Reflections on PRAM

Peter-Paul de Wolf
José M. Gouweleeuw
Peter Kooiman
Leon Willenborg *

Statistics Netherlands
Department of Statistical Methods
P.O. Box 4000
2270 JM Voorburg
The Netherlands.
E-mail: pwof@cbs.nl, jgww@cbs.nl, pkmn@cbs.nl and lwlg@cbs.nl

Abstract

PRAM is a probabilistic, perturbative method for disclosure protection of categorical variables in microdata files. If PRAM is to be applied, several issues should be carefully considered. The microdata file will usually contain a specific structure, e.g., a hierarchical structure when all members of a household are present in the data file. To what extent should PRAM conserve that structure? How should a user of the perturbed file deal with variables that can (logically) be deduced from other variables on which PRAM has been applied? How should the probability mechanism, used to implement PRAM be chosen in the first place? How well does PRAM limit the risk of disclosing sensitive information? In this paper, these questions will be considered.

Keywords: Post RAndomisation Method (PRAM), disclosure limitation, perturbed data, Markov matrix, expectation ratio.

1 Introduction

The Post RAndomisation Method (PRAM) was introduced in [1] as a method for disclosure protection of categorical variables in microdata files. For numerical variables, comparable methods already exist, see for example [2]. PRAM produces a microdata file in which the scores on some categorical variables for certain records in the original file are changed into a different score according to a prescribed probability mechanism. The randomness of the procedure implies that matching a record in the perturbed file to a record of a known individual in the population could, with a certain (high) probability, be a mismatch instead of a true match. However, since the used probability mechanism is known, statistical analyses can still legitimately be performed, be it with a slight adjustment of the standard methods.

* The views expressed in this paper are those of the authors and do not necessarily reflect the policies of Statistics Netherlands.

The theory of applying PRAM is dealt with quite extensively in both [1] and [3]. At Statistics Netherlands, PRAM has been applied to the Dutch National Travel Survey. The application showed that several issues concerning the implementation of PRAM and its resulting policy implications, deserve some extra attention. This paper will discuss some of these issues, will present possible solutions and will indicate directions for further research.

When applying PRAM on a microdata file, several questions arise. First of all, by applying PRAM, certain inconsistencies will occur between variables in a record. How should they be dealt with? Note however, that this is not a problem specific to PRAM but a problem that occurs with any perturbative method. Moreover, often a hierarchical structure is present in the microdata file, for example when records of all the members of a household are contained in the file. Consequently, certain dependencies between these records are present in the data file. The question then becomes how PRAM will behave in the presence of such a structure.

Secondly, before PRAM can be applied, the transition probabilities should be chosen appropriately. The specific choice of the Markov matrix will influence the disclosure risk. Therefore, the process of choosing the Markov matrix will usually turn out to be an iterative process: when a certain matrix is chosen, the effects on the disclosure risk will be considered. If these effects are not satisfactory, the choice of the matrix should be reconsidered. Then the effects on the disclosure risk should be re-examined, etc.

Finally, the effect of PRAM on statistical analyses performed on the perturbed microdata file should be considered. The effects on some analyses were discussed in [1] and [3].

In Section 2 a brief description of PRAM is given, to introduce the notation used throughout this paper. Section 3 deals with the effect of PRAM on structures present in the microdata file under consideration. In Section 4, issues concerning the choice of the Markov matrix will be considered and in Section 5 a methodology to determine the amount of disclosure limitation introduced by applying PRAM will be presented. Section 6 consists of some miscellaneous remarks that were not directly covered by the other sections.

2 A Description of PRAM

In this section we will briefly describe the theory involving PRAM, mainly in order to introduce the notation that will be used throughout this paper. For details we refer to [3].

Let ξ denote a categorical variable in the original file to which PRAM will be applied and let X denote the same categorical variable in the perturbed file. Moreover, assume that ξ, and hence X as well, has K categories numbered $1, \ldots, K$. Define the transition probabilities involved in applying PRAM by $p_{kl} = \mathbb{P}(X = l \mid \xi = k)$, i.e., the probability that an original score $\xi = k$ is changed into the score $X = l$, for all $k, l = 1, \ldots, K$. PRAM is then fully described by the $K \times K$ Markov matrix P whose entries are the transition probabilities p_{kl}. Finally, let $\xi^{(r)}$ and $X^{(r)}$ denote the value of ξ respectively X of the r-th record in the corresponding microdata file. Applying PRAM now means that, given $\xi^{(r)} = k$, the score on $X^{(r)}$ is drawn from the probability distribution $p_{k1}, \ldots, p_{kK}$. For each record in the original file, this procedure is performed independently of the other records.

