sensitive. Specifically, we derive conditions for which one can construct better polychotomous RR designs of a specified structure.

3.2 General Framework

Consider a polychotomous population with k mutually exclusive and exhaustive categories labeled $j = 1, 2, \ldots, k$, of which at least one category is sensitive. The unknown population proportions $\pi_j = P(Y = j)$, are the parameters of interest. Also, consider a RR survey where the response variable Z is also categorical with k categories. The randomization device is such that a respondent belonging to the jth category represented by $Y = j$ reports $Z = i$ with probability $p_{ij} = P(Z = i|Y = j)$ so that $P = ((p_{ij}))$ denotes a transition probability transition matrix. We shall consider all $(k \to k)$ RR surveys with known P such that the probabilities in each column add up to 1.

$$\sum_{i=1}^{k} p_{ij} = 1$$

The response variable is Z with possible values $1, \cdots, k$, and the probability λ_i of the randomized response being i is given by

$$\begin{aligned}\lambda_i = P(Z = i) &= \sum_{j=1}^{k} p_{ij}\pi_j \\ &= \sum_{j=1}^{k-1} p_{ij}\pi_j + p_{ik}(1 - \sum_{j=1}^{k-1} \pi_j)\end{aligned}$$

since $\sum_{j=1}^{k} \pi_j = 1$. In matrix notation, we have $\boldsymbol{\lambda} = P\boldsymbol{\pi}$ where P is a nonsingular transformation probability matrix. Then, the following posterior probabilities determine the level of respondents' privacy:

$$\begin{aligned}P(Y = j|Z = i) &= \frac{p_{ij}\pi_j}{\lambda_i} \\ &= \frac{p_{ij}\pi_j}{\sum_{j=1}^{k-1} p_{ij}\pi_j + p_{ik}(1 - \sum_{j=1}^{k-1} \pi_j)}, \quad i = 1, \cdots, k \end{aligned} \tag{3.1}$$

Let X_i denote the frequency of the response i in the sample, and let n denote the sample size. Then, under SRSWR, $(X_1, \cdots, X_k) \sim Mult(n, \lambda_1, \cdots, \lambda_k)$ with probability mass function

$$\begin{aligned}f(x_1, \cdots, x_k) &= P(X = x_1, \cdots, X_k = x_k) \\ &= \frac{n!}{x_1! \cdots x_k!} \lambda_1^{x_1} \cdots \lambda_k^{x_k}. \end{aligned} \tag{3.2}$$

Note that the posterior probabilities in (3.1), the distribution of $(X_1, \cdots, X_k)$ and the Fisher's information depend on the randomization mechanism only through $((p_{ij}))$. Thus, all $(k \to k)$ RR procedures can be characterized by the values of $\{p_{ij}\}$. This implies that for designing a $(k \to k)$ RR survey we should first determine the values of $\{p_{ij}\}$ and then give a mechanism for implementing them. For a unified approach, we suggest to take $\{p_{ij}\}$ as the RR design parameters and discuss and examine all $(k \to k)$

RR procedures through them.

The ordering of the k true categories and k response categories has no bearing on the substantive properties of a $(k \to k)$ RR design. Evidently, a $(k \to k)$ RR design remains essentially unchanged under any permutation of the categories and responses, that is, permutations of the rows and columns of P. So, for unique characterization of a $(k \to k)$ RR design by $P = (p_{ij})$, we need a convention for representing the matrix. One possibility is to order the population categories first by their level of sensitivity. Essentially, we propose ordering the categories $1, \ldots, k$ with category 1 being the most sensitive and category k is not sensitive. Next, for the most sensitive category 1, order the transition probabilities in decreasing order of magnitude as follows: $p_{11} > p_{12} > \ldots > p_{1k}$. This gives a unique representation of the RR design provided there are no ties. If there are ties, simply order the transition probabilities for the next most sensitive category to determine the tie breaker.

3.3 Infinite Population Estimation

In this section, we discuss the maximum likelihood and method of moments estimators under an infinite population setting. Then, for a finite population setting, we show how to modify homogeneous linear unbiased estimators (e.g., Horvitz-Thompson estimator) based on an open survey can be modified to obtain an unbiased estimator under a polychotomous RR survey.

3.3.1 Maximum Likelihood Estimator

The maximum likelihood estimate (MLE) of $\boldsymbol{\pi} = (\pi_1, ..., \pi_k)$ based on $(X_1, \cdots, X_k)$ is the solution of the set of likelihood equations $\frac{\partial}{\partial \pi_i} ln\, f(x_1, \cdots, x_k) = 0$, $i = 1, ..., k-1$,

with $\pi_k = 1 - \sum_{j=1}^{k-1} \pi_j$. In particular, $ln\, f(x_1, \cdots, x_k) = \sum_{i=1}^{k} x_i ln\, \lambda_i + constant$, and

$$\begin{aligned}
\frac{\partial}{\partial \pi_i} ln\, f(x_1, \cdots, x_k) &= \frac{\partial}{\partial \pi_i} \sum_{i=1}^{k} x_i ln\, \lambda_i \\
&= \frac{\partial}{\partial \pi_i} \sum_{i=1}^{k} x_i ln \left[\sum_{j=1}^{k-1} p_{ij}\pi_j + p_{ik}(1 - \sum_{j=1}^{k-1} \pi_j) \right] \\
&= \sum_{i=1}^{k} \frac{x_i(p_{ij} - p_{ik})}{\sum_{j=1}^{k-1} p_{ij}\pi_j + p_{ik}(1 - \sum_{j=1}^{k-1} \pi_j)} = 0. \qquad (3.3)
\end{aligned}$$

We are interested in a solution of (3.3) provided that it is in the parameter space, $\pi_j > 0$ and $\sum_{j=1}^{k} \pi_j = 1$. Van den Hout and Van den Heijden (2002) suggests to calculate the maximum likelihood estimator via the EM algorithm.

Since $\boldsymbol{\pi}_{(k-1)} = \pi_1, ..., \pi_{k-1}$ are the only parameters free to vary, for $l, m \in \{1, ..., k-1\}$, the (l, m)-th entry of the Fisher information matrix for $\boldsymbol{\pi}_{(k-1)}$ with dimension $(k-1)$ by $(k-1)$, is given by

$$\begin{aligned}
I_{lm}(\boldsymbol{\pi}_{(k-1)}) &= -E \left[\frac{\partial^2}{\partial \pi_l \pi_m} log\, f(x_1, \cdots, x_{k-1}) \right] \\
&= \sum_{i=1}^{k-1} \frac{E(x_i)(p_{lj} - p_{kj})(p_{mj} - p_{kj})}{\lambda_i^2} \\
&= \sum_{i=1}^{k-1} \frac{n\lambda_i(p_{lj} - p_{kj})(p_{mj} - p_{kj})}{\lambda_i^2} \\
&= \sum_{i=1}^{k-1} \frac{n(p_{lj} - p_{kj})(p_{mj} - p_{kj})}{\lambda_i}. \qquad (3.4)
\end{aligned}$$

We can estimate (3.4) by

$$\sum_{i=1}^{k-1} \frac{n(p_{lj} - p_{kj})(p_{mj} - p_{kj})}{\hat{\lambda}_i}$$

where $\hat{\lambda}_i = \frac{x_i}{n}$. The information matrix $I(\boldsymbol{\pi}_{(k-1)}) = \left(\left(I_{lm}(\boldsymbol{\pi}_{(k-1)}) \right) \right)$.

The asymptotic distribution of the MLE is normal, so

$$\sqrt{n}(\hat{\boldsymbol{\pi}}_{(k-1)} - \boldsymbol{\pi}_{(k-1)}) \xrightarrow{L} N(0,\ I^{-1}(\boldsymbol{\pi}_{(k-1)})) \quad \text{as} \quad n \to \infty.$$

Noting that $\boldsymbol{\pi}_{(k)} = (0, \ldots, 0, 1)' + J'\boldsymbol{\pi}_{(k-1)}$, and then applying the delta method, the asymptotic distribution of the MLE $\hat{\boldsymbol{\pi}}$ is normal with

$$\sqrt{n}(\hat{\boldsymbol{\pi}} - \boldsymbol{\pi}) \xrightarrow{L} N(0,\ J'I^{-1}(\boldsymbol{\pi}_{(k-1)})J)$$

where

$$J = \left(\left(\frac{\partial \pi_j}{\partial \theta_i}\right)\right) = \begin{pmatrix} 1 & 0 & \cdots & 0 & -1 \\ 0 & 1 & \cdots & 0 & -1 \\ 0 & 0 & \cdots & 0 & -1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 1 & -1 \end{pmatrix}$$

This asymptotic distribution can be used to construct large sample confidence intervals for $\boldsymbol{\pi}$.

3.3.2 Method of Moments Estimator

The MLE of $\boldsymbol{\pi}$ has been discussed above. Since $\widehat{\boldsymbol{\lambda}} = (\frac{X_1}{n}, ..., \frac{X_k}{n})$ is unbiased for $\boldsymbol{\lambda}$ it can be seen using method of moments (MOM) that an unbiased estimate of $\boldsymbol{\pi}$ is given by

$$\widehat{\boldsymbol{\pi}} = P^{-1}\widehat{\boldsymbol{\lambda}}$$

The variance-covariance matrix is given by

$$\begin{aligned} V(\widehat{\boldsymbol{\pi}}) &= P^{-1}V(\widehat{\boldsymbol{\lambda}})(P^{-1})' \\ &= n^{-1}P^{-1}(D_{\boldsymbol{\lambda}} - \boldsymbol{\lambda}\boldsymbol{\lambda}')(P^{-1})' \\ &= \frac{1}{n}(D_{\boldsymbol{\pi}} - \boldsymbol{\pi}\boldsymbol{\pi}') + \frac{1}{n}P^{-1}(D_{\boldsymbol{\lambda}} - PD_{\boldsymbol{\pi}}P')(P^{-1})' \end{aligned} \tag{3.5}$$

where $D_{\boldsymbol{\lambda}}$ is a diagonal matrix with diagonal elements being $\lambda_1, \ldots, \lambda_k$ and $D_{\boldsymbol{\pi}}$ is defined similarly. The first term on the right side of (3.5) is the variance due to the multinomial scheme and under no randomization, and the second part represents the additional

variance induced by randomization.

Van den Hout and Heijden (2002) showed that the MLE and the MOM estimates coincide when both are in the interior of the parameter space of $\boldsymbol{\pi}$**.** However, since the MLE is one-to-one under transformation and the MOM estimator is not, the MOM estimates π_j can sometimes take on negative values which fall outside the parameter space.

3.4 Finite Population Estimation of $\boldsymbol{\pi}$

In this section we derive estimators based on polychotomous RR data from a finite population and unequal probability sampling. Specifically, we shall show how a linear unbiased estimator based on an open survey can be modified to obtain an unbiased estimator under an RR survey. We also discuss unbiased estimation of variance of the estimators from polychotomous RR survey data.

We shall represent each true category and each randomized response using a k-dimensional vector $\boldsymbol{Y}$. Specifically, if the true category is j then $\boldsymbol{Y} =$ $\boldsymbol{e}_j = (0, 0, \ldots, 1, \ldots 0)'$; the jth element of $\boldsymbol{e}_j$ is 1 and the other elements are zeroes. Thus $\boldsymbol{Y} \epsilon \{e_1, e_2, \ldots, e_k\} = \Omega$. Consider a finite population of N units, labeled $i = 1, \cdots, N$, and let $\boldsymbol{Y}_i$ denote the value of $\boldsymbol{Y}$, which is an indicator vector of the sensitive variable, for unit i. If $\boldsymbol{Y}_i = \boldsymbol{e}_j$ then unit i belongs to the jth category. Let $\boldsymbol{Z}_i$ denote the response of unit i. Note that $\boldsymbol{Z}_i \epsilon \, \Omega$ and suppose that the sample is selected using a non-informative sampling design $p(s)$. So, the data can be represented as $\{(i, \boldsymbol{Z}_i); i \in s\}$, where s is a subset of $\{1, \cdots, N\}$. Our goal is to estimate the vector $\boldsymbol{\pi} = (\sum_{i=1}^{N} \boldsymbol{Y}_i)/N$. Note that while $\boldsymbol{Y}_i$ are fixed, $\boldsymbol{Z}_i$ are random variables, and estimation of $\boldsymbol{\pi}$ is equivalent to estimation of the population vector $T(\boldsymbol{Y}) = \sum_{i=1}^{N} \boldsymbol{Y}_i$.

A common approach to estimation from survey data is to apply survey weights to survey data. These weights, w_{si}, are values assigned to each record i in the data file,

and they are normally used to make statistics estimated from the survey data more representative of the population. Suppose

$$e(s, \boldsymbol{y}) = \sum_{i \in s} w_{si} \boldsymbol{Y}_i \tag{3.6}$$

is a homogeneous linear unbiased estimator of $T(\boldsymbol{Y})$, i.e.,

$$E_p[e(s, \boldsymbol{y})] = \sum_s e(s, \boldsymbol{y}) p(s) = \sum_{i=1}^{N} \boldsymbol{Y}_i \quad \text{for all } \boldsymbol{Y}_1, \cdots, \boldsymbol{Y}_N$$

that is,

$$\begin{aligned} \sum_{i=1}^{N} \boldsymbol{Y}_i &= \sum_s (\sum_{i \in s} w_{si} \boldsymbol{Y}_i) p(s) \\ &= \sum_s \sum_{i \in s} w_{si} \boldsymbol{Y}_i p(s) \\ &= \sum_{i=1}^{N} \boldsymbol{Y}_i \sum_{s \ni i} w_{si} p(s). \end{aligned}$$

Thus

$$\sum_{s \ni i} w_{si} p(s) = 1, \quad i = 1, \cdots, N.$$

Since $P(\boldsymbol{Z} = \boldsymbol{e}_h | \boldsymbol{Y} = \boldsymbol{e}_j) = p_{hj}$, we have

$$E[\boldsymbol{Z} | \boldsymbol{Y} = \boldsymbol{e}_j] = \sum_{h=1}^{k} p_{hj} \boldsymbol{e}_j = \boldsymbol{p}_j$$

where $\boldsymbol{p}_j$ is the jth column vector of P. So we have,

$$E[\boldsymbol{Z} | \boldsymbol{Y}] = P\boldsymbol{Y} \tag{3.7}$$

Since P is nonsingular, it follows that $E_R(P^{-1}\boldsymbol{Z}) = P^{-1}E(\boldsymbol{Z}|\boldsymbol{Y}) = P^{-1}P\boldsymbol{Y} = \boldsymbol{Y}$, where E_R denotes expectation with respect to the randomization mechanism. Thus, $P^{-1}\boldsymbol{Z}$ is an unbiased recovery of $\boldsymbol{Y}$. Using the same weights w_{si} via a natural deperturbation

approach, let

$$e^*(s,z) = \sum_{i\in s} w_{si}\left(\boldsymbol{P}^{-1}\boldsymbol{Z}_i\right), \tag{3.8}$$

Then,

$$\begin{aligned} E[e^*(s,\boldsymbol{z})] &= E_p E_R[\sum_{i\in s} w_{si}\left(\boldsymbol{P}^{-1}\boldsymbol{Z}_i\right)] \\ &= E_p[\sum_{i\in s} w_{si} E_R\left(\boldsymbol{P}^{-1}\boldsymbol{Z}_i\right)] \\ &= E_p[\sum_{i\in s} w_{si}\boldsymbol{Y}_i] \\ &= E_p[e(s,\boldsymbol{y})] = T(\boldsymbol{Y}). \end{aligned}$$

Thus, we have the following:

Theorem 3.1. *For any given sampling design $p(s)$, if $e(s,\boldsymbol{y})$ in (3.6) is a homogeneous linear design unbiased estimator of the population total $T(\boldsymbol{Y}) = \sum_{i=1}^{N}\boldsymbol{Y}_i$ based on the open survey, then the estimator $e^*(s,\boldsymbol{z})$ in (3.8) is a design unbiased estimator of $T(\boldsymbol{Y})$ based on the RR survey with RR design matrix P and sampling design $p(s)$.*

