QÜESTIIÓ, vol. 22, 1, p. 145-156, 1998

THE POST RANDOMISATION METHOD FOR PROTECTING MICRODATA

JOSÉ GOUWELEEUW
PETER KOOIMAN
LEON WILLENBORG
PETER-PAUL DE WOLF
Statistics Netherlands

This paper describes the Post Randomisation Method (PRAM) as a method for disclosure protection of microdata. Applying PRAM means that for each record in the data file according to a specified probability mechanism the score on a number of variables is changed. Since this probability mechanism is known, the characteristics of the latent true data can unbiasedly be estimated from the observed data moments in the perturbed file.

PRAM is applied to categorical variables. It is shown that both cross-tabulation and standard multivariate analysis techniques can easily be adapted to account for PRAM. It only requires pre-multiplication by the transpose of the inverted Markov transition matrix, specifying the randomisation process. Also, estimates for the additional variance introduced by PRAM are given. By a proper choice of the transition probabilities, PRAM can be applied in such a way that certain chosen marginal distributions in the original data file are left invariant in expectation. In that case the perturbed data can be used as if it were the original data. We describe how to obtain such an invariant PRAM process. Finally, some consequences of using PRAM in practice are discussed. The present paper is a shortened version of Kooiman et al. (1997).

Keywords: Post RAndomisation Method (PRAM), disclosure, randomised response, Markov matrix, invariant matrix, perturbed data, data swapping.

* José Gouweleeuw, Peter Kooiman, Leon Willenborg and Peter-Paul de Wolf. Statistics Netherlands. Division Research and Development. Department of Statistical Methods. The Netherlands.

– Received July 1997.
– Accepted October 1997.

1. INTRODUCTION

This paper investigates a suggestion by Särndal *et al.*. (1992), pp. 572-3, to protect microdata files against disclosure by randomisation of individual record data, using the methodology of randomised response techniques due to Warner (1965). This methodology is employed when interviewers have to deal with highly sensitive questions, on which the respondent is not likely to report true values in a face-to-face interview setting. By embedding the question in a pure chance experiment the true value of the respondent is never revealed to the interviewer. By knowing the probabilities involved in the chance experiment, the analyst can nevertheless make inferences about the population frequencies of the characteristic investigated, be it with some loss in precision. The suggestion of Särndal *et al.* amounts to applying the same randomisation technique to reported individual scores, prior to their release as microdata files. Thus, true individual scores will not be revealed, whereas an analyst, by taking the (known) randomisation model into account, can still make valid inferences from the data set as a whole.

The methodology advocated in this paper represents an alternative to data swapping as a technique for disclosure protection of microdata (see Dalenius and Reiss, 1982). In data swapping individual scores on certain variables are interchanged between records in such a way that second order moments are kept more or less intact (first order moments are unchanged automatically).

In this paper we take another approach. We no longer require the perturbed file to mimic the original one. Instead we require that data moments of the original file can unbiasedly be estimated from the perturbed file. This is achieved as follows. For each record in the original microdata file, the score on one or more variables is replaced by an other score according to some probability mechanism. Because of this the moments of the data will change, the true data moments of the original file become latent. Since the probability mechanism that is used for perturbing the scores is completely known to the data protector, it can be shipped to the analyst jointly with the perturbed file. This allows the analyst to reconstruct the latent true data moments to the extent that these moments can unbiasedly be estimated from the observed data moments in the perturbed file. This method for disclosure protection of microdata will be called the Post Randomisation Method (PRAM).

In order to make valid inferences from the perturbed file the analyst has to account for the fact that the true data patterns are hidden behind a veil of errors deliberately introduced to protect the individual records. Thus, he has to apply somewhat more complicated types of statistical analyses, which is a drawback of the method in comparison with other disclosure control techniques as for example recoding or suppression of values. However, in this paper we show that this drawback is minor when PRAM is applied to categorical variables. Most disclosure protection analysis of microdata involves categorical variables. Due to the fact that the analyst knows the complete distribution of the errors, simple corrections that can be routinely applied using standard statistical software are sufficient. As compared to data swapping, the method has the important advantage that it is soundly statistical, not mechanical. This allows us to invoke the complete apparatus of statistical modelling and inference, both to make probability statements about disclosure risks at the individual record level and to calculate the loss in precision of aggregate statistics, due to PRAM.