The previous description of the transition probabilities p_{kl} shows the close relationship

PRAM has with more familiar randomized response techniques as discussed in e.g., [4] and [5]. In the classical randomized response models however, the probability mechanism is applied preceding the survey and independently of the answers of the respondents. PRAM is applied after the survey has been performed and will in general depend on the given answers.

Obviously PRAM can be applied independently to different variables. However, it is sometimes convenient to be able to apply PRAM to different variables simultaneously, e.g., to be able to cope with dependencies between variables. In [3] it is described how this can be implemented. It essentially amounts to considering all the variables to which PRAM will be applied simultaneously, as one compounded variable. That is, a new variable will be defined whose categories consist of the crossings of all the variables on which PRAM will be applied. E.g., if PRAM is to be applied to gender (male/female) as well as marital status (married, unmarried, widowed) the compounded variable will have the categories married male, married female, unmarried male, unmarried female, widowed male and widowed female. As a consequence, certain dependencies between the variables that constitute to the compounded variable can be taken into account. Moreover, a compounded variable may just as well consist of a mixture of variables to which PRAM will be applied and variables that will always be kept unchanged.

So far, no restrictions have been imposed on the Markov matrix P of transition probabilities. In [3] two versions of PRAM are discussed: general PRAM and invariant PRAM. The general version only assumes that the Markov matrix P has an inverse P^{-1}. That inverse can be used to correct contingency tables based on the perturbed file to obtain unbiased estimates of the corresponding tables that would result from the original file. Moreover, it can be used to correct for the effect of PRAM in case of several other statistical analyses, as shown in [1] and [3]. For easy reference, the formulas concerning contingency tables will now be stated. Denote the contingency table of the (compounded) variable ξ in the original file by T_ξ and the corresponding table based on the perturbed file by T_X. Then

$$\mathbb{E}\left(T_X \;\middle|\; \xi^{(1)}, \ldots, \xi^{(n)}\right) = P^t T_\xi, \tag{1}$$

where t denotes transposition and n is the number of records in the microdata file. Note that the dependency on the original file is represented by conditioning on the scores of the considered (compounded) variable in that file.

An unbiased estimator can hence be obtained by defining

$$\hat{T}_\xi = \left(P^{-1}\right)^t T_X \,. \tag{2}$$

As shown in this simple situation of contingency tables, general PRAM involves some extra effort on account of the user of the perturbed microdata file, to obtain unbiased estimates. The other version of PRAM that was discussed in [3], invariant PRAM, imposes extra conditions on the Markov matrix P, but releases the user of the perturbed file of the extra effort to obtain unbiased estimates: the user does not have to use the inverse of the Markov matrix in his analyses, i.e., the user can work with the perturbed file as if it were the original file. The condition needed for invariant PRAM is that the Markov matrix P should be chosen such that

$$P^t T_\xi = T_\xi, \tag{3}$$

for then T_X itself (without the extra multiplication by P^{-1}) can be used as an unbiased estimator of T_ξ. This is easily derived from equation (1). Note that equation (3) implies

that only table T_ξ can directly be estimated by T_X, hence not necessarily every table that can possibly be constructed from the data-file. In order to let the user indeed use the perturbed file as if it were the original file, the variable ξ should be compounded of all variables in the microdata file. Moreover, even in that case it still remains to be investigated whether analyses other than those based on contingency tables can be done directly.

In practice, it may actually be impossible to apply invariant PRAM in such a way that the simultaneous distribution of all the variables in the file is preserved. In that case, ξ has a large number of categories, and many categories will contain only few observations, or none at all. This will in general not allow enough freedom for the choice of P if it also has to satisfy equation (3).

We conclude this section by listing the advantages and drawbacks of invariant PRAM as opposed to general PRAM. If general PRAM is used then the transition matrix P can be any invertible Markov matrix, if invariant PRAM is used, the matrix P should additionally satisfy equation (3). Thus, general PRAM allows more freedom in the choice of P, which makes it easier to obtain a safe file. On the other hand, general PRAM has the drawback that the analysis on the perturbed file becomes more complicated than in the case of invariant PRAM. Another drawback of general PRAM is that when contingency tables are estimated from the perturbed file, these may contain negative estimated frequencies, since P^{-1} will in general not be a Markov matrix. When invariant PRAM is applied, this can not occur.