3.4.1 Variance Estimation

We shall now focus on the variance of the estimator $e^*(s,\boldsymbol{z})$, defined in (3.8). First note that

$$V[\boldsymbol{Z}|\boldsymbol{Y}=\boldsymbol{e}_j] = E[\boldsymbol{Z}\boldsymbol{Z}'|\boldsymbol{Y}=\boldsymbol{e}_j] - E[\boldsymbol{Z}|\boldsymbol{Y}=\boldsymbol{e}_j]E[\boldsymbol{Z}'|\boldsymbol{Y}=\boldsymbol{e}_j]$$

We can express $E[\boldsymbol{Z}\boldsymbol{Z}'|\boldsymbol{Y}=e_j]$ as

$$\sum_{h=1}^{k}\boldsymbol{e}_h\boldsymbol{e}_h' p_{hj} = D_{\boldsymbol{p}_j} \tag{3.9}$$

where $D_{\boldsymbol{p}_j}$ is the diagonal matrix formed by elements of the column vector $\boldsymbol{p}_j$. We note that $D_{\boldsymbol{p}_j} = D_{\boldsymbol{PY}}$ for $(\boldsymbol{Y} = \boldsymbol{e}_j)$. Hence from (3.7) and (3.9),

$$V[\boldsymbol{Z}|\boldsymbol{Y}] = D_{\boldsymbol{PY}} - (\boldsymbol{PY})(\boldsymbol{PY})^{'} = D_{\boldsymbol{PY}} - P\boldsymbol{YY}'P'. \tag{3.10}$$

Using (3.10), the variance-covariance of $e^*(s,z)$ can be written as

$$\begin{aligned}
V(e^*(s,\boldsymbol{z})) &= E_p V_R(e^*(s,\boldsymbol{z})|s,\boldsymbol{Y}) + V_p E_R(e^*(s,\boldsymbol{z})|s,\boldsymbol{Y}) \\
&= E_p[\sum_{i\in s} w_{si}^2 P^{-1}\left\{V[\boldsymbol{Z}_i|\boldsymbol{Y}_i]\right\} P^{'-1}] + V_p(e(s,y))] \\
&= E_p[\sum_{i\in s} w_{si}^2 P^{-1}\left\{D_{\boldsymbol{PY}_i} - P\boldsymbol{Y}_i\boldsymbol{Y}_i^{'}P'\right\} P^{'-1}] + V_p(e(s,y))] \\
&= E_p[\sum_{i\in s} w_{si}^2 \left\{P^{-1}D_{\boldsymbol{PY}_i}P^{'-1}\right\} - \boldsymbol{Y}_i\boldsymbol{Y}_i^{'}] + V_p(e(s,y))]
\end{aligned} \tag{3.11}$$

The first term in (3.11) is the extra variation due to randomization. Noting that in our application $\boldsymbol{Y}_i\boldsymbol{Y}_i^{'} = \boldsymbol{D}_{Y_i}$, we can rewrite (3.11) as

$$V(e^*(s,\boldsymbol{z})) = E_p[\sum_{i\in s} w_{si}^2\left\{P^{-1}D_{\boldsymbol{PY}_i}P'^{-1}\right\} - D_{\boldsymbol{Y}_i}] + V_p(e(s,\boldsymbol{y}))]$$

Also,

$$\begin{aligned}
V_p(e(s,\boldsymbol{y})) &= E_p\{[e(s,\boldsymbol{y})]\,[e(s,\boldsymbol{y})]^{'}\} - \{E_p\,[e(s,\boldsymbol{y})]\}\{E_p\,[e(s,\boldsymbol{y})]\}^{'} \\
&= E_p\{[\sum_{i\epsilon s} w_{si}\boldsymbol{Y}_i][\sum_{i\epsilon s} w_{si}\boldsymbol{Y}_i]'\} - \{E_p[\sum_{i\epsilon s} w_{si}Y_i]E_p[\sum_{i\epsilon s} w_{si}Y_i]'\} \\
&= E_p\{\sum_{i\epsilon s} w_{si}^2\boldsymbol{Y}_i\boldsymbol{Y}_i' + \sum_{\substack{i,j\epsilon s \\ i\neq j}} w_{si}w_{sj}\boldsymbol{Y}_i\boldsymbol{Y}_j')\} - \{(\sum_{i=1}^{N}\boldsymbol{Y}_i)(\sum_{i=1}^{N}\boldsymbol{Y}_i)'\}^2 \\
&= \sum_{i=1}^{N} D_{\boldsymbol{Y}_i}\sum_{s\ni i} w_{si}^2 p(s) + \sum_{\substack{i,j=1 \\ i\neq j}}^{N}\boldsymbol{Y}_i\boldsymbol{Y}_j'\sum_{s\ni i,j} w_{si}w_{sj}p(s) - (\sum_{i=1}^{N}\boldsymbol{Y}_i\boldsymbol{Y}_i' + \sum_{\substack{i,j=1 \\ i\neq j}}^{N}\boldsymbol{Y}_i\boldsymbol{Y}_j') \\
&= \sum_{i=1}^{N} D_{\boldsymbol{Y}_i}\sum_{s\ni i} w_{si}^2 p(s) + \sum_{\substack{i,j=1 \\ i\neq j}}^{N}\boldsymbol{Y}_i\boldsymbol{Y}_j'\sum_{s\ni i,j} w_{si}w_{sj}p(s) - (\sum_{i=1}^{N} D_{\boldsymbol{Y}_i} + \sum_{\substack{i,j=1 \\ i\neq j}}^{N}\boldsymbol{Y}_i\boldsymbol{Y}_j')
\end{aligned}$$

$$= \sum_{i=1}^{N} g_i D_{\boldsymbol{Y}_i} + \sum_{\substack{i,j=1 \\ i \neq j}}^{N} g_{ij} \boldsymbol{Y}_i \boldsymbol{Y}_j^{'}, \quad \text{say,}$$

where,

$$g_i = \sum_{s \ni i} w_{si}^2 p(s) - 1, \quad \text{and} \quad g_{ij} = \sum_{s \ni i,j} w_{si} w_{sj} p(s) - 1.$$

Now we discuss how an unbiased estimator of $V(e^*(s, \boldsymbol{z})$ can be obtained from an unbiased estimator of $V_p(e(s, \boldsymbol{y})$ based on an open survey. Let d_{si} and d_{sij} be such that

$$\sum_{s \ni i} d_{si} p(s) = g_i \quad \text{and} \quad \sum_{s \ni i,j} d_{sij} p(s) = g_{ij},$$

so that

$$t(s, \boldsymbol{y}) = \sum_{i \in s} d_{si} D_{\boldsymbol{Y}_i} + \sum_{\substack{i,j \in s \\ i \neq j}} d_{sij} \boldsymbol{Y}_i \boldsymbol{Y}_j \tag{3.12}$$

is an unbiased estimator of $V_p(e(s, \boldsymbol{y}))$ based on the open survey data. A specific unbiased estimator of $V_p(e(s, \boldsymbol{y}))$ is obtained by using $d_{si} = g_i/\pi_i$ and $d_{sij} = g_{ij}/\pi_{ij}$, where $\pi_i = \sum_{s \ni i} p(s)$ and $\pi_{ij} = \sum_{s \ni i,j} p(s)$. From (3.12) and (3.11) the following result can now be established using the fact that $E_R(P^{-1}\boldsymbol{Z}|\boldsymbol{Y}) = \boldsymbol{Y}$.

Theorem 3.2. *An unbiased estimator of $V(e^*(s, \boldsymbol{z}))$, based on RR survey data, is given by*

$$t^*(s, \boldsymbol{z}) = \sum_{i \in s} d_{si} P^{-1} D_{\boldsymbol{Z}_i} + \sum_{\substack{i,j \in s \\ i \neq j}} d_{sij} \boldsymbol{Z}_i \boldsymbol{Z}_j + [\sum_{i \in s} w_{si}^2 \left\{ P^{-1} D_{\boldsymbol{Z}_i} P'^{-1} \right\} - P^{-1} D_{\boldsymbol{Z}_i}].$$

3.5 Comparison of $(k \to k)$ RR Designs

For comparing two $(k \to k)$ RR procedures one should examine both the respondents' privacy and statistical efficiency afforded by the two procedures. In particular, we should hold some privacy measure at the same level for the two procedures and then compare their statistical efficiencies. Let Y be the true variable with categories $1, \cdots, k$. As

a starting point, we shall consider the situation where only one category is sensitive. Without loss of generality, suppose the first category (i.e., $Y = 1$) is sensitive. See Example 3.1 above where only the first answer is sensitive. Let Z be the response variable also with categories $1, \cdots, k$. Further suppose the transition probabilities are given by $p_{ij} = P(Z = i|Y = j)$, and let $P = ((p_{ij}))$ denote the transition probability matrix. By construction, the columns add up to 1, that is, $1'P = 1'$.

We shall use posterior probabilities to provide a basis for comparing the privacy protection afforded by two designs. Since only category 1 is sensitive for discussing (or controlling) respondents' protection, we only need to consider $P(Y = 1|Z = i)$ for $i = 1, \cdots, k$. For a general design with transition probability matrix P, assumed to be nonsingular, these probabilities are:

$$\begin{aligned} P(Y = 1|Z = i) &= \frac{\pi_1 p_{i1}}{\sum_{l=1}^{k} \pi_l p_{il}} \\ &= \frac{\pi_1}{\pi_1 + \sum_{l=2}^{k} \pi_l \left\{ \frac{p_{il}}{p_{i1}} \right\}} \\ &= h_i, \text{ say.} \end{aligned} \tag{3.13}$$

We note that:

- h_i depends on $\boldsymbol{\pi} = (\pi_1, \pi_2, \cdots, \pi_k)$ and the ith row of P and
- $h_i = 0$ if $p_{i1} = 0$ (and $p_{il} > 0$ for some $l \geq 2$).

Definition 3.1. (Comparison of respondents' privacy): An RR design D_1 is at least as protective as another design D_2 if $P_{D_1}(Y = 1|Z = i) \leq P_{D_2}(Y = 1|Z = i)$ for $i = 1, \ldots, k$.

Next, we consider the efficiency of competing designs. There are many possible approaches for comparing the statistical efficiency. One method discussed by Anderson

(1977) is to compare Fisher's information matrices from designs D_1 and D_2. If the difference of the information matrices $I_{D_1}(\pi) - I_{D_2}(\pi)$ is non-negative definite then D_1 contains more information than D_2. Another method, which we adopt, is based on the sufficiency principle using Blackwell's (1951) ideas for comparing statistical experiments.

Definition 3.2. (Comparison of efficiency): An RR design D_1 is more efficient than (or sufficient for) another design D_2 if the outcome of D_2 is statistically equivalent to randomizing the outcome of D_1.

Comparing $(k \to k)$ RR designs is a difficult problem since it requires satisfying many conditions. Definition 3.1, for instance, has k conditions. Recall that Nayak's (1994) admissibility result for $(2 \to 2)$ RR designs we have an admissible design if all the "Yes" responses stay as "Yes" and some "No" responses are transformed to "Yes" (see Remark 2.2). That is, we protect privacy by introducing false "Yes" responses. We want to explore if this phenomenon holds for $(k \to k)$ RR designs with exactly one sensitive category (category 1). As an analogy, we want to know what happens if all sensitive category 1 cases remain as category 1, and we only randomize the nonsensitive cases; thereby, introducing false category 1 cases. Based on what we know for $(2 \to 2)$ RR designs, we might do the following:

1. Make $p_{11} = 1$ and set the remaining probabilities on the first column of the transition matrix, i.e., $p_{i1} = 0$ for $i \geq 2$.

2. Randomize each of the other categories with just the sensitive category 1 (or rather response category 1 which is the only response for individuals in the true category 1).

The transition probability matrix for designs with the preceding features have the struc-

ture

$$P_* = \begin{pmatrix} 1 & a_2 & \dots & a_k \\ 0 & 1-a_2 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & 1-a_k \end{pmatrix}. \tag{3.14}$$

Here we want to explore if designs with a transition probability of the form (3.14) are sufficient. Specifically, given P, the transition probability matrix for a general design D, we want to find out if there exist a design D^* with transition probability matrix $P_* = \left(\left(p^*_{ij}\right)\right)$ with suitable values $a_2, \cdots, a_k$ such that D^* is better. For P_*, the conditional probabilities in (3.14) reduce to

$$h^*_i = P_D(Y = 1|Z = i) = \begin{cases} 0 & \text{if } i \neq 1 \\ \frac{\pi_1}{\pi_1 + \sum_{l=2}^k \pi_l a_l} & \text{if } i = 1 \end{cases}$$

From Definition 3.1, D_* is at least as protective as D when $P_{D_*}(Y = 1|Z = i) \leq P_D(Y = 1|Z = i), i = 2, \cdots, k$; or $h^*_i \leq h_i$. We note that $p^*_{i1} = 0$ implies $h^*_i = 0 \leq h_i$ for $i \geq 2$. So, we just need to make sure that $h^*_1 \leq h_1$. Let us denote p^*_{1l} by a_l $(l \geq 2)$. Then

$$h^*_1 \leq h_1 \Longleftrightarrow \sum_{l=2}^k \pi_l a_l \geq \sum_{l=2}^k \pi_l \frac{p_{1l}}{p_{11}} \tag{3.15}$$

We note that (3.15) needs to be true for all $\pi_i, i = 1, \ldots, k$, which implies $a_l \geq \frac{p_{1l}}{p_{11}}$. Consider P with $\frac{p_{1l}}{p_{11}} < 1$ for all l. While this is a mathematical restriction, it is common in practice that transition probability matrices typically have $p_{11} > 0.5$. We can choose any a_l satisfying (3.15), and one choice is to take $a_l = \frac{p_{1l}}{p_{11}}$. For $i \geq 2$ and $j \geq 2$, define

$$p^*_{ij} = \begin{cases} 1 - a_i & \text{if } i = j \\ 0 & \text{if } i \neq j. \end{cases}$$

The choices of the elements of P_* satisfy (3.15) and guarantee that $h^*_i \leq h_i$ for all i, so

D_* is at least as protective as D.

In applying Definition 3.2, we shall examine conditions for which data generated by D is equivalent to further randomizing the data generated by D_*. These are conditions for which the data generated by applying transition probability matrix P can be produced by using a known randomization of the data generated by P_*. Let U denote the response variable for P_*. Also, let $R = ((r_{il}))$ be a transition probability matrix and consider applying R after applying P_*. Let W be the resulting response variable. Then, $r_{il} = P(W = i|U = l)$ and

$$P(W = i|Y = j) = \sum_{l=1}^{k} P(W = i|U = l)P(U = l|Y = j)$$

$$= \sum_{l=1}^{k} r_{il} p^*_{lj} = \lambda_{ij}$$

In matrix form this can be expressed as $\Lambda = RP_*$. We shall now investigate if there exists a transition probability matrix R such that $\Lambda = P$, that is $P = RP_*$. In R exists then P adds "random noise" to the data generated by P_*, from which it follows that P_* is more informative than P.

Since P_* is nonsingular by construction, $R = PP_*^{-1}$ is unique. R is a transition matrix only if $\mathbf{1}'R = \mathbf{1}'$ and $0 \leq r_{il} \leq 1$ for every l, i. We note that $\mathbf{1}'R = \mathbf{1}'PP_*^{-1} = \mathbf{1}'P_*^{-1}$, but since $\mathbf{1}'P_* = \mathbf{1}' \Longrightarrow \mathbf{1}'P_*P_*^{-1} = \mathbf{1}'P_*^{-1}$ or $\mathbf{1}' = \mathbf{1}'P_*^{-1}$, we can conclude that $\mathbf{1}'R = 1'$.