The remainder of this paper is organised as follows. In Section 2, PRAM is introduced by an example. In Section 3, the technique is worked out for the most elementary type of analysis: cross-tabulation of categorical variables. The probability mechanism used for PRAM can be chosen in such a way that the distribution of the variables in the perturbed file equals that in the original file, in which case the perturbed file can be used directly for analysis. Section 4 describes how to construct such a probability mechanism. In Section 5 the consequences of using PRAM in practice are discussed and finally Section 6 contains some conclusions and topics for further research.

2. DISCLOSURE PROTECTION BY RANDOMISATION

Consider a $\{0, 1\}-$variable, e.g. gender with 0 = male, 1 = female. Scores are randomised, independently for each record, by using the following (known) probability matrix $P_x = \{p_{kl}\}$:

$$P_x = \begin{pmatrix} \theta_0 & 1-\theta_0 \\ 1-\theta_1 & \theta_1 \end{pmatrix},$$

where p_{kl} represents the probability that the reported randomised score equals l given that the latent true score equals k. In the sequel we denote the true score with ξ, and the randomised score with x. Assume that a data file representing a simple random sample of n records is available. We want to estimate the true population fractions of both categories. From the original file this would be achieved by calculating $T\xi/n$, $T\xi$ being the total of ξ in the data file. From the perturbed file we can similarly calculate T_x. The relation between both totals can be computed as:

$$E(T_x|\xi) = T_\xi \theta_1 + \left(n - T_\xi\right)(1-\theta_0),$$

where $E(.|.)$ denotes the conditional expectation. An unbiased estimator for T_ξ is directly obtained as

$$\hat{T}_\xi = \frac{T_x - n(1-\theta_0)}{\theta_1 + \theta_0 - 1}. \qquad (2.1)$$

Clearly

$$V\left(\hat{T}_\xi|\xi\right) = \frac{V\left(T_x|\xi\right)}{\left(\theta_1+\theta_0-1\right)^2}, \tag{2.2}$$

where $V(.|.)$ indicates the conditional variance. Note that we have to assume that $\theta_1 \neq 1-\theta_0$, since otherwise the denominator in (2.1) as well as (2.2) becomes 0. In practice, it is desirable that θ_1 is not too close to $1-\theta_0$, since a value of θ_1 close to $1-\theta_0$ would lead to a large variance as can be seen from formula (2.2).

Formula (2.2) can be rewritten as follows. Note that $T_x|\xi$ is distributed as the sum of two independent binomially distributed random variables with parameters (T_ξ, θ_1) and $(n-T_\xi\,\theta_0)$ respectively. Consequently $V(T_x|\xi) = T_\xi\,\theta_1(1-\theta_1)+(n-T_\xi)\theta_0(1-\theta_0)$, so that

$$V\left(\hat{T}_\xi|\xi\right) = \frac{T_\xi\,\theta_1(1-\theta_1)+\left(n-T_\xi\right)\,\theta_0(1-\theta_0)}{\left(\theta_1+\theta_0-1\right)^2},$$

which for $\theta_0=\theta_1(=\theta)$ reduces to

$$V\left(\hat{T}_\xi|\xi\right) = \frac{n\,\theta(1-\theta)}{(2\,\theta-1)^2}.$$

This demonstrates that under the assumptions stated the standard deviation does not depend on the true proportion T_ξ/n. Consequently the coefficient of variation $V\left(\hat{t}_\xi|\xi\right)^{1/2}/T_\xi$ is inversely proportional with T_ξ. Thus, the distortion is relatively large when the true score T_ξ (or $n-T_\xi$) is (very) low. This fits nicely into our general purpose to protect rare scores, since these are the most vulnerable to disclosure.

We conclude this section with an example indicating the effect of PRAM on protecting the individual scores. Consider a microdata file of n records, representing a simple random sample of a population of size N. The data set contains exactly one *surgeon*, whose gender is given as *female*. PRAM has been applied to the gender variable, though. Independently for each record, the gender score has been kept intact with probability 0.9, and has been changed to the opposite score with probability 0.1. Other variables in the file are not perturbed. Suppose an intruder knows that the population contains 1 female surgeon and 99 male surgeons. He can derive that the (posterior) odds are 11:1 in favour of a perturbed male surgeon in the data file. When the population contains 9 male surgeons besides the one female surgeon, the odds are 1:1. So without additional information, the intruder can not conclude that he has identified the female surgeon.