However, note that general PRAM and invariant PRAM represent two extremes. It is also possible to use a mixture of both. This entails that only some simultaneous distributions in the original data file are preserved, implying that some analyses can be performed directly on the perturbed file, while others are more complicated. In practice, it may be possible that even though only some simultaneous distributions are preserved in the PRAM process, the simultaneous distribution of all variables in the perturbed data file is close to that in the original data file. In that case, the perturbed file can be treated as if it were the result of applying invariant PRAM.

3 PRAM and Data Integrity

When PRAM has been applied to a microdata file, the resulting perturbed file may contain inconsistencies. This can give an intruder a clue as to which scores have been altered as a result of applying PRAM, and he may thus be able to (partly) undo the effect of PRAM. Obviously, this is undesirable from the point of view of data protection and therefore, inconsistencies in the perturbed file should be avoided as much as possible. Moreover, in the process of data collection at a statistical office, certain edit-rules are usually checked in order to detect (accidentally) recorded inconsistencies. Obviously, these rules can be of help in determining the inconsistencies to be considered when applying PRAM. In this section, different kinds of inconsistencies that may occur are discussed, and suggestions on how to avoid them are given. This will have implications for the way PRAM is applied to the data file.

First of all, if the file contains a hierarchical structure, inconsistencies can occur between records. A file contains a hierarchical structure if the variables in the file are measured on two or more different levels, i.e., each record contains a key that can be used

to link several records uniquely to a group of records. (A key does not necessarily have to be some identification number, in some cases for example a weight can be used as a key.) Moreover, each record contains information on the group of records it belongs to, as well as information specific to itself. Since PRAM is applied to each record independently, several records that belong to the same group of records can have different scores on the same group-variable. In some cases an intruder may thus be able to (partly) undo the disclosure limitation. Obviously this is an undesirable situation.

An possible solution is to consider each group of records as one large record consisting of all the information of the members of the group that are present in the microdata file, to which PRAM will be applied. If PRAM can be applied to the group- and record-variables independently, then this is a good solution and analyses on the perturbed file can proceed as described in [1]. It is not clear how to apply PRAM to a group- and a record-variable simultaneously, as for example is the case when we want to apply invariant PRAM in such a way that the simultaneous distribution of a group- and a record-variable is preserved, and these variables are correlated. This is a topic for further research.

The situation described so far occurs whenever the data file has some hierarchical structure, i.e., whenever the variables in the data file are measured on two (or more) different levels.

Another problem concerning inconsistencies that can occur in files with a hierarchical structure, is illustrated by the following example of a household survey. Consider the variable 'size of the household'. If the microdata file is such that it only contains households from which all members have responded, then applying PRAM to the variable 'size of the household' may lead to a situation where this variable has the value 4, while 5 records of members of this household are present in the data file. If not all the members of a household have responded, then the number of records corresponding to one household should always be smaller than the value of the variable 'size of the household'. PRAM can then be applied to that variable, as long as this rule is followed, which implies that the value of 'size of the household' can only be replaced by a value that is equal or larger. In this case, it is easy to see that it is impossible to apply invariant PRAM.

Inconsistencies can not only occur between different records, but also between different variables within one record. The first situation in which this type of inconsistency can occur, is when the score on a variable can be derived from the score on one or more other variables in the same record. This happens for instance if a variable is a refinement of a second variable (in which case the first variable can be derived from the second) or if a variable is a combination of two or more variables. If PRAM is applied to a variable then this may introduce inconsistencies with all the variables that can be derived from this variable. Again this can help an intruder to (partly) undo the effects of PRAM, which is undesirable. This problem can be solved easily by removing all the variables that can be derived from one or more other variables from the original data file and then applying PRAM on the resulting reduced file. The derived variables should then be treated as new variables that have to be matched to the perturbed file. In [1] it is described how this could be done. When invariant PRAM is applied, this implies that the score on a derived variable can directly be computed for all the records in the perturbed file, and these variables can be added back to the perturbed file. For general PRAM the situation is a little more complicated.

A special variable that is derived from other variables, is a (sampling) weight, that is present in most microdata files. A weight can be derived through the variables in the weighting model. If PRAM is applied to one or more variables in the weighting model, and the weight is copied unaltered to the perturbed data file, then this can in some cases give an intruder a clue as to in which records the scores on the variables of the weighting model have been altered. If this is the case, some action needs to be taken. Since the theory of [1] does not apply in this case, it is not completely clear what should be done. When invariant PRAM is used, it seems reasonable to remove the weights from the data file, apply PRAM and recalculate the weights. At this moment, further research is necessary on the effect of PRAM on the weights. For the moment, it seems wise to avoid perturbing the variables in the weighting model whenever possible, until further results are available.