We shall now examine when the conditions $0 \leq r_{il} \leq 1$ are satisfied:

$$R = \begin{pmatrix} p_{11} & p_{12} & \dots & p_{1k} \\ p_{21} & p_{22} & \dots & p_{2k} \\ \vdots & \vdots & \dots & \vdots \\ p_{k1} & p_{k2} & \dots & p_{kk} \end{pmatrix} \begin{pmatrix} 1 & a_2 & \dots & a_k \\ 0 & 1-a_2 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & 1-a_k \end{pmatrix}^{-1}$$

$$= \begin{pmatrix} p_{11} & p_{12} & \dots & p_{1k} \\ p_{21} & p_{22} & \dots & p_{2k} \\ \vdots & \vdots & \dots & \vdots \\ p_{k1} & p_{k2} & \dots & p_{kk} \end{pmatrix} \begin{pmatrix} 1 & \frac{-a_2}{1-a_2} & \dots & \frac{-a_k}{1-a_k} \\ 0 & \frac{1}{1-a_2} & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & \frac{1}{1-a_k} \end{pmatrix}$$

$$= \begin{pmatrix} p_{11} & \frac{p_{12}-a_2p_{11}}{1-a_2} & \dots & \frac{p_{1k}-a_kp_{11}}{1-a_k} \\ p_{21} & \frac{p_{22}-a_2p_{21}}{1-a_2} & \dots & \frac{p_{2k}-a_kp_{21}}{1-a_k} \\ \vdots & \vdots & \dots & \vdots \\ p_{k1} & \frac{p_{k2}-a_2p_{k1}}{1-a_2} & \dots & \frac{p_{kk}-a_kp_{k1}}{1-a_k} \end{pmatrix}.$$

For R to be a transition probability matrix, $0 \le r_{il} \le 1 \Longrightarrow 0 \le \frac{p_{il}-a_lp_{i1}}{1-a_l} \le 1$, or $a_l \le \frac{p_{il}}{p_{i1}}$ and $a_l \le \frac{1-p_{il}}{1-p_{i1}}$. Thus, we need

$$a_l \le \begin{cases} min_i \left\{ \frac{p_{il}}{p_{i1}}, \frac{1-p_{il}}{1-p_{i1}} \right\} & \text{if } 0 \le p_{i1} \le 1 \\ 1 - p_{il} & \text{if } p_{i1} = 0 \\ p_{il} & \text{if } p_{i1} = 1 \end{cases}, \tag{3.16}$$

for $i = 1, \dots, k$ and $l = 2, \dots, k$. If we take $a_l = \frac{p_{1l}}{p_{11}}$ then (3.16) reduces to

$$\frac{p_{1l}}{p_{11}} \le min_i \left\{ \frac{p_{il}}{p_{i1}}, \frac{1-p_{il}}{1-p_{i1}} \right\}. \tag{3.17}$$

We can summarize the above discussion with the following theorem:

Theorem 3.3. *Let D be a given design whose transition probability matrix P satisfies conditions (3.17), then we can find a better design D_*, with transition probability matrix P_*, which provides at least the same level of respondent protection as D, and simultaneously yields more information than D .*

One consequence of theorem 3.3 is that if the original transition probability matrix P satisfies conditions (3.17) then it is not admissible. In examining conditions (3.17) and attempting to characterize P that satisfy them, we note that (3.17) imposes the

following restrictions

- The ratio $\frac{p_{1l}}{p_{11}}$ needs to be sufficiently small. That is, matrix P should be such that p_{11} is large (but less than 1) and p_{1l} ($l \geq 2$) should be small. Having a large p_{11} makes sense since we are typically interested in introducing minimal random noise that is necessary to protect respondents' privacy but simultaneously preserve much of the statistical information.

- For $l > 2$, if at least one p_{il} is zero, $\frac{p_{1l}}{p_{11}}$ will have to be zero to satisfy (3.17).

Example 3.3. The $(3 \to 3)$ RR design D with the following transition probability matrix

$$P = \begin{pmatrix} 0.8 & 0.2 & 0.1 \\ 0.1 & 0.7 & 0.3 \\ 0.1 & 0.1 & 0.6 \end{pmatrix}$$

satisfies (3.17), since

$$\frac{p_{12}}{p_{11}} = \frac{2}{8} \leq min\left\{\frac{p_{22}}{p_{21}}, \frac{1-p_{22}}{1-p_{21}}, \frac{p_{32}}{p_{31}}, \frac{1-p_{32}}{1-p_{31}}\right\} = min\left\{\frac{7}{1}, \frac{3}{9}, \frac{1}{1}, \frac{1}{1}\right\} = \frac{3}{9} \quad \text{and}$$
$$\frac{p_{13}}{p_{11}} = \frac{1}{8} \leq min\left\{\frac{p_{23}}{p_{21}}, \frac{1-p_{23}}{1-p_{21}}, \frac{p_{33}}{p_{31}}, \frac{1-p_{33}}{1-p_{31}}\right\} = min\left\{\frac{3}{1}, \frac{7}{9}, \frac{6}{1}, \frac{4}{9}\right\} = \frac{4}{9}.$$

Thus, D is not admissible and is dominated by D_* whose transition probability matrix is

$$P_* = \begin{pmatrix} 1 & \frac{2}{8} & \frac{1}{8} \\ 0 & \frac{6}{8} & 0 \\ 0 & 0 & \frac{7}{8} \end{pmatrix}.$$

Example 3.4. Consider another $(3 \to 3)$ RR design with the following transition probability matrix

$$P = \begin{pmatrix} 0.7 & 0.1 & 0.1 \\ 0.2 & 0.9 & 0.1 \\ 0.1 & 0 & 0.8 \end{pmatrix}.$$

This matrix does not satisfy (3.17), since

$$\frac{p_{12}}{p_{11}} = \frac{1}{7} \nleq min\left\{\frac{p_{22}}{p_{21}}, \frac{1-p_{22}}{1-p_{21}}, \frac{p_{32}}{p_{31}}, \frac{1-p_{32}}{1-p_{31}}\right\} = min\left\{\frac{9}{2}, \frac{1}{8}, 0, \frac{10}{9}\right\} = 0.$$

So we cannot construct a better design using Theorem 3.3.

Chapter 4

Post-Randomization Technique for Limiting Statistical Disclosure

4.1 Introduction

The Post Randomization Method (PRAM), introduced by Kooiman et al. (1997) and further discussed by Gouweleeuw et al. (1998), Willenborg and De Waal (2001) and others, is an important disclosure control technique for categorical variables. The PRAM stochastically transforms each record in a data set using pre-selected probabilities. This deliberate misclassification of the original responses introduces uncertainty about the true category of any respondent. On the other hand, since the misclassification probabilities are known, valid statistical inferences can be derived from the PRAMed data with suitable adjustment of standard methods. In practice, a data agency should release the PRAMed data along with the transition probability matrix P to the public, so that users can derive valid statistical inferences from the released data.

As noted by Gouweleeuw et al. (1998) and Van den Hout and Van der Heijden (2002), mathematically, the PRAM is similar to randomized response (RR) surveys (e.g., Warner, 1965; Chaudhuri and Mukerjee, 1988; Nayak, 1994). So, many concepts, theoretical results and methods developed for RR surveys can also be used in the context of PRAM. Both methods are concerned with respondents' privacy protection and statistical efficiency. One difference is that in RR surveys, the responder randomizes his

response at data gathering stage, whereas in PRAM, randomization is carried out by the surveyor after the data are collected.

In Section 4.2, we describe several variations of the PRAM procedure and give examples. We demonstrate that any PRAM procedure can be regarded as a PRAMing the cross-classification of all the variables in the microdata set. In Section 4.3, we discuss the connection of PRAM and RR and note that the estimators originally developed in the previous chapter, for RR of polychotomous populations, can be used for PRAM. Next, in Section 4.4, we discuss a special case of PRAM known as invariant PRAM and introduce the notion of a strongly invariant PRAM. We review methods for constructing invariant PRAM matrices in Section 4.5, and we clarify certain perceptions of invariant PRAM that are not fully justified. We also discuss estimation from an invariantly PRAMed data in Section 4.6. Finally in Section 4.7, we examine the effectiveness of PRAM for limiting statistical disclosure.

4.2 The PRAM Procedure

Let X be a categorical variable with categories $c_1, \cdots, c_k$. The basic ideas of PRAM are: (i) select a transition probability matrix, also known as PRAM matrix, $P = ((p_{ij}))$, where $\sum_i p_{ij} = 1$ for $j = 1, \cdots, k$, and then (ii) randomize each record on X in such a way that if an original category is c_j, $(j = 1, \cdots, k)$, then it changes to c_i with probability p_{ij}. The randomization step is performed for each record in the data set, independently of the other records. We shall denote the transformed variable by Z, in which case $P(Z = c_i | X = c_j) = p_{ij}$.

Operationally, PRAM can be applied in several ways. For example, one may

1. PRAM some of the variables independently (see Example 4.1), or

2. PRAM certain variables conditionally within each category of other variables (see

Example 4.3).

Example 4.1. Consider Case 1 and suppose we have a data set containing two variables (X, Y) and we PRAM only X, where X has two categories and Y has three categories. And, let $\widetilde{X}$ be the variable resulting from PRAMing X. Let x_1 and x_2 denote the categories of X and $\widetilde{X}$; and, let y_1, y_2 and y_3 denote the categories of Y. If we define $\alpha_{ij} = P(\widetilde{X} = i|X = j)$ so that $\sum_{i=1}^{2} \alpha_{ij} = 1$, the transition probability matrix of X to $\widetilde{X}$ is given by

$$P_1 = \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{pmatrix}.$$

Since only X is being PRAMed, this is equivalent to PRAMing a new variable (with 2 x 3=6 categories) obtained by cross classifying X and Y. The transition probability matrix for this new variable is given by

$$P = \begin{pmatrix} \alpha_{11} & 0 & 0 & \alpha_{12} & 0 & 0 \\ 0 & \alpha_{11} & 0 & 0 & \alpha_{12} & 0 \\ 0 & 0 & \alpha_{11} & 0 & 0 & \alpha_{12} \\ \alpha_{21} & 0 & 0 & \alpha_{22} & 0 & 0 \\ 0 & \alpha_{21} & 0 & 0 & \alpha_{22} & 0 \\ 0 & 0 & \alpha_{21} & 0 & 0 & \alpha_{22} \end{pmatrix}.$$

Thus, we can write a single transition probability matrix by cross classifying variables X and Y as $P = P_1 \otimes I_3$, where I_3 is the identity matrix of dimension 3.

Example 4.2. In a more general case of the previous example, the PRAM procedure can be applied independently to more than one variable, where there is a transition probability matrix for each variable. If in addition to PRAMing X as in above, we also PRAMed Y using

$$P_2 = \begin{pmatrix} \beta_{11} & \beta_{12} & \beta_{13} \\ \beta_{21} & \beta_{22} & \beta_{23} \\ \beta_{31} & \beta_{32} & \beta_{33} \end{pmatrix}.$$

Then, this is equivalent to PRAMing a new variable obtained by cross classifying X and Y, whose transition probability matrix is given by

$$P_1 \otimes P_2 = \begin{pmatrix} \alpha_{11}\beta_{11} & \alpha_{11}\beta_{12} & \alpha_{11}\beta_{13} & \alpha_{12}\beta_{11} & \alpha_{12}\beta_{12} & \alpha_{12}\beta_{13} \\ \alpha_{11}\beta_{21} & \alpha_{11}\beta_{22} & \alpha_{11}\beta_{23} & \alpha_{12}\beta_{21} & \alpha_{12}\beta_{22} & \alpha_{12}\beta_{23} \\ \alpha_{11}\beta_{31} & \alpha_{11}\beta_{32} & \alpha_{11}\beta_{33} & \alpha_{12}\beta_{31} & \alpha_{12}\beta_{32} & \alpha_{12}\beta_{33} \\ \alpha_{21}\beta_{11} & \alpha_{21}\beta_{12} & \alpha_{21}\beta_{13} & \alpha_{22}\beta_{11} & \alpha_{22}\beta_{12} & \alpha_{22}\beta_{13} \\ \alpha_{21}\beta_{21} & \alpha_{21}\beta_{22} & \alpha_{21}\beta_{23} & \alpha_{22}\beta_{21} & \alpha_{22}\beta_{22} & \alpha_{22}\beta_{23} \\ \alpha_{21}\beta_{31} & \alpha_{21}\beta_{32} & \alpha_{21}\beta_{33} & \alpha_{22}\beta_{31} & \alpha_{22}\beta_{32} & \alpha_{22}\beta_{33} \end{pmatrix}.$$

By the independence property, we can also estimate joint probabilities of X and Y by using $P_1 \otimes P_2$ for analysis of the cross-classified variable. Also the transition probabilities of the cross-classified variable can be decomposed as

$$P(\widetilde{X} = x_i, \widetilde{Y} = y_i, |X = x_j, Y = y_j) = P(\widetilde{X} = x_i | X = x_j) P(\widetilde{Y} = y_i | Y = y_j).$$

We noted a similar result for RR designs in Example 3.2 from Chapter 3.

Example 4.3. In Case 2, let x_1 and x_2 denote the categories of X (which is either one variable or a cross-classification of several variables); and, let y_1, y_2 and y_3 denote the categories of Y. Let the transition probability matrix of X to $\widetilde{X}$ when $Y = y_1$ be given by P_1. Similarly let the transition probability matrices of X to $\widetilde{X}$ when $Y = y_2$ and when $Y = y_3$ be given

$$P_3 = \begin{pmatrix} \delta_{11} & \delta_{12} \\ \delta_{21} & \delta_{22} \end{pmatrix}$$

and

$$P_4 = \begin{pmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{pmatrix}$$

respectively. Then if we PRAM the categories of X within each category of Y, the

transition matrix of a new variable W, with categories x_1y_1,x_2y_1,x_1y_2,x_2y_2, x_1y_3, x_2y_3, derived from crossclassifying X and Y is given by

$$P = \begin{pmatrix} \alpha_{11} & 0 & 0 & \alpha_{12} & 0 & 0 \\ 0 & \delta_{11} & 0 & 0 & \delta_{12} & 0 \\ 0 & 0 & \gamma_{11} & 0 & 0 & \gamma_{12} \\ \alpha_{21} & 0 & 0 & \alpha_{22} & 0 & 0 \\ 0 & \delta_{21} & 0 & 0 & \delta_{22} & 0 \\ 0 & 0 & \gamma_{21} & 0 & 0 & \gamma_{22} \end{pmatrix}. \tag{4.1}$$

In summary, the PRAM procedure can be applied to more than one categorical variable, independently or jointly on the cross-classification of all variables (see, Gouweleeuw et al., 1998). Conceptually, any PRAM procedure can be regarded as being applied to the combined variable created by cross-classifying all variables. The structure of the transition probability matrix is determined by the specific PRAMing procedure. If the variables are PRAMed independently, the transition probability matrix for the combined variable can be expressed as a Kronecker product of the transition probability matrices for the individual variables. For estimating the joint probabilities, it is both convenient and appropriate to think in terms of the PRAM procedure for the cross-classified variable.

4.3 Estimation from PRAM

Let $\pi_i (i = 1, \cdots, k,)$ denote the proportion of population units with $X = c_i$ and let $\boldsymbol{\pi} = (\pi_1, \cdots, \pi_k)'$. Let n denote the sample size and N_i denote the sample frequency of category i. Most papers on PRAM assume multinomial sampling, i.e., the original data are collected by random sampling from an infinite population or by simple random sampling with replacement (SRSWR) if the population is finite. Under multinomial sampling, $\boldsymbol{N} = (N_1, \cdots, N_k) \sim Mult(n, \pi)$, but note that $N_1, \cdots, N_k$ are not

computable from the PRAMed data. Let, M_i denote the frequency of category i in the PRAMed data set, $\lambda_i = P(Z = c_i), i = 1, \cdots, k$ and $\boldsymbol{\lambda} = (\lambda_1, \cdots, \lambda_k)'$. Then, $\boldsymbol{M} = (M_1, \cdots, M_k) \sim Mult(n, \lambda)$, where

$$\boldsymbol{\lambda} = P\boldsymbol{\pi}, \tag{4.2}$$

and it follows easily that if $\tilde{\boldsymbol{\lambda}}$ is an unbiased estimator of $\boldsymbol{\lambda}$ and P is nonsingular, then $\tilde{\boldsymbol{\pi}} = P^{-1}\tilde{\boldsymbol{\lambda}}$ is an unbiased estimator of $\boldsymbol{\pi}$. The MLE (and UMVUE) of $\boldsymbol{\lambda}$ is $\hat{\boldsymbol{\lambda}} = \boldsymbol{M}/n$ and it leads to the following estimator of $\boldsymbol{\pi}$:

$$\hat{\boldsymbol{\pi}} = P^{-1}\hat{\boldsymbol{\lambda}} = P^{-1}\left(\boldsymbol{M}/n\right). \tag{4.3}$$

Obviously, $\hat{\pi}$ is an unbiased estimator of π and it can be seen that

$$Var(\hat{\boldsymbol{\pi}}) = \frac{\left(D_{\boldsymbol{\pi}} - \boldsymbol{\pi}\boldsymbol{\pi}'\right)}{n} + \frac{\left[P^{-1}D_{\boldsymbol{\lambda}}(P^{-1})' - D_{\boldsymbol{\pi}}\right]}{n} \tag{4.4}$$

where D_π is a diagonal matrix with diagonal elements being $\pi_1, \cdots, \pi_k$ and D_λ is defined similarly (see Chaudhuri and Mukerjee, 1988, p. 43). The first term on the right side of (4.4) is the variance under no randomization and the last term is the additional variance due to PRAMing.