3. CROSS-TABULATION

We now turn to the problem of deriving *valid* cross-tabulations from a perturbed data file with categorical variables. The example of Section 2, which refers to the simplest non-trivial tabulation possible, i.e. a (2×1) table, will be generalised step by step. First consider a categorical variable ξ, with categories $\xi^{(k)}$, $k = 1,\ldots,K$. To protect a data file containing ξ, we perturb the scores $\xi^{(k)}$. In particular, $\xi^{(k)}$ is transformed into a score $x^{(k)} = \xi^{(l)}$ with probability p_{kl}, for $k,l = 1,\ldots,K$. The matrix $P_x = \{p_{kl}\}$ is a Markov matrix, i.e. $P_x\,\iota = \iota, \iota$ being a $(K \times 1)$ vector of ones. Since in practical situations we only want to change a small minority of the true values the diagonal elements of P_x dominate strongly (these will in general be in the range of $0.9-1.0$), so that P_x certainly has full rank. Now let T_ξ be the $(K \times 1)$ vector of frequencies of the K categories of ξ, observed in the original data file, and similarly, let T_x be the vector of frequencies in the perturbed file. It is easy to verify that

$$E\left(T_x|\xi\right) = P_x^T\,T_\xi, \tag{3.1}$$

where the superscript $'T'$ indicates transposition.

Thus T_ξ can unbiasedly be estimated by

$$\hat{T}_\xi = \left(P_x^{-1}\right)^T\,T_x.$$

Note that the matrix P_x has to be non-singular in order for $\hat{T}_\xi$ to be well defined. This implies that the matrix P_x can not contain two equal rows, which means that each value of ξ corresponds uniquely with a distribution of the perturbed variable over the categories, as determined by the rows of P_x. The conditional variance of $\hat{T}_\xi$ is given by

$$V\left(\hat{T}_\xi|\xi\right) = \left(P_x^{-1}\right)^T\,V\left(T_x|\xi\right)P_x^{-1}.$$

Due to the recordwise independence of the multinomial transition process we obtain the covariance matrix of T_x as

$$V\left(T_x|\xi\right) = \sum_{k=1}^{K} T_\xi(k)\,V_k,$$

where, for $k = 1,\ldots,K$, V_k is the $(K \times K)$ covariance matrix of the outcomes $x^{(l)}$, $l = 1,\ldots,K$, of the multinomial transition process of an element with true score $\xi^{(k)}$:

$$V_k(l,\,j) = \begin{cases} P_{x,kl}\left(1-P_{x,kl}\right) & \text{if} \quad l = j \\ -P_{x,kl}\,P_{x,kj} & \text{if} \quad l \neq j \end{cases}, \quad \text{for} \quad l,j = 1,\ldots,K. \tag{3.2}$$

Substituting the estimator $\hat{t}_\xi$ for the unknown true frequencies T_ξ we obtain an estimator for the uncertainty introduced by the noise process:

$$\hat{V}\left(T_x|\xi\right) = \sum_{k=1}^{K} \hat{T}_\xi(k)\,V_k. \tag{3.3}$$

The derivations show that univariate frequencies can straightforwardly be corrected for the noise added to the file. It just requires pre-multiplication with the transpose of the inverted transition probability matrix. This matrix can be supplied with the (perturbed) data file so that analysis only requires a matrix multiplication as an extra step in the tabulation. Using the same information it is also possible to estimate covariances of the estimated true frequencies. This result generalises to multivariate distributions (see Kooiman *et al.*, 1997).

A problem related to cross-tabulation that is of practical importance is the following. Some users of statistical microdata sets want to match extra variables to the data set supplied. We give a simplified example. Suppose the data set supplied to a client contains the dichotomous variable car *ownership* C (0: does not possess a car; 1: does possess a car) and the highly detailed geographical classification indicating *place of living* G (say, 1000 localities). The client wants to analyse the relationship between car ownership and the presence of a railway station in the place of living. For each of the 1000 localities of the geographical classification the client knows whether there is a railway station or not. When the file is not perturbed it is easy to match this data to the file and add a third *variable railway station* R (0: has no direct access to railway facilities; 1: does have direct access to railway facilities). Then, the analysis could be done by studying the 2 (2 cross classification of the variables C and R. The variable G is only used as an intermediate variable; it does not play a role in the subsequent analysis.