Inconsistencies between variables in a record may also occur when certain combinations of scores are impossible as for example a record corresponding to a pregnant male, or a two year old that has been unemployed for the last three years. PRAM should be applied in such a way that these combinations will never occur in the perturbed file, which can always be done by applying PRAM to two or more variables simultaneously while not allowing certain transitions to occur. Apart from impossible combinations of scores, there also are unlikely combinations of scores, as for example a 16 year old person that is married. It is still an open question whether these combinations should be avoided when applying PRAM.

Some of the aforementioned problems originating from the hierarchical structure of the data-file can be dealt with using a so called normal form of a database (see e.g., [6]). In that situation however, it is still to be investigated whether PRAM can be considered to be applied independently on all the variables and in what way this will interact with statistical analyses on the perturbed file.

4 Choice of the Markov Matrix

Before applying PRAM, a number of decisions have to be made. Each decision has its consequences for the choice of the Markov matrix P that is used for PRAM. In this section several decisions with their consequences will be discussed.

The first decision that has to be made when applying PRAM, is to which variables PRAM will be applied. This depends on the criteria that the perturbed file has to satisfy in order to be called safe, as will be discussed in Section 5. Suppose for the moment that PRAM will be applied to m variables $\xi_1, ..., \xi_m$. The next step then is to partition those variables into groups $V_1, ..., V_H$ for some $1 \leq H \leq m$, in such a way that PRAM is applied to each group V_h independently of the other groups. The variables within a group are considered to be one compounded variable, and thus within a group, dependencies can be taken into account. The process of determining the Markov matrix P for this situation thus entails that a Markov matrix $P^{(h)}$ has to be determined for each group of variables V_h.

It is still an open problem how these groups should be constructed. It seems wise to construct the groups in such a way that the possible inconsistencies as described in Section 3 do not occur. It would be nice if the groups could be constructed in such a way that ξ_i

and ξ_j are independent (or only weakly correlated) whenever $\xi_i \in V_{h_1}$ and $\xi_j \in V_{h_2}$ and $h_1 \neq h_2$. If this would be the case and invariant PRAM would be applied to each group, then the simultaneous distribution of all the variables $\xi_1, ..., \xi_m$ is preserved, and not only the simultaneous distribution of the ξ_i within each group V_h, and thus it would be as if we applied invariant PRAM to $\xi_1, ..., \xi_m$.

The covariance matrix of the original data file provides some insight into the dependency structure of the variables. Multivariate techniques like e.g., factor analysis and principal component analysis can then be used in the construction of the groups of variables mentioned in the previous paragraph. Another aid in that construction may be graphical models (see [7] for an overview of such models). Research still needs to be done as to whether there exists some criterion to define an optimal partitioning of the variables. Another problem is to find an efficient algorithm to construct the groups.

Since the Markov matrix can be determined for each group V_h ($h = 1, ..., H$) separately and since the variables within a group can be considered as one compounded variable, it will be assumed without loss of generality that PRAM is applied to a single variable ξ. For this variable ξ, the transition matrix P has to be determined. First of all, it has to be decided which category can be replaced by which other categories. This determines the skeleton of the matrix, i.e., it determines whether an entry p_{kl} is zero or not.

One way to accomplish this is to partition the categories $\{1, ..., K\}$ into groups $C_1, ..., C_G$ in such a way, that each category can only be replaced by a category within the same group. It is up to the data protector to decide how these groups should be constructed. This can be done in such a way that a group always contains categories that are (in some sense) similar, which could prevent some of the inconsistencies from Section 3 from occurring, but this is by no means necessary. It can also be done in such a way that the disclosure risk (that will be described in Section 5) is minimised, given the expected number of changes in the perturbed file.

Such a partition immediately determines the structure of the Markov matrix P: the categories $1, ..., K$ can now be reordered in such a way that P can be written as

$$\begin{pmatrix} P_1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & P_G \end{pmatrix}. \tag{4}$$

The only thing that is left to be done is to choose the values of p_{kl} for k and l in the same group C_g. If general PRAM is applied, then any set of values for these p_{kl} will suffice, as long as the resulting matrix is invertible. When invariant PRAM is applied, the matrix P should also satisfy (3). It is still an open problem if the p_{kl} can be determined in some optimal way. We return to this problem in Section 6. What can be done is the following. Start with some initial choice for the p_{kl} (for example by letting $p_{kk} \geq 0.9$ for all k). With these probabilities, it can be determined whether the perturbed file is safe. If the resulting file is not safe, then the probabilities p_{kl} should be adjusted in such a way that p_{kk} decreases, implying more changes in the perturbed file. If the perturbed file is safe, then we can increase the p_{kk} (implying less changes in the perturbed file) to see if this still leads to a safe file. This process is repeated until the perturbed file is safe, while the original file is not too much perturbed.