Note that (4.4) is the same as (3.5) from Chapter 3. Other estimators, under infinite and finite population settings, developed for RR in sections 3.3 and 3.4 also apply to PRAM.

4.4 Invariant PRAM

Without any PRAMing, i.e., based on the original data, the MLE (and the UMVUE) of π is $\hat{\pi}_0 = f = \boldsymbol{N}/n$ and the estimator $\hat{\pi}$ in (4.3) can be thought of an unbiasedly

recovered version of $\hat{\pi}_0$, as $E[P^{-1}\boldsymbol{M}|\boldsymbol{N}] = \boldsymbol{N}$. However, if $E[\boldsymbol{M}|\boldsymbol{N}] = \boldsymbol{N}$, i.e.,

$$P\boldsymbol{N} = \boldsymbol{N} \quad \text{or equivalently} \quad P\hat{\boldsymbol{\pi}}_0 = \hat{\boldsymbol{\pi}}_0 \tag{4.5}$$

then $\boldsymbol{M}/n$ is also an unbiased estimator of π, which does not involve P or its inverse. Also, $\boldsymbol{M}/n$ is always a probability vector, while $\hat{\pi}$ in (4.3) may not be so. These facts motivated Gouweleeuw et al. (1998) to call such procedures invariant PRAMs. Specifically, a PRAM is called an invariant PRAM if P satisfies (4.5). We note that (4.5) ensures invariance of a specific estimator, viz., unbiasedness of the observed relative frequency vector ($\boldsymbol{M}/n$) for estimating π, but not necessarily of all estimators. In particular, $\boldsymbol{N}$ and $\boldsymbol{M}$ may not have a common distribution (unless $P = I$), and consequently (i) the covariance matrices of ($\boldsymbol{N}/n$) and ($\boldsymbol{M}/n$) may be different and (ii) unbiased estimation of some functions of π (e.g., π_1^2) may not be possible if P is not given.

For broader validity of standard inferences, one may wish to choose P suitably so that all inferential methods remain invariant under PRAMing of the data using P, i.e., for making inferences about π, one can legitimately treat the PRAMed data set as the original data set. Then, an user would not need to know or use the transition probability matrix P and would be able to simply use any procedure that is valid for the original data. This goal requires the frequency counts based on original and PRAMed data, respectively, i.e., $(N_1, \cdots, N_k)$ and $(M_1, \cdots, M_k)$, to have the same distribution, viz., $Mult(n, \pi)$. This condition holds if and only if P satisfies the condition

$$P\boldsymbol{\pi} = \boldsymbol{\pi}. \tag{4.6}$$

We shall call a PRAM a *strongly invariant* PRAM if P satisfies (4.6). For such a procedure, there would be no loss of statistical information and the PRAMed data could be analyzed without any adjustment for post-randomization, i.e., PRAMing would have

no effect on either data utility or data analysis. However, such a strong situation is usually unattainable in practice because typically π, being the population probability vector, is unknown and so (4.6) cannot be solved for P (except for $P = I$). One situation where the true π is known is when the data come from a census. Thus, a strongly invariant PRAM is applicable when releasing a random subset census data for a large population.

In Section 4.5, we review two general classes of solutions of equations (4.5) and (4.6) and present a new approach. We may also note that if a data set contains several variables and they are PRAMed independently and invariantly, the marginal probabilities can be estimated from the PRAMed data without any further adjustment, but not the joint probabilities. This is because, independent invariant PRAMing of individual variables does not imply invariant PRAMing of the cross-classified variable, unless the variables are independently distributed. So, for estimating the joint probabilities, we would need to obtain the transition probability matrix for the combined variable and use its inverse. For a full invariant PRAM, one would need to apply an invariant PRAM to the combined variable, which is cumbersome when the total number of cells is large. As we discuss the next section, invariant PRAMing of a subset of variable within the cells of remaining variables may be a convenient approach.

4.5 Construction of Invariant PRAM

For constructing an invariant PRAM, one needs to solve (4.6) for P. The first relevant practical issue for a surveyor is to determine what to use in place of an unknown π in (4.6). Then, the surveyor can use various techniques to find an appropriate invariant PRAM matrix P based on the chosen π. Common methods for estimating π include:

- Prior estimates assuming a survey is performed periodically

- Estimate π from original data

To construct a strongly invariant PRAM, for a given π, one needs to solve (4.6) for P and choose a solution. Similarly, constructing an invariant PRAM requires solving (4.5). However, since (4.5) and (4.6) are very similar in nature, we shall discuss solutions of only (4.6). Note that for given π, (4.6) may have many solutions for P, the identity matrix being an obvious solution. Another solution is

$$P = \begin{pmatrix} \pi_1 & \pi_1 & \cdots & \pi_1 \\ \pi_2 & \pi_2 & \cdots & \pi_2 \\ \vdots & \vdots & \vdots & \vdots \\ \pi_k & \pi_k & \cdots & \pi_k \end{pmatrix}, \tag{4.7}$$

from which it can be seen easily that $P\pi = \pi$. In general, if P_1 and P_2 satisfy (4.6), then $aP_1 + (1-a)P_2$ also satisfies (4.6) for all $0 \leq a \leq 1$ and hence the solutions of (4.6) form a convex set.

Applying P on a microdata set to obtain an invariantly PRAMed dataset is akin to creating a synthetic dataset without any covariates. This approach for generating synthetic data was presented in a general framework by Fienberg et al. (1998). Their framework discusses generating synthetic datasets for categorical data from cumulative distribution functions derived from the original microdata set. The idea of creating synthetic microdata for public releases is appealing to statistical agencies. The proponents of this method claim that the released data is simulated data and not the actual data, but it preserves the analytical properties of the data. Currently, the most popular approach for generating synthetic data is via imputing microdata unit-level data using models fit with covariates from the original microdata. This method was proposed by Rubin (1993) and it uses appropriate estimation methods based on the concepts of multiple imputation (Rubin, 1987). Other papers on this topic include Reiter (2002) and

Raghunathan et al. (2003).

For generating other solutions (4.6), Gouweleeuw et al. (1998) presented two methods, which are briefly reviewed next.

Method 1

The first method is direct and involves setting up several simultaneous equations. If we write (4.6) in long form,

$$\sum_j p_{ij}\pi_j = \pi_i$$
$$\sum_i p_{ij} = 1$$

where $i = 1, \ldots, k$, $0 \leq p_{ij} \leq 1$, we can see that we have more unknown quantities than equations. This non-uniqueness property allows for flexibility in choosing P.

Let us look at the simplest case of $k = 2$, so that $\pi_2 = 1 - \pi_1$. Noting that

$$P = \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix}$$

and $\pi = (\pi_1, 1 - \pi_1)'$, here, (4.6) reduces to:

$$p_{11}\pi_1 + p_{12}(1 - \pi_1) = \pi_1, \quad \text{and}$$
$$p_{21}\pi_1 + p_{22}(1 - \pi_1) = (1 - \pi_1). \tag{4.8}$$

We also have:

$$p_{12} + p_{22} = 1 \quad \text{and} \quad p_{11} + p_{21} = 1$$

Note that (4.8) follows from other equations. Thus we have three independent equations with four unknowns which yields many solutions and, hence, flexibility in the choice of P. If we fix $p_{12} = \theta$, then $p_{11} = 1 - \frac{\theta(1-\pi_1)}{\pi_1}$, $p_{21} = \frac{\theta(1-\pi_1)}{\pi_1}$, $p_{22} = 1 - \theta$. Let categories 1 and 2 be the minority and majority categories respectively. We can interpret θ as the proportion of the n_2 cases of category 2 that is swapped to category 1. Likewise $p_{21} = \frac{\theta(1-\pi_1)}{\pi_1}$ is the proportion of n_1 category 1 cases that is swapped into category 2. In the next section we provide more guidance for choosing θ in a beneficial way, considering privacy protection.

The results above for $k = 2$ has been extended for $k > 2$ categories (Gouweleeuw et. al., 1998). Suppose all cell probabilities are positive and for notational simplicity assume that $0 < \pi_k \leq \pi_i$ for $i = 1, \ldots, k$. Then, let $p_{ii} = 1 - \theta(\frac{\pi_k}{\pi_i})$, and $p_{ij} = \theta[\frac{\pi_k}{(k-1)\pi_j}]$ if $i \neq j$, where $0 \leq \theta \leq 1$. The resultant P can be seen to satisfy (4.6) for all $0 \leq \theta \leq 1$. These solutions can be interpreted easily. Here, $P(Z \neq c_j | X = c_j) = \theta(\frac{\pi_k}{\pi_j}), j = 1, \cdots, k$. So, $P(Z \neq c_k | X = c_k) = \theta$, i.e., θ is the probability of changing the true response when it is c_k. If the true response is c_j and $j \neq k$, it changes to another category with probability $\theta(\frac{\pi_k}{\pi_j})$. Moreover, it follows easily that for $i \neq j, P(Z = c_i | X \neq Z) = 1/(k-1)$, which means that when a true response is changed, the perturbed value is selected at random form one of the remaining $(k-1)$ categories. We may note that this method, as stated above, requires the smallest cell probability to be positive, which may not hold in case of a cross-classified variable due to structural zeroes, i.e., theoretically empty cells. In such cases, one can apply the basic idea to the nonempty cells to obtain solutions of (4.6). Without loss of generality, suppose, $\pi_1 > \cdots > \pi_l > 0 = \pi_{l+1} = \cdots = \pi_k$. Then, taking $\pi_{ii} = 1, i = l+1, \cdots, k$ (and hence $p_{ij} = 0$ if $i = l+1, \cdots, k$ and $i \neq j$), $p_{ll} = 1 - \theta(\frac{\pi_l}{\pi_i}), i = 1, \cdots, l$, and $p_{ij} = \theta[\frac{\pi_l}{(l-1)\pi_j}]$ for $i \neq j$ and $i, j = 1, \cdots, l$, it can be seen that the resulting P satisfies (4.6) for all $0 \leq \theta \leq 1$. We can interpret θ here as the proportion of cases in the minority category, say k, that will be swapped to into

categories $1, \ldots, k-1$. Since p_{ij} is the proportion of n_j category j cases that is swapped into category i, $\theta = \sum_{i=1}^{k-1} p_{ik}$. We note that for a given π, the invariant transition matrix P, satisfying (4.6), is not unique since it could vary depending on the choice of θ.

Method 2

Another method for solving (4.6) is as follows (see De Waal et al., 1998; Willenborg and De Waal, 2001). Start with any transition probability matrix $R = ((r_{ij}))$ such that all components of $R\pi$ are positive. Let $q_{ij} = [r_{ji}\pi_i]/[\sum_l r_{jl}\pi_l]$ and $Q = ((q_{ij}))$. Then, it can be verified that $P = QR$ satisfies the equation $P\pi = \pi$. We can show that this procedure produces an invariant matrix as follows:

Starting with $r_{ij} = P(Z = i|X = j)$, where X is the original category and Z represents the PRAMed variable. The problem now reduces to finding Q such that $QR\pi = \pi$. Then, $P = QR$, the invariant matrix we are interested in, is invariant.

We derive Q as follows:

$$R\pi = \begin{pmatrix} r_{11} & r_{12} & \ldots & r_{1k} \\ r_{21} & r_{22} & \cdots & r_{2k} \\ \vdots & \vdots & \cdots & \vdots \\ r_{k1} & r_{k2} & \cdots & r_{kk} \end{pmatrix} \begin{pmatrix} \pi_1 \\ \pi_2 \\ \vdots \\ \pi_k \end{pmatrix} = \begin{pmatrix} \sum_l r_{1l}\pi_l \\ \sum_l r_{2l}\pi_l \\ \vdots \\ \sum_l r_{kl}\pi_l \end{pmatrix}$$

and

$$QR\pi = \begin{pmatrix} q_{11} & q_{12} & \ldots & q_{1k} \\ q_{21} & q_{22} & \cdots & q_{2k} \\ \vdots & \vdots & \cdots & \vdots \\ q_{k1} & q_{k2} & \cdots & q_{kk} \end{pmatrix} \begin{pmatrix} \sum_l r_{1l}\pi_l \\ \sum_l r_{2l}\pi_l \\ \vdots \\ \sum_l r_{kl}\pi_l \end{pmatrix} = \begin{pmatrix} \sum_j q_{1j} \sum_l r_{jl}\pi_l \\ \sum_j q_{2j} \sum_l r_{jl}\pi_l \\ \vdots \\ \sum_j q_{kj} \sum_l r_{jl}\pi_l \end{pmatrix}$$

If we set $q_{ij} = \frac{\phi_j \pi_i}{\sum_l r_{jl}\pi_l}$, where $0 \leq \phi_j \leq 1$ and $\sum_j \phi_j = 1$, we get $QR\pi = \pi$. So,

$\sum_j q_{ij} \sum_l r_{jl}\pi_l = \sum_j \frac{\phi_j \pi_i}{\sum_j r_{jl}\pi_l} \sum_l r_{jl}\pi_l = \sum_j \phi_j \pi_i = \pi_i$. De Wolf et al., 1997 presented this result in a elegant form by choosing $\phi_j = r_{ji}$, and since $\sum_l r_{jl}\pi_l = \sum_i r_{ji}\pi_i$, $q_{ij} = \frac{r_{ji}\pi_i}{\sum_i r_{ji}\pi_i} = P(X = i|Z = j)$. Thus, they chose Q to be a matrix of posterior probabilities. This method can be viewed as a two stage procedure. The preceeding discussion shows that at the first stage an arbitrary transition probability matrix, R, can be applied to the original microdata set. Then, a suitable transition probability matrix Q can be applied at the second stage, to make the final outcome an invariant PRAM.

We note that here R need not be a square matrix, i.e., for any transition probability matrix R of order $m \times k$ with all components of $R\pi$ being positive, the final matrix P is a strongly invariant PRAM matrix. However, P would be singular, if $m < k$ or $rank(R) < k$.

When there are several variables, independent invariant PRAMing of the variables may not yield an invariant PRAM of the compounded variable as we demonstrate in below in Section 4.5.1. So, the two methods discussed above need to be applied to the cross-classification of all variables, which may be unwieldy when the total number of cells is large.

4.5.1 Invariantly PRAMing Several Variables Separately

Here, we shall examine some special cases or procedures under which PRAMing variables separately could yield an invariant PRAM of the compounded variable. Let X and Y be two categorical variables that are PRAMed invariantly and independently. We show below that the transition matrix P of a cross classified variable is invariant if the variables X and Y are themselves independent. In this case (of independence) we would not require P for estimating joint probabilities. Otherwise, P is needed.