The question arises whether this process, which occurs frequently in practice, is still feasible when the microdata file supplied to the client has been perturbed. The answer is: yes, it is still possible, but not by matching the new variable to the (perturbed) microdata file itself. What can be done is the following. First estimate the (2×1000) table T_{CG}. Then an unbiased estimate of the true table is

$$\hat{T}_{CR} = \left(P_C^{-1}\right)^T \tilde{T}_{CG} P_G^{-1} T_{GR},$$

where $\hat{T}_{CG} = \left(P_C^{-1}\right)^t \tilde{T}_{CG} P_G^{-1}$ is the estimated $C \times G$ table.

In fact the T_{GR} table acts as an aggregator of the rows of T_{CG}. The process of matching additional variables as indicated above, therefore amounts to the addition of specific aggregation keys for detailed classificatory variables present in the data set. A similar example is occupation, which could for instance be aggregated according to the distinction between white collar and blue collar work.

So far we have demonstrated that tabular analysis of perturbed microdata sets consisting of categorical variables poses no fundamental problems. The frequency tables summarise all available information in the data set. Multivariate analysis for categorical data, like loglinear modelling or correspondence analysis, can therefore proceed from the estimates of the latent true tables. The presence of a small bit of extra variance in these estimates will generally pose no problems to the analyst, as the extra variance just adds to the sampling variance that is present in the data anyhow.

A somewhat more serious problem might be that for small frequencies in the original table the estimated frequencies can turn out to be negative. Indeed, when a true frequency is zero and its estimate is unbiased, positive and negative estimates are equally likely to occur. When the analyst is interested in the cross-tabulation per se, he might truncate estimated frequencies at zero: although a biased estimate results, the mean squared error of the cells involved decreases by the truncation. However, this procedure introduces an upward bias in the row and column totals associated with these cells.

A final problem is that subdomain analysis can not proceed as usual, as subdomains can no longer be properly identified when the domain indicator involved has been affected by PRAM. The solution is simple, though. We add the domain indicator to the set of variables under analysis, and (re)construct the relevant tables, including the domain indicator as an extra variable in the tabulations. The table entries pertaining to the subdomain of interest give unbiased estimates of the original subdomain tabulations.

4. INVARIANT PRAM

So far, it was seen that applying PRAM implies that the perturbed tables have to be pre-multiplied by $(P_x^{-1})^t$. In this section, it will be demonstrated that the analysis of the perturbed data file can be simplified if P_x is chosen to be invariant with respect to the distribution of the variable that is to be perturbed. Here distribution can refer to the distribution in the data file (the sample) as well as the distribution in the population. The consequence of this choice of P_x will be discussed in this section.

First consider the case where the data file contains one categorical variable ξ with categories $\xi^{(k)}$, $k = 1, \dots, K$. As before, $\xi^{(k)}$ is transformed into a score $x^{(k)} = \xi^{(l)}$ with probability p_{kl} for $k, l = 1, \dots, K$. The matrix $P_x = \{p_{kl}\}$ now should be chosen in such a way that the distribution of ξ over the different categories in the original data file is invariant with respect to P_x, i.e. P_x should satisfy .

(4.1) $$P_x^T T_\xi = T_\xi.$$

The $K \times K$ identity matrix I always satisfies this equation, but this is not very interesting, since the perturbed data file will be the same as the unperturbed data file. In general, there will be at least one other solution P_x for this set of equations (since there are $K(K-1)$ unknowns and K equations). For example, P_x can be chosen as follows. Assume without loss of generality that the categories are ordered in such a way that $T_\xi(k) \geqslant T_\xi(K) > 0$ for $k = 1, \dots, K$

$$p_{kl} = \begin{cases} 1 - \left(0.1\, T_\xi(K)/T_\xi(k)\right) & \text{if} \quad l = k \\ 0.1\, T_\xi(K)/T_\xi(k) & \text{if} \quad l = (k+1) \bmod K \\ 0 & \text{otherwise} \end{cases} \tag{4.2}$$

Then a simple computation shows that this matrix P_x satisfies

$$P_x^T \iota = \iota \quad \text{and} \quad P_x^T T_\xi = T_\xi.$$

There are other choices of P_x possible.