When invariant PRAM is applied, it will in general be tedious to determine the p_{kl} in such a way that the system of equations as defined by (3) is satisfied. This can be facilitated if some special form for each block in (4) is chosen. For example, it can be assumed that within a block P_g all the off-diagonal elements in the same row are equal (implying that a category in C_g can be replaced by any other category with the same probability). Now suppose that the categories in C_g are numbered $1, ..., K_g$ in such a way that $T_\xi(k) \geq T_\xi(K_g) > 0$ for $k = 1, ..., K_g$, where $T_\xi(k)$ denotes the k-th entry of table T_ξ. If the above assumption is made and P also has to satisfy (3), then it can be computed that for $k, l \in C_g$

$$p_{kl} = \begin{cases} 1 - \theta_g T_\xi(K_g)/T_\xi(k) & \text{if} \quad k = l \\ \theta_g T_\xi(K_g)/(K_g - 1)T_\xi(k) & \text{if} \quad k \neq l \end{cases} ,$$

where $\theta_g \in (0, 1)$ is a parameter that can still be chosen. Any choice of θ_g $(g = 1, .., G)$ now leads to a matrix P for which (3) is satisfied.

A drawback of this solution is that it only allows P to have a special form. In order to allow for more general matrices, a procedure called 'two-stage PRAM' was developed. The idea behind two-stage PRAM is the following. Suppose that ξ is perturbed using an arbitrary Markov matrix P. Then given the values in the perturbed file, the probability distribution of ξ in the original file can be computed. This probability distribution can then be used to transform the perturbed file back. The twice perturbed file is generally not exactly the same as the original, but it can be seen as the result of applying invariant PRAM.

More formally, assume that a variable ξ is perturbed using any Markov matrix P. Let X denote the value of ξ in the perturbed file. Now for a given value of X in some record, the probability distribution of the original value ξ in that record can be determined. Let $p_{lk}^{\leftarrow}$ denote the probability that the original value of ξ is k, given that $X = l$. Then it is easy to compute (using Bayes formula) that

$$p_{lk}^{\leftarrow} = \mathbb{P}(\xi = k \mid X = l) = \frac{p_{kl} T_\xi(k)}{\sum_j p_{lj} T_\xi(j)} . \tag{5}$$

Let $P^{\leftarrow}$ be the matrix that has $p_{lk}^{\leftarrow}$ as its (l,k)-th entry. $P^{\leftarrow}$ is obviously again a Markov matrix. This Markov matrix can be used to apply PRAM to the perturbed file. Note that the transition probabilities $\mathbb{P}(X' = k \mid X = l)$ thus are defined to be $p_{lk}^{\leftarrow}$. That way, we obtain a file that is twice perturbed. Let X' denote the value of ξ in this twice perturbed file. Then it is easy to deduce that the probability distribution of X' is the same as that of ξ, since

$$\mathbb{P}\left(X' = k\right) = \sum_{l=1}^{K} \sum_{j=1}^{K} \mathbb{P}\left(X' = k \mid X = l\right) \mathbb{P}\left(X = l \mid \xi = j\right) \mathbb{P}(\xi = j) .$$

Substitution of the transition probabilities and re-arranging terms then yields $\mathbb{P}(X' = k) = \mathbb{P}(\xi = k)$. Furthermore, let $R = PP^{\leftarrow}$. It can be computed that R is an invariant matrix. Moreover, note that the entries of R equal

$$r_{kl} = \mathbb{P}(X' = l \mid \xi = k) .$$

If we would have applied R to the original file, then we would have obtained the same result as we now have in two steps! So starting with an arbitrary matrix P, if we then apply $P^{\leftarrow}$ to the perturbed file, we have in fact applied invariant PRAM.