Let π_{XY} represent a vector of joint probabilities of X and Y categories. Then, the

cross classified variable is invariant if (4.6) is satisfied, i.e., $P\pi_{XY} = \pi_{XY}$. Further, let P_X and P_Y be the transition probability matrices to be applied to variables X and Y respectively. If

$$P_X = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1k} \\ p_{21} & p_{22} & \cdots & p_{2k} \\ \vdots & \vdots & \cdots & \vdots \\ p_{k1} & p_{k2} & \cdots & p_{kk} \end{pmatrix}$$

then

$$P_X \otimes P_Y = \begin{pmatrix} p_{11}P_Y & p_{12}P_Y & \cdots & p_{1k}P_Y \\ p_{21}P_Y & p_{22}P_Y & \cdots & p_{2k}P_Y \\ \vdots & \vdots & \cdots & \vdots \\ p_{k1}P_Y & p_{k2}P_Y & \cdots & p_{kk}P_Y \end{pmatrix}$$

We note that, in general, $(P_X \otimes P_Y)\pi_{XY} \neq \pi_{XY}$. One exception is when X and Y are independent, so that $\pi_{XY} = \pi_X \otimes \pi_Y$. In this case, $(P_X \otimes P_Y)\pi_{XY} = (P_X \otimes P_Y)(\pi_X \otimes \pi_Y) = (P_X \otimes \pi_X)(P_Y \otimes \pi_Y) = (\pi_X \otimes \pi_Y) = \pi_{XY}$. Thus if the original variables themselves are independent, invariantly PRAMing each variable separately yields an invariant PRAM of the compounded variable. However, if X and Y are not independent then $P_X \otimes P_Y$ may not be an invariant PRAM of the cross classified variable.

In the next subsection, we shall present a new approach procedure where invariantly PRAMing X and Y separately will yield an invariant PRAM of the two variables jointly even if the variables are not independent.

4.5.2 A New Approach for Designing Invariant PRAM

The following results provide another mechanism for devising strongly invariant PRAMs of several variables jointly.

Theorem 4.1. *Consider two categorical variables X and Y and suppose X is strongly invariantly PRAMed within each category of Y, generating a transformed variable Z*

where Y is kept unchanged. Then, the stochastic transformation from (X, Y) to (Z, Y) is a strongly invariant PRAM of the cross classification of X and Y.

Proof. Suppose X and Z have k categories denoted $c_1, \cdots, c_k$ and Y has m categories labeled $d_1, \cdots, d_m$. Since X is strongly invariantly PRAMed within each category of Y, we have $P(X = c_i|Y = d_j) = P(Z = c_i|Y = d_j)$ for all $i = 1, \cdots, k, j = 1, \cdots, m$. Then considering the joint probabilities, we get

$$\begin{aligned} P(Z = c_i, Y = d_j) &= P(Y = d_j)P(Z = c_i|Y = d_j) \\ &= P(Y = d_j)P(X = c_i|Y = d_j) \\ &= P(X = c_i, Y = d_j) \end{aligned}$$

for all $i = 1, \cdots, k$ and $j = 1, \cdots, m$, which completes the proof. □

Similarly, strongly invariant PRAMing of Y within the categories of X also yields a strongly invariant PRAM of (X, Y). Also, if P_1 and P_2 are two strongly invariant PRAMs, then P_1P_2 is also a strongly invariant PRAM. Using these we get the following.

Theorem 4.2. *Consider two variables X and Y and suppose X is transformed to Z as in Theorem 4.1, creating (Z, Y). Suppose then Y is strongly invariantly PRAMed within each category of Z. Let W denote the transformed version of Y. Then, the stochastic transformation from (X, Y) to (Z, W) is a strongly invariant PRAM.*

Remark 4.1. We note that the process in theorems 4.1 and 4.2 can be repeated several times. For instance, after several applications of the procedures to (X, Y), we obtain the following stochastic transformations:

$$(X, Y) \to (Z, Y) \to (Z, W) \to (Z^*, W) \to (Z^*, W^*) \to (U, V) \quad \text{and so on.}$$

At every stage we would still have an invariant PRAM. This is a mathematical result,

but in practice, we do not see a need for repeating this step more than twice.

Theorem 4.1 would suffice in most applications. The theorem also reduces constructing an invariant PRAM with many cells to construction of several invariant PRAMs, each with a much smaller number of cells. Both X and Y may be compound variables. So, when there are several variables in a data set, each variable must be a part of either X or of Y.

In some applications of Theorem 4.1, one does not want the Y variable to be perturbed, so that there is no information loss for the marginal distribution of Y. In such situations, we can think of Y as a control variable and X as a randomization variable. This type of requirement should be taken into account when assigning each variable to either X or Y. Shlomo and De Waal (2008) discuss an evaluation of PRAM while preserving edit constraints, such as verifying that only people of legal age are married. Their evaluation data set is drawn from the Israel Census sample data which include two categorical variables *age* (X) and *marital status* (Y). They are interested in methods for implementing PRAM that will place controls on the perturbation process such that edit failures and information loss are both minimized. If no controls are incorporated into the PRAMing process, edit failures resulting in inconsistent combinations such as married small children may occur. The controls in the PRAM process are defined by categorical variable, Y, whose cells or categories define groupings within which PRAMing is allowed. That is, *age* is PRAMed within *marital status,* which is the control variable. Specifically, the X is invariantly PRAMed within each category d_j of Y, $j = 1, \cdots, m$, using an invariant transition probability matrix P_{d_j}. Each P_{d_j} is then placed in the main

diagonal of the overall transition probability matrix P for the data set:

$$P = \begin{pmatrix} P_{d_1} & 0 & \cdots & 0 \\ 0 & P_{d_2} & \cdots & 0 \\ \vdots & 0 & \ddots & \vdots \\ 0 & 0 & \cdots & P_{d_m} \end{pmatrix}.$$

In practice, there may be many edit constraints and it is often difficult to address all of them.

4.6 Estimation from Invariant PRAM

As noted in Section 4.4, under multinomial sampling all analytical methods for the original data are equally valid for the perturbed data when a strongly invariant PRAM, as defined by (4.6) is used. However, a strongly invariant PRAM cannot be constructed in most practical situations as π is unknown. So, here we shall discuss estimation based on invariantly PRAM data, as defined by (4.5). Let $P = [P_1 : \cdots : P_k]$ be an invariant PRAM matrix, which depends on $\boldsymbol{N}$ and hence is a random matrix. Also note that (4.5) implies that

$$\sum_{i=1}^{k} N_i P_i = \boldsymbol{N} \tag{4.9}$$

From (4.9), note that if $N_i = 0$ (for any i), then P_i can be arbitrary and p_{ij} must be 0 for all j with $N_j \neq 0$, in which case $M_i = 0$. This implies that if a cell has zero frequency in the original data, it will also have zero frequency in any invariantly PRAMed data. Let F_{ij} denote the number of units with original category i and perturbed category j, and let $\boldsymbol{F_i} = (F_{i1}, \cdots, F_{ik})$. Then, $\boldsymbol{M} = \sum_i^k \boldsymbol{F_i}$ and given $\boldsymbol{N}$, $\boldsymbol{F_i} \sim Mult(N_i, P_i), i = 1, \cdots, k$ and they are independent. Thus, the distribution of the frequency vector ($\boldsymbol{M}$) based on the perturbed data, is a mixture of multinomial distributions. Then, letting $\hat{\boldsymbol{\pi}}_* = \boldsymbol{M}/n$,

we get

$$E(\hat{\pi}_*|\boldsymbol{N}) = \frac{1}{n}\sum_{i=1}^{k} N_i P_i = \frac{1}{n}\boldsymbol{N} = \hat{\boldsymbol{\pi}}_0 \tag{4.10}$$

by (4.9), and

$$V(\hat{\pi}_*|\boldsymbol{N}) = \frac{1}{n^2}\sum_{j=1}^{k} N_i[D_{P_i} - P_i P_i'] = \frac{1}{n}[D_{\hat{\pi}_0} - \sum_{i=1}^{k}(\frac{N_i}{n})P_i P_i'], \tag{4.11}$$

Under multinomial sampling, i.e., when $\boldsymbol{N} \sim Mult(n, \pi)$, it follows that $E(\hat{\pi}_*) = \pi$, so that $\hat{\pi}_*$ is an unbiased estimate of π, and

$$\begin{aligned} V(\hat{\pi}_*) &= VE(\hat{\pi}_*|N) + EV(\hat{\pi}_*|N) \\ &= V(\hat{\pi}_0) + \frac{1}{n}[D_\pi - E\left\{\sum_{i=1}^{k}(\frac{N_i}{n})P_i P_i'\right\}], \end{aligned} \tag{4.12}$$

where $V(\hat{\pi}_0) = [D_\pi - \pi\pi']/n$. Thus, the relative frequency vector $\hat{\pi}_*$ from invariantly PRAMed data is an unbiased estimator of π and the second term on the right side of (4.12) is the variance inflation matrix, which is different from the second term on the right side of (4.4).

Estimation of $V(\hat{\pi}_*)$ poses certain challenges. First, the transition probability matrix P depends on $\boldsymbol{N}$, through (4.5), but it is not a function of $\boldsymbol{N}$, and consequently the expectation in (4.12) cannot be calculated unless the specific algorithm for obtaining P from (4.5) is given. If P is reported along with the perturbed data, one may estimate $V(\hat{\pi}_*)$ by

$$\hat{V}(\hat{\pi}_*) = \frac{1}{n}[D_{\hat{\pi}_*} - \hat{\pi}_*\hat{\pi}_*'] + \frac{1}{n}[D_{\hat{\pi}_*} - \sum_{i=1}^{k}(\frac{N_i}{n})P_i P_i'].$$

However, the reporting of P is problematic from disclosure perspective. As Gouweleeuw et al. (1998) noted, $\boldsymbol{N}$ is an eigenvector of P corresponding to the eigenvalue 1 and one might be able to deduce the original frequencies from P without any error. In the following we present bounds on the variance of any linear combination of $\hat{\pi}_*$ and also

provide some suggestions about the choice of P.

A natural estimator of a linear combination $a'\pi$ of π is $a'\hat{\pi}_*$ and its variance is

$$V(a^{'}\hat{\pi}_*) = \frac{1}{n}a^{'}[2D_\pi - \pi\pi']a - \frac{1}{n^2}E\left\{\sum_{i=1}^{k} N_i \left(a^{'}P_i\right)^2\right\}. \tag{4.13}$$

Now using (4.5) and the Cauchy-Schwarz inequality, $(x'y)^2 \leq (x'Ax)(y'A^{-1}y)$, with $x' = (N_1, \cdots, N_k), y' = (a'P_1, \cdots, a'P_k)$ and $A = diag(1/N_1, \cdots, 1/N_k)$, we get

$$\begin{aligned}[\sum_{i=1}^{k} N_i(a'P_i)]^2 &\leq \left(\sum_{i=1}^{k} N_i\right)[\sum_{i=1}^{k} N_i(a'P_i)^2] \\ &= n[\sum_{i=1}^{k} N_i(a'P_i)^2]. \end{aligned} \tag{4.14}$$

We note that

$$\begin{aligned}[\sum_{i=1}^{k} N_i(a'P_i)]^2 &= (a'\sum_{i=1}^{k} N_iP_i)^2 \\ &= (a'\boldsymbol{N})^2 \\ &= (a'\boldsymbol{N}\boldsymbol{N}^{'}a) \end{aligned} \tag{4.15}$$

Using (4.15) in (4.14), taking expectations and dividing by n^3, we get

$$\begin{aligned}\frac{1}{n^2}E\left\{\sum_{i=1}^{k} N_i(a'P_i)^2\right\} &\geq \frac{1}{n}(a'\frac{\boldsymbol{N}}{n}\frac{\boldsymbol{N}^{'}}{n}a) \\ &= \frac{1}{n}a^{'}[E(\hat{\pi}_0\hat{\pi}_0^{'})]a \\ &= \frac{1}{n}a^{'}[V(\hat{\pi}_0) + E(\hat{\pi}_0)E(\hat{\pi}_0^{'})]a \\ &= \frac{1}{n}a^{'}\left[\left(\frac{D_\pi - \pi\pi'}{n}\right) + \pi\pi^{'}\right]a. \end{aligned} \tag{4.16}$$

Then, using (4.16) in (4.13), we get

$$V(a^{'}\hat{\pi}_{*}) \leq (2-\frac{1}{n})a^{'}\left[\frac{D_{\pi}-\pi\pi'}{n}\right]a. \tag{4.17}$$

The upper bound in (4.17) can be estimated by replacing π by $\hat{\pi}_{*}$. Also note that the upper bound equals $(2-1/n)V(a'\hat{\pi}_0) \approx 2V(a'\hat{\pi}_0)$ for large n. An obvious lower bound for $V(a'\hat{\pi}_{*})$ is $V(a'\hat{\pi}_0) = a'[\{D_{\pi}-\pi\pi'\}/n]a$. Thus, if an invariant PRAM is used and the PRAM matrix is not released, a user can estimate any linear combination $a'\pi$, without any adjustment, and lower and upper bounds for its variance.

Given a, $V(a'\hat{\pi}_{*}) = V(a'\hat{\pi}_0)$ or equivalently $V(a'\boldsymbol{M}) = V(a'\boldsymbol{N})$ if and only if $a'\boldsymbol{M} = a'\boldsymbol{N}$ with probability 1. We can see this from the fact that $V(a'\boldsymbol{M}) = E[V(a'\boldsymbol{M}|\boldsymbol{N})] + V[E(a'\boldsymbol{M}|\boldsymbol{N})] = E[V(a'\boldsymbol{M}|\boldsymbol{N})] + V(a'\boldsymbol{N})$. But, $E[V(a'\boldsymbol{M}|\boldsymbol{N})] = 0$ if and only if $a'\boldsymbol{M}|\boldsymbol{N}$ is degenerate for $\boldsymbol{N}$. This implies that given $\boldsymbol{N}$, $a^{'}\boldsymbol{M} = constant(=\text{c})$ $w.p.$ 1, and $E(a^{'}\boldsymbol{M}) = a^{'}E(\boldsymbol{M}) = c$ or $a^{'}\boldsymbol{N} = c$. That is, for each $\boldsymbol{N}$, $a^{'}\boldsymbol{M} = a^{'}\boldsymbol{N}$ $w.p.1$ or $a^{'}(\boldsymbol{M}-\boldsymbol{N}) = 0$ $w.p.1$.

Let $\boldsymbol{e}^{'} = (e_1, \cdots, e_n)^{'} = (\boldsymbol{M}-\boldsymbol{N})$, then $e_i = 0, \pm 1, \pm 2, \ldots$ We want to make sure that for certain marginal totals or total frequencies of a subgroup, $a^{'}\boldsymbol{M} = a^{'}\boldsymbol{N}$. That is, $a^{'}(\boldsymbol{M}-\boldsymbol{N}) = \sum_{i=1}^{dj} e_i = 0$ for each subgroup d_j $w.p.1$, and consequently $\sum_{i=1}^{n} e_i = 0$ where $\sum_{j=1}^{k} d_j = n$. We can achieve this by choosing an invariant PRAM matrix of the following structure

$$P = \begin{pmatrix} D_1 & 0 & \cdots & 0 \\ 0 & D_2 & \cdots & 0 \\ \vdots & 0 & \ddots & \vdots \\ 0 & 0 & \cdots & D_k \end{pmatrix} \tag{4.18}$$

where D_j is itself an invariant PRAM matrix for the records in subgroup d_j. If we set $b_1^{'} = (1, \cdots, 1, 0, \cdots, 0)$ where the first d_1 elements are 1's, and set

$b_2' = (0, \cdots, 0, 1, \cdots, 1, 0, \cdots, 0)$ where the first d_1 elements are 0's and the next d_2 elements are 1's, and so on. Then, $V(b_1'e) = 0 = V(b_2'e) = \ldots = V(b_k'e)$. Therefore, given constants $c_1, \cdots, c_k$, for any $a = c_1b_1 + c_2b_2 + \ldots + c_kb_k$, $V(a'e) = 0$ $w.p.1$. This implies that if P is block diagonal (possibly after reordering the cells), then the frequency of all cells within any given block remains unchanged under invariant PRAMing.

4.7 Disclosure Protection

As mentioned earlier in Section 4.1, the statistical inferential issues and methods are similar for RR surveys and PRAM. However, the physical processes of RR surveys and PRAM are different. In RR surveys, it is the responder who randomizes his or her true category during the data gathering stage; whereas, in PRAM, it is the data agency that randomizes the data after the data are collected. It is this difference in the RR and PRAM physical processes that in turn creates differences in their confidentiality or privacy protection issues.