Now suppose that P_x is chosen such that (4.1) is satisfied. In that case

$$E\,(T_x|\xi) = P_x^T T_\xi = T_\xi.$$

Here the first equality follows from formula (3.1) and the second equality follows from the choice of P_x. This means that T_ξ can unbiasedly be estimated by

$$\hat{T}_\xi = T_x.$$

Note that a transformation satisfying (4.1) is indeed invariant with respect to the (sample) distribution of ξ. However this invariance does not entail that the transformation is invariant with respect to crossings of ξ with other variables in the file.

5. APPLICATION OF PRAM

Now that we have introduced PRAM, the question that remains to be considered is how to apply this technique in practice. This essentially means that a choice should be made for the Markov matrices to be applied. This involves several aspects that are discussed in separate subsections.

5.1. Markov Matrix Classes

For practical purposes it is important to study a few special classes of Markov matrices that might be considered for use in PRAM. Below we study three types of such matrices. Other choices are also possible, however. What type of Markov matrix to choose for a variable or set of variables in a particular application of PRAM depends on such things as the initial distribution to be left invariant (if any) and requirements implied by the acceptability of certain combinations of values (only if combinations of variables have to be considered in the light of integrity checks). In fact it takes some further study into the structure of these matrices so as to inform the practitioner who wants to make a motivated choice among the various possibilities.

5.1.1. Type I: two different non-zero values per row

Suppose that the number of categories for a variable on which PRAM is applied, is K. A type I Markov matrix P is one that has nonzero values on the main diagonal and for each row there is at least one other nonzero entry. Nonzero entries outside the main diagonal are all the same within a row (they may differ among rows) and they will in general not be equal to the entry on the diagonal in the same row. In fact if in row i there are $k_i - 1$ nonzero off-diagonal elements, they are all equal to $(1 - p_{ii})/(k_i - 1)$, where p_{ii} is the $i-$th diagonal element. Furthermore we require that P is non-singular, so that the inverse of P exists. Note that each row in P has two unknowns, implying that, because the row entries should add to 1, there are K unknowns. Note that the matrix that is defined in formula (4.2) is a type I matrix. In this case $k_i = 2$ for all i.

Special cases of Type I matrices are given in (5.1) below. In both cases k_i does not depend on i. In the first case $k_i = 2$, and in the second case $k_i = K$.

(4.3)

$$\begin{pmatrix} p_1 & \overline{p}_1 & 0 & \dots & \dots & 0 \\ 0 & p_2 & \overline{p}_2 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & \dots & 0 & p_{k-1} & \overline{p}_{k-1} \\ \overline{p}_k & 0 & \dots & \dots & 0 & p_k \end{pmatrix} \quad \begin{pmatrix} p_1 & \tilde{p}_1 & \dots & \dots & \tilde{p}_1 \\ \tilde{p}_2 & p_2 & \tilde{p}_2 & \dots & \tilde{p}_2 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \tilde{p}_{k-1} & \dots & \tilde{p}_{k-1} & p_{k-1} & \tilde{p}_{k-1} \\ \tilde{p}_k & \dots & \dots & \tilde{p}_k & p_k \end{pmatrix}$$

where $\overline{p}_i = (1 - p_i)$, in the first matrix, and $\tilde{p}_i = (1 - p_i)/(K - 1)$ in the second matrix, for $i = 1, \dots, K$.

5.1.2. Type II: more than two different non-zero values per row

In this case we assume well-defined functional relationships between diagonal elements and the nonzero off-diagonal elements in the same row as the respective diagonal element. In this light the Type I matrices are a subclass of the Type II matrices: each off-diagonal element is a linear function of the corresponding diagonal element, and besides these linear relationships are the same for the off-diagonal elements in the same row. This removes the tight constraint for the Type I Markov matrices namely that of the equality of the nonzero off-diagonal elements in the same row.

5.1.3. Block matrices

This is not a new type in the sense of the previous two cases, but rather a type that can be used with the blocks of e.g. Type I or of Type II. A block matrix has blocks P_j on the diagonal and all off-diagonal blocks have only zero entries. A block matrix can be used in case the vector of probabilities to be left invariant can be split into two or more parts that can, or should, be considered independent from each other. A motivation for this could be derived from the values in the probability vector to be left invariant. It might be more attractive to group these value into blocks consisting of categories for whom the associated probabilities of the probability vector to remain invariant are (almost) equal.