Two-stage PRAM is a very convenient method if we want to implement PRAM as an additional tool in some software package for statistical disclosure control, such as e.g., μ-ARGUS (see [8]). If PRAM is added to such a package, then the data protector using the package has to determine the Markov matrix. This will be difficult, especially when invariant PRAM is applied, since the data protector (or the package) has to check a lot of restrictions on the matrix. However, if two-stage PRAM is used as an implementation of invariant PRAM, then the data protector can choose any invertible Markov matrix to start with, and the package will compute the corresponding invariant matrix!

On the other hand, when a Markov matrix P is provided the effect of the resulting invariant matrix R, defined in the above way, on the disclosure risk is not clear. This has to be investigated, since the data protector will have to fine-tune the probabilities in P to achieve the desired level of disclosure limitation. Moreover, it still has to be investigated under which conditions on the matrix P (and T_ξ) the matrix R will be invertible.

5 PRAM and Disclosure Risk

As discussed in the introduction, the process of choosing the Markov matrix with transition probabilities is intertwined with determining the level of disclosure risk. In this section a methodology will be introduced, that can be of use in determining the safety of a microdata file. The main idea is that there should be enough uncertainty as to whether a rare combination of scores in the perturbed data file is just the result of applying PRAM, or that it corresponds to the same rare combination of scores in the population.

Several methods exist in order to determine whether a microdata file is safe or not. The ideas in this section are inspired by the type of statistical disclosure rules currently in use at Statistics Netherlands, as described in [9], Chapters 4 and 5. These entail that certain rare combinations of scores on variables should not be released. The rules prescribe which combinations have to be checked. When traditional disclosure control methods like global recoding and local suppression are used, the resulting file is considered to be safe, if no rare combinations of scores exist in that file. The definition of the term 'rare' as 'being present in the data file less then a certain number of times', assumes that no measurement errors are present in the data file. A method like PRAM however, deliberately introduces artificial measurement errors to the data file. (Since PRAM is applied to categorical data, these measurement errors amount to misclassification.) Therefore, the traditional definition of 'rare' is no longer applicable when PRAM is involved. Moreover, when PRAM is applied to a microdata file the resulting file may or may not still contain rare combinations of scores, depending on the exact realisation of the process. The decision whether the perturbed file can be considered to be safe with respect to the traditional rules hence also depends on the coincidentally obtained realisation of the perturbed file. Obviously this is an undesirable situation. The safety of an originally rare combination of scores should only be determined by the transition probabilities used by PRAM. The concept of expectation ratios as described in [3] quantifies that idea.

The expectation ratio of a rare score k of a (compounded) variable ξ is related to the

posterior odds ratio of that rare score, defined in for example [10] as

$$PO(k) = \frac{\mathbb{P}(\xi = k | X = k)}{\mathbb{P}(\xi \neq k | X = k)} = \frac{p_{kk}\mathbb{P}(\xi = k)}{\sum_{l \neq k} p_{lk}\mathbb{P}(\xi = l)}, \tag{6}$$

with ξ, X and p_{lk} as defined in Section 2. In case of a self-weighting design, the probabilities $\mathbb{P}(\xi = k)$ can be estimated by $T_\xi(k)/n$, the frequency of score k in the original file. More complicated sampling designs yield more complex formulas for those probabilities and hence for their estimators. Plugging in the observed frequencies does in some sense reflect a worst case scenario as well: in case an intruder does not only have information on certain individuals in the population, but knows who has participated in the survey as well (the worst case), the frequencies in the sample are all that matters to him. To cover this case and to simplify matters, the notion of expectation ratios was introduced in Gouweleeuw et al. (1997):

$$ER(k) = \frac{p_{kk}T_\xi(k)}{\sum_{l \neq k} p_{lk}T_\xi(l)}\,. \tag{7}$$

Note that this ratio is just the ratio of the expected number of records in the perturbed file with score k equal to its score in the original file, and the expected number of record in the perturbed file with score k but with a different score in the original file. Hence, the smaller the value of the expectation ratio, the more likely it is that a record in the perturbed file with score k did not originally belong to this category, and thus the safer the perturbed file is. That is, disclosure limitation rules should imply bounds on the expectation ratios of the unsafe (i.e., rare) combinations of scores in the original file. A possible formulation of a disclosure limitation rule based on that idea is the following. Demand that, for some integer m, at least $a_j\%$ of the expectation ratios is smaller than b_j $(j = 1, \ldots, m)$ with $0 < a_1 \leq a_2 \leq \cdots \leq a_m \leq 100$ and $0 < b_1 \leq b_2 \leq \cdots \leq b_m$. In case m is taken to be equal to 1, the expectation ratios may be such that even though they are all less than e.g., 10, they all equal 9.99. To eliminate this kind of behaviour, the number of bounds m as well as the constants a_j and b_j should be chosen appropriately. This choice should be motivated by the policy a statistical agency uses in her disclosure limitation methods.