In RR, the responders do not reveal their true category to the surveyor, so the surveyor knows who the respondents are but not their true categories. Also, the surveyor is typically honest in that he or she is only interested in statistical information and not in the private data of respondents. In RR context, the posterior probabilities are logical determinants of the level of respondents' privacy as discussed in sections 2.3 and 2.8.

In PRAM, the surveyor or statistical agency knows the true values of all respondents. The issue for the statistical agency is in the determination of what information from the survey can be released to the public without revealing private information about any survey participant.

4.7.1 Privacy Matters in Data Release

Statistical agencies disseminate information primarily using summary or tabular data products. For instance, the U.S. Census Bureau's American Factfinder is a web query application that mostly generates tables based on users' data requests. However, experienced users and researchers also need access to microdata sets to allow them to perform various types of data analysis that may be different from the analysis performed by the statistical agency in producing the tables in standard reports.

In the context of SDC, tabular data are classified into two categories: tables of frequency (or count) data and tables of magnitude data (Federal Committee on Statistical Methodology, 2005). Tables containing frequency data show the percent of the population with certain characteristics, or equivalently, the number in the population with certain characteristics. If a cell has only a few respondents whose characteristics are sufficiently distinctive, then it may be possible for an intruder to identify the individuals in the population. For tables of frequency data, SDC methods are applied to cells with fewer than a specified threshold number of respondents in order to minimize the risk that individuals can be identified from the microdata set. SDC may be applied before or after tabulation. Methods that can be applied after tabulation include controlled rounding and cell suppression. Methods can also be applied before tabulation using SDC techniques such as data perturbation and data swapping. In tables of magnitude data, the survey items published are aggregates of nonnegative reported values. The values reported by respondents may vary widely, with some extremely large values and some small values. Here, the disclosure issue relates to ensuring that an intruder cannot use the published cell totals and other publicly available data to estimate an individual respondent's value too accurately. SDC methods are applied to cells for which a linear sensitivity measure (Cox, 1981) indicates that some respondent's data may be estimated too closely. For tables of magnitude data, cell suppression is the most widely

used method after tabulation, while microdata protection techniques such as noise addition can be applied before tabulation. Although tabular data has many interesting disclosure issues as mentioned above, our main focus in this section is to examine the effectiveness of PRAM in limiting disclosure risk in a publicly released microdata set.

A microdata set consists of a series of records, each containing information about an individual unit. In its basic form, it may be represented as a two dimensional data matrix, where the rows correspond to the units and the columns to the variables. Typically, a microdata set contains records of n individuals (or sampling units) on k variables, some of which are confidential or sensitive. For privacy protection, all obvious direct identifiers, such as full name, passport number, and social security number, must be removed prior to data release.

However, simply removing direct identifiers is not sufficient for disclosure avoidance as it may be possible to identify the record of a survey participant by matching the values of variables such as gender, birth date and zip code, that are easily available from other sources. Such variables are generally non-sensitive and will be called key variables, following Bethlehem et al. (1990), and are useful attributes for identification. Broadly speaking, disclosure occurs when the values of some confidential variables for a specific unit can be predicted too accurately based on the values of the key variables. Therefore, the original microdata need to be further modified or masked for protection against disclosure. With PRAM, values of variables for many units of the original microdata are usually perturbed.

The most serious type of disclosure is identity disclosure, which happens when the intruder can identify the record of a unit, by matching his or her independent information with values of the key variables on the dataset. Identity disclosure reveals the values of all confidential variables of an identified subject. Identity disclosure is closely related to a unit being unique in the population (or sample) with respect to the key variables

(Bethlehem et al., 1990; Greenberg and Zayatz, 1992). If a unit is not unique in the sample, its identity cannot be ascertained with certainty. Measures of identity disclosure are typically based on the proportion of sample or population units whose records can be identified by matching on the key variables. Assessment of identity disclosure risk has been discussed by Bethlehem et al. (1990), Greenberg and Zayatz (1992), Willenborg and De Waal (2001), Skinner and Elliot (2002), Reiter (2005) and others. In particular, Skinner and Elliot (2002) discussed three measures and recommended using the proportion of correct matches among those population units which match a sample unique microdata record for assessing identity disclosure risk.

Another type of disclosure is predictive disclosure, which occurs if the released data enable the intruder to infer a respondent's value of a confidential variable with high accuracy. This type of disclosure is subject to some uncertainty. It is assumed that the intruder undertakes prediction by combining prior information with the released information in the microdata, and the prediction is usually represented by a prediction interval within which the intruder infers that the true value must lie. An extreme case of predictive disclosure is attribute disclosure, where a confidential variable value can be predicted completely accurately. Attribute disclosure can occur without identity disclosure. For example (cf., Nayak, 2008), suppose in a data set, a confidential variable has the same value, say X_0, for all respondents with a specific value, say Y_0, of the key variables. Then, if the original data are released, an intruder would know surely the confidential variable value (X_0) of any respondent with $Y = Y_0$, but not the identity of the respondent in the data set. The essence of predictive disclosure is: an intruder gaining too much new information about specific respondents from the released data. So, predictive disclosure depends not only on the released data set but also on the intruder's prior knowledge, and hence predictive disclosure risk can be assessed appropriately by comparing the intruder's knowledge before and after data release; see Duncan and Lam-

bert (1986, 1989), Lambert (1993) and Keller-McNulty et al. (2005) for measures of predictive disclosure risk, based on the predictive distribution of X given Y_0.

The main issue of disclosure avoidance for many statistical agencies like the U.S. Census Bureau is the identification of the record of one unit based on known values of certain key variables. A lot of research has focused on identity disclosure which is the worst kind of disclosure. Since the knowledge of values of key variables could vary substantially by intruder, this makes disclosure avoidance a difficult topic (cf., Lambert, 1993). This is because the intruder's information can occur in various forms associated with different disclosure scenarios (cf., Willenborg and De Waal, 2001).

The application of any SDC method will lead to some loss in data quality or statistical information even for a legitimate user of the masked microdata set. In SDC, there is no distinction between a legitimate user and an intruder, unlike in cryptography where a legitimate user is given an encryption key to unlock the masking done on a microdata set. In general, SDC methods may change the information in the microdata set causing some dilution and introduction of bias. While the user can still obtain unbiased estimates from a data set modified by PRAM, the additional noise in the data may introduce variance inflation in such estimates as discussed in Section 4.6. Thus, the statistical agency has a difficult problem of balancing privacy protection and maintaining accuracy of statistical inferences that can be made from the masked microdata set.

4.7.2 A Similarity Between RR disclosure and PRAM disclosure

We mentioned above that the privacy issues of RR and PRAM are different. In this subsection, we note a special case where their privacy issues are similar though not exactly the same. Suppose an intruder has already identified or can identify the unit of interest by matching on values of key variables (some of which could be quantitative) in the microdata. Then if the sensitive variable X is not perturbed, the intruder knows

the value of X. However, if X is PRAMed, the intruder knows both the identity of the unit and PRAMed X which we shall denote as Z. This disclosure scenario is the same as the one encountered in an RR design setup; in both cases an outsider (interviewer or intruder) knows the perturbed value of a variable for a specific unit and may use it to predict the true value. Unlike in RR surveys, the microdata set to be PRAMed may not have a sensitive group. However, we may still want to protect private information such as income categories of the survey participants.

As with RR, the issue is how accurately one can predict the value of X from the values of Z. The accuracy of a prediction is loosely related to how concentrated are the predictive distribution $P(X = i|Z = j), j = 1, \cdots, k$. If the predictive distribution is not concentrated or the largest predictive probability is not too large then predictive disclosure is limited. Therefore, posterior probabilities are relevant as a criterion for determining the level of disclosure.

Typically, $p_{ii} > 0.5$ and $P(X = i|Z = i) > P(X = j|Z = i)$ for all $j \neq i$, so we want to keep $P(X = i|Z = i), i = 1, \cdots, k$ small. Also, some intruders may not know or use conditional probabilities and will simply choose $X = i$ whenever $Z = i$. As Nayak (2008) suggested, we will keep

$$P(X = i|Z = i) < \beta. \tag{4.19}$$

Since $P(X = i|Z = i) = \frac{P(Z=i|X=i)P(X=i)}{P(Z=i)} = \frac{P(Z=i|X=i)\pi_i}{\lambda_i}$, (4.19) implies that $\frac{P(Z=i|X=i)\pi_i}{\lambda_i} < \beta$. Both $\pi_i = P(X = i)$ and $\lambda_i = P(Z = i)$ can be estimated from the original data. Thus, we can choose the p_{ii} element of PRAM transition probability matrix satisfying $p_{ii} = P(Z = i|X = i) < \frac{\beta\hat{\lambda}_i}{\hat{\pi}_i}$.

Another criterion suggested by Gouweleeuw et al (1998) is to keep the posterior odds

$$PO(i) = \frac{P(X = i|Z = i)}{P(X \neq i|Z = i)} \tag{4.20}$$

lower than a selected value, say ζ. We note that if $\zeta = \frac{\beta}{1-\beta}$ then (4.20) and (4.19) are equivalent.

4.7.3 One Approach to Choosing an Invariant PRAM

Here, we will continue with the scenario described in the previous section where the intruder already knows the identity of a unit. Typically, an intruder identifies the unit of interest by matching on key variables which can be either continuous or a cross classification of many variables on the microdata set, and he or she is trying to predict the value of a sensitive variable X. Since X is sensitive, we want to transform X into another categorical variable Z with same exact categories via an invariant PRAM transition probability matrix. So, the intruder will have to predict the value of X based on the value of Z, and the natural issue here is how we can design a PRAM matrix to make his prediction difficult.

Suppose an intruder picks a unit from the invariantly PRAMed dataset at random, observes its value of $(Z = x)$, and then attempts to predict the true category $(X = x)$ of that unit using a prediction rule based on only Z. In order to protect the privacy of the respondents, we would like to make the intruder's probability of correct prediction, P(correct prediction), as small as possible. This P(correct prediction) depends on $\pi = P(X = x)$, the prediction rule, and a fixed PRAM transition probability matrix. So, we want to choose a PRAM transition probability matrix that minimizes the P(correct prediction) for an optimum rule, as defined below.

Definition 4.1. For a given invariant PRAM transition probability matrix P, the optimum prediction rule maximizes the probability of correct prediction, D_P.

Definition 4.2. The optimum PRAM rule minimizes the probability of correct predic-

tion yielded by the optimum prediction rules. Thus, the optimum PRAM rule yields $min_P D_P$.

Case of a Binary Sensitive Variable

We present some results for a binary sensitive variable X. Suppose for a given value of Y we could have several units with different values of binary sensitive characteristic X. These sensitive X values are then transformed into Z values. The context under which we discuss disclosure or privacy protection is as follows: Suppose we apply an invariant PRAM procedure to the microdata set using Method 1 of Section 4.5. Recall that $\pi_1 = P(X = x_1)$ where $\pi_1 > \pi_2 = 1 - \pi_1$, and P is characterized by θ. So, the probability of a unit's X value being transformed from category 2 to category 1 is θ, and with probability $\frac{\theta(1-\pi_1)}{\pi_1}$ a unit's X value is transformed from category 1 to category 2. Suppose that the intruder chooses one unit at random and then predicts the value of X based on the transformed value Z using a nonrandomized or deterministic rule R such that all units with $Z = 0$ are assigned to the same X value. Likewise, all units with $Z = 1$ are assigned to a single, but different, X value. The four possible rules R_1, R_2, R_3 and R_4 for predicting X based on only Z are defined in Table 4.1 below.

Theorem 4.3. *An optimum invariant PRAM rule, indexed by θ, is such that*

$$\frac{1}{2} \leq \theta \leq min(\frac{\pi_1}{2(1-\pi_1)}, 1).$$

Proof. The probability of correct prediction $P(\text{correct prediction}) = P(C|X = 1)P(X = 1) + P(C|X = 0)P(X = 0)$. Using the nonrandomized rules, the $P(\text{correct prediction})$, $D_P = D_\theta$, is summarized in Table 4.1 below.

The maximum $P(\text{correct prediction})$ is given by

Table 4.1: Probabilities of Correct Prediction Under Four Nonrandomized Rules

Prediction Rule	Description: $(Z \to \hat{X})$	D_θ
R_1	$0 \to 0,\ 1 \to 1$	$\pi_1 p_{11} + (1-\pi_1)p_{22} = 1 - 2\theta(1-\pi_1)$
R_2	$0 \to 1,\ 1 \to 0$	$\pi_1 p_{21} + (1-\pi_1)p_{12} = 2\theta(1-\pi_1)$
R_3	$0 \to 0,\ 1 \to 0$	$\pi_1 p_{11} + \pi_1 p_{21} = \pi_1$
R_4	$0 \to 1,\ 1 \to 1$	$(1-\pi_1)p_{12} + (1-\pi_1)p_{22} = (1-\pi_1)$

$$D_\theta = max\left\{1 - 2\theta(1-\pi_1), 2\theta(1-\pi_1), \pi_1, (1-\pi_1)\right\}$$

$$= max\left\{1 - 2\theta(1-\pi_1), 2\theta(1-\pi_1), \pi_1\right\}$$

if we assume, without loss of generality, that $\pi_1 \geq (1-\pi_1)$. Therefore,

$$D_\theta = \begin{cases} 1 - 2\theta(1-\pi_1) & \text{if } 0 \leq \theta \leq \frac{1}{2} \\ \pi_1 & \text{if } \frac{1}{2} \leq \theta \leq min(\frac{\pi_1}{2(1-\pi_1)}, 1) \\ 2\theta(1-\pi_1) & \text{if } min(\frac{\pi_1}{2(1-\pi_1)}, 1) \leq \theta \leq 1 \end{cases}$$

and, D_θ attains minimum value when $\frac{1}{2} \leq \theta \leq min(\frac{\pi_1}{2(1-\pi_1),}1)$. □

If $\theta = 1/2$, the resulting invariant PRAM transition probability matrix is

$$P = \begin{pmatrix} 1 - \frac{1-\pi_1}{2\pi_1} & 1/2 \\ \frac{1-\pi_1}{2\pi_1} & 1/2 \end{pmatrix}$$

where π_1 can be estimated from the original data as $\hat{\pi}_1$.

This minimax approach for designing a PRAM transition probability matrix protects against disclosure by limiting an intruder's probability of correct prediction to $min_\theta[max\, D_\theta] = \pi_1$. We can achieve the same level of protection if we choose (see (4.7))

$$P = \begin{pmatrix} \pi_1 & \pi_1 \\ \pi_2 & \pi_2 \end{pmatrix}$$

or if the intruder simply predicts $X = 1$, without using Z, since category 1 is dominant. We note that if the original microdata set was released without PRAMing then the intruder's probability of correct prediction is one. So, if π_1 is large (close to one) implying that the intruder can predict the value of X too accurately then, in such a situation, invariant PRAM would not be effective in protecting against disclosure.

Case of a Polychotomous Sensitive Variable

Here, we would like to investigate the minimax approach discussed above for limiting disclosure when an intruder tries to predict the value of a sensitive polychotomous variable. If we take the invariant PRAM solution (4.7),

$$P = \begin{pmatrix} \pi_1 & \pi_1 & \cdots & \pi_1 \\ \pi_2 & \pi_2 & \cdots & \pi_2 \\ \vdots & \vdots & \vdots & \vdots \\ \pi_k & \pi_k & \cdots & \pi_k \end{pmatrix},$$

then $P(Z = i|X = j) = \pi_i$ and X and Z are independent. Suppose that the prediction rule is to classify all records belong to the cell that has the highest probability. Then $[max\ D_P] = \pi_{(1)}$, which is sufficient to justify that

$$min_P[max\ D_P] \leq \pi_{(1)} \tag{4.21}$$

where $\pi_{(1)}$ is the largest cell probability. On the other hand, for any given invariant P,

$$max\ D_P \geq \pi_{(1)} \tag{4.22}$$

if the category for which the frequency is largest remains the same. Therefore, combining (4.21) and (4.22), we have

$$min_P[max\ D_P] = \pi_{(1)}. \tag{4.23}$$

This minimum may be attained by many PRAM transition probability matrices. A full characterization for a general k-category PRAM, like we did in the binary case above, would require minimizing the maximum probability of correct prediction of k^k different possible rules. This is a difficult problem as the number of rules may be unwieldy when k is large. While the transition probability matrix from (4.7) is invariant and achieves the minimax, it is not attractive because it is singular. In addition, since X and Z generated from this matrix are independent, the desired dependence structure between X and other variables in the microdata set may be very difficult (if not impossible) to recover from the PRAMed microdata set. In the following example we shall show another PRAM transition probability matrix, different from (4.7), that also attains (4.23).