In case the Markov matrix with transition probabilities is shipped along with the perturbed microdata file, like in case of general PRAM, a possible intruder can estimate the expectation ratios and hence obtain additional information on the possibilities that a record in the perturbed file with score k, possessed the same score in the original file. Therefore the percentage a_m in the above mentioned rule should in that case be equal to 100 and the corresponding upper bound b_m should not be too high. On the other hand, in case the Markov matrix is unknown to the intruder, the $ER(k)$ can not be estimated and the rule could somewhat be relaxed in taking e.g., $a_m = 90$.

6 Additional Remarks

Policy implications

Before a statistical agency wants to apply PRAM, the way of delivering the data should be considered. If invariant PRAM is used, just furnishing the perturbed file ensures that a user is able to obtain (approximately) unbiased estimates of contingency tables.

If general PRAM is used, the user should either get the Markov matrix P, its inverse P^{-1} or its backwards version $P^{\leftarrow}$ along with the perturbed file, to be able to adjust the analyses to obtain (approximately) unbiased results. However, if $P^{\leftarrow}$ is shipped with the perturbed file, a user could use that matrix to obtain multiple versions of an invariantly perturbed microdata file, which us undesirable as discussed in the following remark (*'Multiple versions'*).

In Section 5 the disclosure limitation rules currently in use at Statistics Netherlands were the basis of the introduced methodology to determine the amount of disclosure risk. Another approach can be found in [11]. In that paper, a model is presented to estimate the probability of disclosure per record in the microdata file. The model could probably be extended to include the effect of applying PRAM and hence give the disclosure probability per record in the perturbed microdata file. That disclosure probability could then be a starting point for deriving disclosure limitation rules.

Using a notion of per record disclosure risk, it is tempting to apply PRAM only to the records with an unacceptable disclosure risk. However, additional uncertainty is introduced when PRAM is also applied to (some of) the safe records. Indeed, in Skinner's proposed disclosure risk per record (see [11]), the number of occurences in the data file of certain combinations of variables is used. That number could be influenced by applying PRAM not only to the unsafe records, but to (some of) the safe records as well.

Multiple versions
Instead of shipping the Markov matrix along with the perturbed microdata file in order to be able to estimate the variance added by applying PRAM, one might be tempted to provide the user with multiple realisations of the perturbed microdata file. In that situation the Markov matrix with transition probabilities does not need to be shipped along for variance estimation, since these multiple realisations would allow the user to estimate the added variance, even though the Markov matrix is unknown to him. However, from the point of view of disclosure limitation there is a major drawback. The transition probabilities will usually be chosen in such a way that the probability of changing the same score of a variable of a specific record, into an alternative score in more than one realisation is quite small. That is, comparing the same record in all the provided data files and picking the score that is present in the majority of them, will -with high probability- result in the true score. Hence, the disclosure limitation accomplished by PRAM is essentially eliminated in this way.

Actually, if X_k counts the number of times a specific variable in a specific record is not changed by PRAM, X_k is binomially distributed with parameters m (the number of different files) and p_{kk} (the probability that score k is *not* changed by PRAM). Hence, if $p_{kk} > 1/2$, binomial tables show that the probability that the majority of the values of that variable in the m files is indeed the true value, can get arbitrarily close to 1 by increasing m.

Alternative estimator
The definition of $P^{\leftarrow}$ as given in Section 4 as a tool in determining an invariant Markov matrix of transition probabilities, yields a solution to the problem of negatively estimated frequencies in case of general PRAM as well. Define an alternative estimator of the contingency table T_ξ by

$$\tilde{T}_\xi = \left(P^{\leftarrow}\right)^t T_X \,. \tag{8}$$

Since $P^{\leftarrow}$ is a Markov matrix itself, the estimator $\tilde{T}_\xi$ will only contain non-negative values. Moreover, using equation (1) it is easily seen that it is also an unbiased estimator of T_ξ:

$$\mathbb{E}\left(\tilde{T}_\xi \;\middle|\; \xi^{(1)}, \ldots, \xi^{(n)}\right) = \left(PP^{\leftarrow}\right)^t T_\xi = T_\xi, \tag{9}$$

where the last equality follows from the observation made in Section 4 that $PP^{\leftarrow}$ is an invariant Markov matrix. Whether the matrix $P^{\leftarrow}$ can also be used to correct for the effect of PRAM on other statistical analyses such as e.g., factor analysis, still remains to be investigated. In [12] (pp. 650ff) similar estimators were introduced to deal with misclassification of categorical data using calibration probabilities. Note that this is not surprising, since the effect of PRAM can be viewed as deliberately imposed misclassification, see remark in section 5.

Information versus disclosure limitation
In Section 4 the choice of the Markov matrix needed to apply PRAM was discussed. It was argued that this will usually result in an iterative procedure. A particular choice of the matrix induces a certain amount of disclosure limitation. If that amount is not satisfactory, that particular choice should be reconsidered and will in turn imply a certain amount of disclosure limitation. That process of choosing a matrix and determining the amount of disclosure limitation should then be iterated, until it hopefully will converge. At the moment this still amounts to 'educated guessing'. An alternative approach would be to first only assume a certain structure of the Markov matrix (e.g., a block-diagonal matrix or a band-matrix) without actually computing the exact transition probabilities. Ideally, the disclosure limitation rules, together with some measure for information loss, should then lead to transition probabilities that maximise the amount of disclosure limitation and at the same time minimise the information loss within the class of matrices defined by the chosen structure. Whether this goal is indeed attainable or not, is again a topic for further research.

Combining PRAM with other techniques
In this paper, PRAM has been considered as a disclosure control method on its own. However, it seems reasonable to combine this method with other methods like global recoding and local suppression. In that case, it is not at all clear what the implications would be on disclosure limitation rules, how an optimal mixture of disclosure control methods could be obtained, nor what the effects would be on statistical analyses.

Maintaining marginal totals
In [13] it is suggested that it is desirable to exactly maintain marginal totals of (some) contingency tables. In our approach this is not possible, since PRAM is applied on each record independently, whereas the approach in [13] implies dependency between certain records on the induced misclassifications.

When applying invariant PRAM we are able to maintain (some) marginal totals *in expectation*. From a statistical point of view, this should be sufficient since it supplies unbiased estimates of the (population) marginals, just as the original sample itself does.

If the marginals are of variables that occur in the model that produced the weights, it might be appropriate to maintain these weighted marginals. As discussed in section 3, for the moment we suggest to avoid perturbing such variables whenever possible, until further research is conducted.

References

[1] **Kooiman, P., Willenborg, L.C.R.J. and Gouweleeuw, J.M.**, 'PRAM: a method for disclosure limitation of microdata', "Research paper no. 9705", Statistics Netherlands, (1997).

[2] **Fuller, W.A.**, 'Masking procedures for microdata disclosure limitations', "Journal of Official Statistics", 9, 383–406, (1993).

[3] **Gouweleeuw, J.M., Kooiman, P., Willenborg, L.C.R.J. and Wolf, P.P. de**, 'Post Randomisation for Statistical Disclosure Control: Theory and Implementation', "Research paper no. 9731", Statistics Netherlands, (1997).

[4] **Warner, S.L.**, 'Randomized response; a survey technique for eliminating evasive answer bias', "Journal of the American Statistical Association", 57, 622–627, (1965).

[5] **Chauduri, A. and Mukerjee, R.**, 'Randomized response, theory and techniques', Marcel Dekker Inc., New York, (1988).

[6] **Ullman, J.D.**, 'Principles of Database Systems', Computer Science Press, Inc., (1982).

[7] **Whittaker, J.**, 'Graphical models in applied multivariate statistics', John Wiley and Sons, New York, (1990).

[8] **Gemerden, L. van, Wessels, A. and Hundepool, A.** 'μ-ARGUS, Users Manual Version 2.0', Statistics Netherlands, (1997).

[9] **Willenborg, L.C.R.J. and de Waal, T.** 'Statistical Disclosure Control in practice', "Lecture notes in statistics 111", Springer Verlag, New York, (1996).

[10] **Zellner, A.**, 'An introduction to Bayesian inference in Econometrics', John Wiley and Sons, New York, (1971).

[11] **Skinner, C.**, 'Estimating the re-identification risk per record in microdata', "Proceedings of the third international seminar on statistical confidentiality, Bled, 1996", (1997).

[12] **Kuha, J. and Skinner, C.**, 'Categorical data analysis and misclassification', "Survey measurement and Process quality", chapter 28, John Wiley and Sons, New York, (1997).

[13] **Duncan, G.T. and Fienberg, S.E.**, 'Obtaining Information While Preserving Privacy: A Markov Perturbation Method for Tabular Data', Current proceedings, (1998).