Example 4.4. Suppose X has three categories with $n_1 = 12$, $n_2 = 8$ and $n_3 = 3$ units respectively. From Section 4.7.2, we mentioned that the accuracy of a prediction is loosely related to how concentrated are the predictive distribution $P(X = i|Z = j), j = 1, \cdots, k$. If the predictive distribution is not concentrated or the largest predictive probability is not too large then predictive disclosure is limited. So, to achieve maximum privacy protection, we would want to PRAM the microdata such that each category has a uniform mix of the true original category and each of the other categories. That is, in this example case we would want $\hat{\pi}_1$, $\hat{\pi}_2$ and $\hat{\pi}_3$ to each be around $\frac{1}{3}$. Clearly for our example, it is impossible to achieve this kind of uniform distribution in each category since the categories are not equally populated. So, we suggest starting by ensuring uniformity in the smallest or minority category, and then sequentially getting as close to uniformity as possible in the remaining categories. We provide a simple algorithm to

achieve this as follows:

- Start by PRAMing the units from the minority category 3 into categories 1 and 2 in a uniform manner. That is, we expect about two-thirds of category 3 units are transformed evenly into category 1 and 2.

- Of the remaining unperturbed units in the next minority category 2, about half of them are expected to be transformed into category 1 via PRAMing.

We can demonstrate that the suggested algorithm will decrease disclosure by observing that the maximum probability of correct prediction is given by $\frac{12}{23}$ with optimum prediction rule $1 \to 1$, $2 \to 1$, $3 \to 1$. The resulting PRAM matrix is

$$P_X = \begin{pmatrix} 7/12 & 4/8 & 1/3 \\ 4/12 & 3/8 & 1/3 \\ 1/12 & 1/8 & 1/3 \end{pmatrix}$$

With $\pi = (12/23, 8/23, 3/23)$ we can easily check that $P\pi = \pi$. That is, P is invariant while yielding maximum protection possible.

We would like to note here that the class of solutions for an invariant PRAM suggested by (Gouweleeuw et. al., 1998) will not always provide maximum privacy protection. Recall that $p_{ii} = 1 - \theta\frac{\pi_k}{\pi_i}$, and $p_{ij} = \theta[\frac{\pi_k}{(k-1)\pi_i}]$ if $i \neq j$, where $0 \leq \theta \leq 1$. For our working example, setting $\theta = 2/3$ guarantees that only units in the minority category are maximally protected yielding a transition probability matrix

$$P_X^* = \begin{pmatrix} 10/12 & 1/8 & 1/3 \\ 1/12 & 6/8 & 1/3 \\ 1/12 & 1/8 & 1/3 \end{pmatrix}.$$

However, overall the microdata does not have maximum protection since a simple rule $1 \to 1$, $2 \to 2$, $3 \to 3$ yields a probability of correct prediction of $17/23$ which is higher

than that of our algorithm.

Applying the method akin to a synthetic data set but with no covariates solution via the transition probability matrix in (4.7) (see Section 4.5),

$$P_X^{**} = \begin{pmatrix} 12/23 & 12/23 & 12/23 \\ 8/12 & 8/23 & 8/23 \\ 3/23 & 3/23 & 3/23 \end{pmatrix} = \begin{pmatrix} 6.26/12 & 4.17/8 & 1.57/3 \\ 4.17/12 & 2.78/8 & 1.04/3 \\ 1.57/12 & 1.04/8 & 0.39/3 \end{pmatrix}$$

Using optimum prediction rule $1 \to 1$, $2 \to 1$, $3 \to 1$, the maximum probability of correct prediction ($12/23$) is the same as that of our suggested method above. However, the resulting masked microdata are not PRAMed such that each category has a uniform mix of the true original category and each of the other categories.

4.7.4 Effect of PRAM on Identity Disclosure

We shall discuss identity disclosure in the context of a single categorical variable. Let us suppose a microdata file contains a simple random sample of n records. Further, let X be a categorical variable with categories and suppose that an intruder knows that his or her target unit is in the sample and has value $X = x$. Suppose the original microdata were released and the intruder selects at random one of the units in the original microdata for which $X = x$ and attributes it to the target unit. If the data set has n_x such units then the probability of correct match is

$$P(\text{correct match}) = \frac{1}{n_x}.$$

If there are many units having the value $X = x$, i.e., n_x is large, then the probability of getting the correct match will be low. In this case a surveyor can justify releasing the original microdata file without applying any masking technique to protect disclosure. On the other hand, if n_x is small then the category $X = x$ is considered sensitive, since

the probability of getting a correct match would be high. In this case, it is necessary to mask the microdata before it is released in order to minimize the probability of an intruder correctly identifying the target unit.

In order to evaluate the effect of PRAM on the identification of a target unit for which $X = x$, we shall consider masking the data set by applying the PRAM method on variable X using a transition probability matrix P

$$P = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1k} \\ p_{21} & p_{22} & \cdots & p_{2k} \\ \vdots & \vdots & \cdots & \vdots \\ p_{k1} & p_{k2} & \cdots & p_{kk} \end{pmatrix}$$

Let Z be the variable that results from PRAMing X. Typically, $p_{ii} \geq 0.5$ and often close to 1. So, with large probability, the original category remains unchanged. Naturally, if an intruder knows that $X = x$ then he or she selects one of the records with $Z = x$. We shall assume that at least one record has $Z = x$.

In the masked data, Z is released in place of X, and let T denote the number of units for which $Z = x$. Note that T is a random variable and its value depends on each realized application of PRAM. If m_j is the number of units which are transformed from another category j to category $X = x$ due to PRAMing, then $T = \sum_{j=1}^{k} m_j$. So, $\{T = t\} = \left\{(m_1, \ldots, m_k) : \sum_{j=1}^{k} m_j = t\right\}$ and $P(T = t) = \sum_{(m_1,\ldots,m_k):\sum m_j = t} P(m_1, \ldots, m_k)$ where $P(m_1, \ldots, m_k)$ is made up of independent binomial models. For identification of the target unit, suppose the intruder selects one of the T records for which $Z = x$ in the masked data. Then, the probability that the intruder gets a correct match, given that $T = t > 0$, is:

$$P(\text{correct match}|T = t) = \frac{P(\text{correct match}, T = t)}{P(T = t)}.$$

The numerator is

$$P(\text{correct match}, T = t, E) + P(\text{correct match}, T = t, E^c),$$

where E is the event that the target unit remains in its original category x after it has been perturbed. We note that $P(\text{correct match}, T = t, E^c) = 0$ since there cannot be a correct match if the target unit does not remain in its original category. Now, $P(\text{correct match}, T = t, E) = P(\text{correct match}|T = t, E)P(T = t, E) = \frac{1}{t}P(T = t, E)$. So,

$$\begin{aligned}
P(\text{correct match}|T = t) &= \frac{\frac{1}{t}P(T = t, E)}{P(T = t)} \\
&= \frac{\frac{1}{t}P(E)P(T = t|E)}{P(T = t)} \\
&= \frac{\frac{1}{t}P(E)P(T = t|E)}{P(T = t)} \\
&= \frac{\frac{1}{t}P(E)P(\sum_j m_j = t)}{P(T = t)} \\
&= \frac{\frac{1}{t}P(E)\sum_{A_t} P(m_1, \ldots, m_k)}{P(T = t)} \\
&= \frac{\frac{1}{t}P(E)\sum_{A_t} P(m_1)P(m_2)\ldots P(m_k)}{P(T = t)}
\end{aligned}$$

$$= \frac{\frac{1}{t}p_{xx}\sum_{A_{t-1}}\{\begin{pmatrix} n_x - 1 \\ m_x - 1 \end{pmatrix} p_{xx}^{m_x}(1 - p_{xx})^{n_x - 1 - m_x}\prod_{\substack{j=1 \\ j\neq x}}^{k}\begin{pmatrix} n_j \\ m_j \end{pmatrix} p_{xj}^{m_j}(1 - p_{xj})^{n_j - m_j}\}}{\sum_{A_t}\prod_{j=1}^{k}\begin{pmatrix} n_j \\ m_j \end{pmatrix} p_j^{m_j}(1 - p_j)^{n_j - m_j}}, \tag{4.24}$$

where $A_t = \{(m_1, \ldots, m_k) : \sum_{j=1}^{k} m_j = t\}$ and $P(E) = p_{xx} = P(Z = x|X = x)$ is the probability that category $X = x$ remains unchanged after PRAMing.

For very large t, the value of (4.24) could be small, depending on the other terms in the numerator, and there would be low disclosure risk. However, a small t could result in a higher disclosure risk.

As Nayak (2008) observed, the value of t, which determines the disclosure risk, depends on a specific realization of the PRAM procedure. On the other hand, at the time of selecting the transition probability matrix P, the value of t is unknown. So, for evaluating and constructing a P, we should consider both the probability distribution of T and the conditional probabilities of the correct match. Our approach is to consider the likely values of t within a 95% confidence interval. That is, we consider values t for category $X = x$ such that $P(t_l < t < t_u) = 0.95$ and calculate $P(\text{correct match}|T = t)$. Now given $P(t_l < t < t_u) = 0.95$, if $P_0 = max_t P(\text{correct match}|T = t) > \nu$, where ν is a fixed threshold value for an acceptable level of disclosure risk, then we shall reduce p_{xx} accordingly in order to protect the privacy of the target unit. This process needs to be carried out for all categories of X, or at least for all the sensitive categories of X, in order to specify the entries of a PRAM transition probability matrix P .

4.7.5 Effect of PRAM on Predictive Disclosure

Predictive disclosure can occur under several scenarios. We shall consider two scenarios of predictive disclosure: with or without covariates.

4.7.5.1 Predictive Disclosure Without Covariates

Let X be a sensitive categorical variable of interest with categories, $i = 1, \cdots, k$. Let us further suppose that there are no covariates or additional information on our dataset. The issue is this: An intruder knows that a unit of interest is a record in the data set or in sample, and he or she wants to know which category the unit belongs to. Assuming no identity disclosure has occurred, the best an intruder could do is to predict that a unit of interest belongs to the mostly likely category determined by the category with

the largest estimated proportion. The intruder is likely to predict that $X = i$ with $\hat{\pi}_i$, which is the relative frequency of $X = i$. Suppose $max\{\hat{\pi}_1, \cdots, \hat{\pi}_k\} = \hat{\pi}_{(1)} = \hat{\pi}_j$, i.e., the jth category has the highest relative frequency, the intruder may conclude that the most likely category is j and his confidence in this conclusion ($X = j$) increases with $\hat{\pi}_j$. We would like to emphasize that in order to limit disclosure, we want to make not just the intruder's prediction difficult but also lower the confidence level of his or her prediction.

After the data has been PRAMed, the intruder will not have access to the actual frequency counts for the different categories. Instead the intruder obtains $\tilde{\pi}$ from the PRAMed data set, and he or she should recognize that $\tilde{\pi}$ is an estimate of $\hat{\pi}$ (and also of π) and it would be wrong to conclude that $X = i$ simply based on $\tilde{\pi}_i$. However, based only on the knowledge that the unit of interest is in the microdata set and that the true probability that $X = i$ is π_i, the intruder may calculate a confidence interval (CI) around π_i, say $(\tilde{\pi}_{il}, \tilde{\pi}_{iu})$. Then, he or she may conclude that $X = i$, with a certain confidence level, is between $\tilde{\pi}_{il}$ and $\tilde{\pi}_{iu}$. Thus, PRAM adds uncertainty to the intruder's prediction.

Since $\tilde{\pi}$ is an unbiased estimate of $\hat{\pi}$, the most likely category shown by $\tilde{\pi}$ and $\hat{\pi}$ may be the same. Suppose, $\hat{\pi}_1 = max\{\hat{\pi}_1, \cdots, \hat{\pi}_k\}$ and $\tilde{\pi}_1 = max\{\tilde{\pi}_1, \cdots, \tilde{\pi}_k\}$. Then, in both cases, based on the original and PRAMed data, the intruder will conclude that the most likely category for the unit is $X = 1$. Even in this case, the intruder's confidence in the conclusion will be different depending on whether the original or PRAMed data are released. For actual data, he knows that the conclusion that $X = 1$ is correct with probability $\hat{\pi}_1$, whereas for PRAMed data, the probability of the conclusion being correct may be believed to be between $\tilde{\pi}_{1l}$ and $\tilde{\pi}_{1u}$. The CI for the components of $\boldsymbol{\pi}$ may overlap and this also complicates an intruder's finding of the most likely category. For the actual data set, the category with the largest proportion of units may be category

1. However, in the PRAMed data, if the confidence intervals around $\tilde{\pi}_1$ and $\tilde{\pi}_2$ overlap then the intruder cannot be confident in predicting that the unit of interest belongs to category 1 even with $\tilde{\pi}_1 > \tilde{\pi}_2$.

Remark 4.2. Suppose the intruder does not know if the unit is in the sample or not. So even if the original data are released, the probability that $X = i$ is not $\hat{\pi}_i$. The true probability, π_i, is unknown. One should recognize $\hat{\pi}$ as an estimate of π. Then, the CI for π_i is $(\hat{\pi}_{il}, \hat{\pi}_{iu})$. Naturally, the intruder can make more precise predictions if the unit is in the sample and he or she knows that.

4.7.5.2 Predictive Disclosure With Covariates

In general, suppose we have two variables X and Y. Let X be the sensitive variable and Y the nonsensitive identifying variable. In a more general case, X could be the cross-classification of all sensitive variables and Y could be the cross-classification of all nonsensitive variables. For our purposes of examining the effect of PRAM on privacy protection, $\pi_{i|j} = P(X = i|Y = j)$ is of direct relevance.

Consider the case where the intruder is interested in one unit in the microdata set, and the intruder knows that $Y = j$. The intruder does not know $\pi_{i|j}$; however, from the actual data set, he or she can predict $\pi_{i|j}$ from $\hat{\pi}_{i|j} = \frac{\hat{\pi}_{ij}}{\hat{\pi}_{\cdot j}}$, where $\hat{\pi}_{\cdot j}$ is the marginal proportion of $Y = j$.

After the microdata set is PRAMed, an intruder can estimate $\tilde{\pi} = \{\tilde{\pi}_{ij}\}$ from the publicly released data. Subsequently, he or she can obtain estimates for predicting $X = i$, $\pi_{i|j}$, as $\tilde{\pi}_{i|j} = \frac{\tilde{\pi}_{ij}}{\tilde{\pi}_j}$. All previous discussion items in Section 4.7.5.1 are applicable here to the conditional relative frequency. So, even if the most likely categories match before and after PRAMing, the uncertainty levels would be different. That is, PRAM introduces uncertainty about $\hat{\pi}$. We also note that there are two sources of variation in the estimate from the PRAMed data:

- Variation from the actual sampling
- Variation from the randomization process (PRAMing) based on sample data

Only Sensitive Variable is PRAMed

We can estimate $P(X = i|Y = j) = \frac{\pi_{ij}}{\pi_j}$ by $\frac{\tilde{\pi}_{ij}}{\tilde{\pi}_j}$ from the PRAMed data and compare with $\frac{\hat{\pi}_{ij}}{\hat{\pi}_j}$ from the original data. If their distributions are different then some uncertainty is introduced in the estimates derived from the PRAMed data, and subsequently limiting disclosure. We shall examine their expectations. If only X is PRAMed (and Y is left unchanged) then $\tilde{\pi}_j = \hat{\pi}_j$, so

$$E_R(\frac{\tilde{\pi}_{ij}}{\tilde{\pi}_j}) = \frac{E_R(\tilde{\pi}_{ij})}{\hat{\pi}_j} = \frac{\hat{\pi}_{ij}}{\hat{\pi}_j} \tag{4.25}$$

provided $\hat{\pi}_j \neq 0$, say. From (4.25), $\frac{\tilde{\pi}_{ij}}{\tilde{\pi}_j}$ is an exact unbiased estimate of $\frac{\hat{\pi}_{ij}}{\hat{\pi}_j}$. We note that this case is similar to predictive disclosure with no covariates since X is PRAMed within Y.

Only Nonsensitive Variable is PRAMed

On the other hand when Y is the PRAMed variable, $E_R(\frac{\tilde{\pi}_{ij}}{\tilde{\pi}_j})$ cannot be reduced further without knowing the distribution of the ratio. Thus, unlike in (4.25) above,

$$E_R(\frac{\tilde{\pi}_{ij}}{\tilde{\pi}_j}) \neq (\frac{\hat{\pi}_{ij}}{\hat{\pi}_j}). \tag{4.26}$$

Here, $\frac{\tilde{\pi}_{ij}}{\tilde{\pi}_j}$ is not exactly unbiased for $\frac{\hat{\pi}_{ij}}{\hat{\pi}_j}$ from (4.26). Apart from the uncertainty in the estimates of the conditional relative frequencies as discussed in Section 4.7.5.1, there may be additional uncertainty introduced by the bias $E_R(\frac{\tilde{\pi}_{ij}}{\tilde{\pi}_j}) - (\frac{\hat{\pi}_{ij}}{\hat{\pi}_j})$. This indicates that for protection of privacy with PRAM, the statistical agency may want to consider PRAMing nonsensitive variables over PRAMing only sensitive variables.

Both Sensitive and Nonsensitive Variables are PRAMed

The comments above when "Only Nonsensitive Variable is PRAMed" would apply if both X and Y are PRAMed.

Chapter 5

Conclusions and Future Research

5.1 Summary of the Dissertation

In this dissertation, we studied two connected research topics in the area of survey respondents' privacy protection and confidentiality. In our first topic, we presented a unified framework for discussing RR surveys of dichotomous populations with multiple response categories. Several RR procedures have appeared in the literature, but typically, each procedure has been discussed within its own framework. We believe this has hindered systematic thinking about the core statistical issues and has led to erroneous conclusions. A common framework is helpful, and perhaps necessary, for abstraction and formalization of the key elements relating to respondents' privacy and statistical efficiency and comparison of various procedures. We hope the ideas put forward in this dissertation will be helpful in developing a unified theory of RR surveys, comparing various procedures and reaching valid conclusions.

We also studied RR designs for polychotomous populations. We extended the theory and framework for RR surveys of dichotomous populations to RR surveys of polychotomous populations, including to situations under finite population settings. We compared a class of polychotomous RR designs where only one category is sensitive, and we derived conditions for which one can construct better polychotomous RR designs of a specified structure. While we discussed only polychotomous response variables, we believe our

ideas can be extended to RR surveys of dichotomous and polychotomous populations with quantitative response variables, such as the procedures discussed in Franklin (1989) and Chua and Tsui (2000). Other RR design analogues for quantitative data involve noise addition or multiplication (see Warner (1971), Pollock and Bek (1976), and Duffy and Waterton (1984)). An RR technique based on a multiplicative model is described by Poole (1974) where the responder is asked to multiply his or her true response X by a random number Y and report only the product (or randomized response) $Z = XY$ to the surveyor. Let $F(.)$ and $W(.)$ represent the cummulative distribution functions of X and Z respectively, and let $w(.)$ be the density of Z. If Y is uniformly distributed on [0, 1], say, then

$$F(x) = W(x) - xw(x). \tag{5.1}$$

Using (5.1) an estimate of F can be obtained once W and w are estimated from sample observations of Z. This result was extended to estimate the multivariate distribution of a k-dimensional vector of continuous variables (see Poole and Clayton, 1982). Other papers such as Kim and Flueck (1987) and Poole and Clayton (1982) provide solutions for RR techniques based on an additive model.

A unified theory is also helpful for recognizing connections of RR surveys to other areas of statistics, such as comparison of experiments (see Remark 2.6) and estimation from open surveys. RR surveys are closely related to our second research topic - The PRAM for controlling statistical disclosure. This method is concerned with protecting respondents' privacy while releasing microdata (already collected) for public use. It stochastically transforms the values of categorical variables in a data set using a known Markov matrix. As noted by Van den Hout and Van der Heijden (2002), mathematically the PRAM is equivalent to an RR procedure. Both are concerned with protection of respondents' privacy and statistical efficiency; only difference is that in RR surveys, the responder randomizes the response during the data gathering stage whereas in PRAM

randomization is carried out by the surveyor after the data are collected. Thus, the results for RR surveys can be used beneficially in statistical disclosure control.

We described several variations of the PRAM procedure, and we demonstrated that any PRAM procedure can be regarded as a PRAMing of the cross-classification of all the variables in a data set. We studied he connections of PRAM and RR and noted that the estimators developed for RR of polychotomous populations can be used for PRAM. We contributed some theory to a special case of PRAM known as invariant PRAM and introduced the notion of a strongly invariant PRAM. We developed a new approach for constructing invariant PRAM matrices, and we clarified certain perceptions of invariant PRAM that were not fully justified in past literature. We presented estimation results from an invariantly PRAMed data, and finally we examine the effectiveness of PRAM for limiting statistical disclosure.

5.2 Future Research

There are several topics that we identified in the course of this research that are either connected to our work, or that we did not have time to complete.

1. In Section 2.6.1, we mentioned that following Chaudhuri (2001, 2004), one can express $V(e^*(s,z))$ in other forms and thence obtain other unbiased estimators of it. One undesirable feature of $V(e^*(s,z))$ is that it may be negative. Another form of variance estimator is the nonnegative estimator introduced in Rao (1979) and extended in Chaudhuri and Pal (2002). We would like to explore expressing $V(e^*(s,z))$ in this form.

2. We presented a limited result in our comparison of $(k \to k)$ RR designs in Section 3.5. Developing a general framework for comparing $(k \to k)$ RR designs is an important issue, and we would like to look into it further.

3. Data swapping (Dalenius and Riess, 1982) resembles PRAM. It is another method for limiting disclosure in microdata sets that contain categorical variables. The main underlying idea of data swapping is to transform or mask a microdata set by exchanging values of senstitive variables among individual records in such a way that marginal frequency counts are preserved. Like in the PRAM, such masking limits disclosure by introducing uncertainty about the sensitive data and at the same time maintaining accuracy of statistical inferences by preserving certain summary statistics of the microdata. Several variants of data swapping have been proposed and investigated in the literature, e.g., Rank-based Swapping (Moore, 1996), Data Shuffling (Muralidhar and Sarathy, 2006), and others; see Fienberg and McIntyre (2005) for a discussion of many of these procedures. We would like to explore the connections between data swapping and PRAM, and develop a general framework that encompasses both methods.

4. In section 4.6, we only consider estimation from invariant PRAM under an infinite population setting. We would like to extend our results to probability sampling designs under a finite population setting.

5. In Section 4.7, we discussed the effect of PRAMing on a microdata set under several disclosure scenarios. In particular, we discussed the impact of PRAM on an intruder's ability to accurately predict the value of a sensitive charateristic of a given unit with or without covariates, and we developed some asymptotic results to determine the extra variance from PRAMing variables in a data set when there are covariates. We would like to consider real applications of our findings on survey data so as to gain further insight into how well this variance inflation due to randomization can help with preventing disclosure.

References

[1] Abul-Ela, Abdel-Latif, A., Greenberg, B. G., and Horvitz, D. G. (1967). A multiproportions RR model. *Journal of the American Statistical Association*, **62**, 990-1008.

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

[3] Anderson, H. (1977). Efficiency versus protection in a general randomized response model. *Scandinavian Journal of Statist*ics, **4**, 11-19.

[4] Bethlehem, J.G., Keller, W.J. and Pannekoek, J. (1990). Disclosure control of microdata. *Journal of the American Statistical Association*, **85**, 38-45.

[5] Blackwell, D. (1951). Comparison of experiments, *Proceedings of Second Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, Berkeley, 93-102.

[6] Blackwell, D. (1953). Equivalent comparison of experiments, *Annals of Mathematical Statist*ics, **24**, 265-272.

[7] Bourke, P. D., and Dalenius, T. (1976). Some new ideas in the realm of randomized enquiries. *International Statistical Rev*iew, **44**, 219-221.

[8] Bycroft, C. and Merrett, K. (2005) Experience of using a Post Randomisation Method at the Office for National Statistics, *United Nations Economic Commission for Europe Work Session on Statistical Data Confidentiality.*

[9] 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**, 37-42.

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

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

[12] Chaudhuri, A. and Pal, S. (2002). On certain alternative mean square error estimators in complex survey sampling. *Journal of Statistical Planning and Inference*, **104 (2)**, 363-375.

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

[14] Christofides, T.C. (2005). Randomized response in stratified sampling. *Journal of Statistical Planning and Inference,* **128**, 303-310.

[15] Chua, T.C. and Tsui, A.K. (2000). Procuring honest responses indirectly. *Journal of Statistical Planning and Inference*, **90**, 107-116.

[16] Cox, L.H. (1981). Linear sensitivity measures in statistical disclosure control. *Journal of Statistical Planning and Inference,* **5**, 153-164.

[17] Dalenius, T. and Reiss, S.P. (1982). Data-Swapping: A technique for disclosure control. *Journal of Statistical Planning and Inference,* **6**, 73-85.

[18] De Waal, T. and Quere, R. (2003). A Fast and Simple Algorithm for Automatic Editing of Mixed Data. *Journal of Official Statistics*, **19**, 383–402.

[19] De Wolf, P.P., Gouweleeuw, J.M., Kooiman, P., and Willenborg, L.C.R.J. (1997). Reflections on PRAM. *Statistical Data Protection Conference Proceedings*, Lisbon, Eurostat, Luxembourg, 337-349.

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

[21] Duffy, J.C. and Waterton, J. J. (1984). Randomized Response Models for Estimating the Distribution Function of a Quantitative Character. *International Statistical Review,* **52**, 165-171.

[22] Duncan, G.T. and Lambert, D. (1986). Disclosure-limited data dissemination. *Journal of the American Statistical Association*, **81 (393)**, 10-19.

[23] Duncan, G.T. and Lambert, D. (1989). The risk of disclosure for microdata. *Journal of Business and Economic Statist*ics, **7**, 207-217.

[24] Federal Committee on Statistical Methodology. (1994, revised 2005). Report on Statistical Disclosure Limitation Methodology. *Statistical Policy Working Paper 22 (Revised).* <http://www.fcsm.gov/working-papers/spwp22.html>

[25] Fienberg, S.E. and McIntyre, J. (2005). Data Swapping: Variations on a Theme by Dalenius and Reiss. *Journal of Official Statistics*, **9**, 383–406.

[26] Fienberg, S. E., Makov, E. U., and Steel, R. J. (1998). Disclosure Limitation using Perturbation and Related Methods for Categorical Data. *Journal of Official Statistics*, **14**, 485-502.

[27] 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 Meth*ods, **6**, 1511-1524.

[28] Franklin, L.A. (1989). Randomized response sampling from dichotomous populations with continuous randomization. *Survey Methodology*, **15**, 225-235.

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

[30] 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.

[31] Greenberg, B. and Zayatz, L. (1992). Strategies for measuring risk in public use microdata files. *Statistica Neerlandica*, **46**, 33-48.

[32] Gross, B., Guiblin, P. and Merrett, K. (2004). Risk Assessment of the Individual Sample of Anonymised Records (SAR) from the 2001 Census, *Office for National Statistics*. http://www.ccsr.ac.uk/sars/events/2004-09-30/slides/index.html

[33] Keller-McNulty, S., Nakhleh, C.W. and Singpurwalla, N.D. (2005). A Paradigm for Masking (Camouflaging) Information. *International Statistical Rev*iew, **73**, 331-349.

[34] Kim, J-I. and Flueck, J. A. (1978). An additive randomised response model. *Proceedings of the Survey Research Section, American Statistical Association*, 351-355.

[35] Kim, J-M. and Warde, W.D. (2004). A stratified Warner's randomized response model. *Journal of Statistical Planning and Inference*, **120**, 155-165.

[36] Kooiman, P., Willenborg, L., and Gouweleeuw, J. (1997). A method for disclosure limitation of microdata. *Research paper 9705*, Statistics Netherlands, Voorburg.

[37] Konnu, J. (2007). The use of protected micro data in tabulation: A case of SDC-methods, microaggregation and PRAM. *United Nations Economic Commission for Europe Work Session on Statistical Data Confidentiality.*

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

[39] Lambert, D. (1993). Measure of disclosure risk and harm. *Journal of Official Statistics*, **9 (2)**, 313331.

[40] Lanke, J. (1976). On the degree of protection in randomized interviews. *International Statistical Rev*iew, **44**, 197-203.

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

[42] Liu, P. T., and Chow, L. P. (1976). A new discrete quantitative RR model. *Journal of the American Statistical Association*, **71**, 72-73.

[43] Mangat, N.S. (1994). An improved randomized response strategy. *Journal of the Royal Statistical Society*, **56**, 93-95.

[44] Mangat, N.S. and Singh, R. (1990). An alternative randomized response procedure. *Biometrika*, **77**, 439-442.

[45] Moore, R.A. (1996). Controlled data swapping techniques for masking public use microdata sets. *RR 96-05*. U.S. Bureau of the Census, Washington, DC.

[46] Muralidhar, K. and Sarathy, R. (2006). Data shuffling - A new masking approach for numerical data. *Management Science*, **52 (5)**, 658-670.

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

[48] Nayak, T.K. (2008). Reflections on Data Perturbation and Post-randomization for Statistical Disclosure Control. *Journal of Statistics and Applications*, **3**, 303-313.

[49] 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.

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

[51] Pollock, K. H. and Bek, Y. (1976). A comparison of three randomized response models for quantitative data. *Journal of the American Statistical Association*, **71 (356)**, 994-886.

[52] Poole, W. K. (1974). Estimation of the Distribution Function of a Continuous Random Variable Through Randomized Response, *Journal of the American Statistical Association*, **69**, 1002-1005.

[53] Poole, W. K. and Clayton, A. C. (1982). Generalisations of continuous contamination model for continuous type random variables. *Communications in Statistics - Theory and Meth*ods, **11**, 1733-1742.

[54] Raghunathan, T. E., Reiter, J. P., and Rubin, D. B. (2003). Multiple Imputation for Statistical Disclosure Limitation. *Journal of Official Statistics,* **19**, 1–16.

[55] Rao, J.N.K. (1979). On Deriving Mean Square Errors and other Non-Negative Unbiased Estimators in Finite Population Sampling. *Journal of Indian Statististical Assoc*iation, **17**, 125–136.

[56] Reiter, J. P. (2002). Satisfying Disclosure Restrictions with Synthetic Data Sets. *Journal of Official Statistics*, **18**, 531-543.

[57] Reiter, J.P. (2005). Estimating identification risk in microdata. *Journal of the American Statistical Association*, **100**, 1101-1113.

[58] Rubin, D.B. (1987). *Multiple imputation for nonresponse in surveys*. J. Wiley & Sons, New York.

[59] Rubin, D. B. (1993). Satisfying Confidentiality Constraints through the Use of Synthetic Multiply-imputed Microdata. *Journal of Official Statistics*, **91**, 461-468.

[60] Shlomo, N. and De Waal, T. (2008). Protection of Micro-data Subject to Edit Constraints Against Statistical Disclosure. *Journal of Official Statistics*, **24 (2)**, 1-26.

[61] Skinner, C.J. and Elliot, M.J. (2002). A measure of disclosure risk for microdata. *Journal of the Royal Statistical Society*, Series B, **64**, 855-867.

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

[63] Warner, S.L. (1965). Randomized Response: a Survey Technique for Eliminating Answer Bias. *Journal of the American Statistical Association*, **66**, 884-888.

[64] Warner, S.L. (1971). The Linear Randomized Response: a survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association*, **60**, 63-69.

[65] Willenborg, L.C.R.J. and De Waal, T. (2001). *Elements of Statistical Disclosure Control*. New York: Springer.