5.2. Disclosure Risk

We discuss the disclosure risk criterion that can be used to judge the effectiveness of a PRAM procedure. The idea behind the risk measure that we propose is based on the paradigm used by Statistics Netherlands (see Willenborg and De Waal, 1996). Suppose that we consider a particular combination of key values in a record in the perturbed file, given the PRAM techniques that have been applied to the variables whose values are considered in the combination. For this combination the set of original values that could yield it can be determined, as well as their probabilities. In fact, we may consider the combination observed in the perturbed file equivalent to the set of original combinations, together with their probabilities. To judge whether this set of combinations is safe we apply a reasoning borrowed from the case of data without measurement errors. In that case one can apply the principle that the frequency of a particular combination in the population is used to decide about the safety or unsafety of that combination. If the combination occurs more frequently than a threshold value it is considered safe, otherwise it is considered unsafe.

Instead of considering the expected frequency of the observed scores, the effectiveness of PRAM can also be judged by considering the posterior odds, as was done in the example at the end of Section 2. With posterior odds, we mean the (relative) probability that a rare score in the perturbed file corresponds with a rare score in the original file. These posterior odds should be small, so as to confuse a potential intruder. How small these odds have to be, is a topic of further research.

5.3. Information Loss

In this subsection we consider a method to quantify the information loss due to the application of PRAM. A straightforward measure is the increase of variance of the estimates due to the measurement error introduced by PRAM. Appropriate variance formulas have been derived in Sections 3 and 4. Another approach that could be used for this purpose is by considering the concept of entropy, as introduced by Shannon in communication theory in the 1940's (see Shannon and Weaver, 1949). In information theory it is assumed that a message is to be transmitted across a noisy channel, that is a communication channel that may perturb part of the message. As a result of the noisiness of the channel the message received at the other end of the channel may be distorted. The problem then is to restore the original message from the received message.

We can use the paradigm of the noisy channel through which a message is sent, to calculate the information loss in case of the application of disclosure control techniques such as PRAM[1]. It is a reasonable approach to produce a safe microdata file from an unsafe one by applying disclosure control techniques in such a way to the original microdata set that the resulting file is safe (according to the criteria applied) while the amount of information loss due to the modification is minimised.

6. CONCLUDING REMARKS

In this paper, PRAM was introduced for categorical variables. It was shown how cross-tabulations for the unperturbed variables could be unbiasedly estimated from the perturbed file and how the additional variance introduced by the PRAM process could be estimated. Furthermore, it was shown that the analysis could be simplified if an invariant matrix P_x was used for PRAM. However it remains to be studied whether it is possible to choose P_x in such a way that all cross-tabulations are (nearly) preserved. A difficulty with the PRAM method is the preservation of consistency in the data. Inconsistent data are undesirable for statistical purposes. Moreover inconsistencies in the data might give a potential intruder a clue about the values that have possibly been perturbed in a record. This knowledge might help such a person to (partly) undo the protective effect of PRAM applied to a microdata file.

A number of topics remain for further research. For instance it should be clear what Markov matrices should be used to apply PRAM: how should the entries be chosen, especially in the case where P_x is invariant with respect to the distribution of some variables: if a block matrix is used, how should the blocks be constructed. This issue is, of course, directly related to the degree of safety that one tries to achieve with a particular PRAM method applied to a microdata file, and also with the information loss that one is willing to accept.

[1]The information theory description is mainly colourful imagery. One can just as well assume that the measurement error is specified so that the observed measurements can be used to estimate the underlying, latent values. Contrary to the usual situation encountered when dealing with measurement errors in statistics, the error process in PRAM is exactly known.

Furthermore it remains to be studied what is more convenient in practice: general PRAM or invariant PRAM. The first method is easier for the data protector to apply but it requires extra work on the part of the user, whereas for the second method things are reversed. It is also possible to use a mixture of these two: a matrix which is invariant for a number of combinations in the perturbed data file, but not for all. Some cross-tabulations can then directly be derived from the perturbed data file, and for some a matrix multiplication has to be performed. In that case it is not possible to reconstruct the original microdata file from P_x.

An important topic for further research is to check which standard statistical analysis techniques survive PRAM, and how these techniques could be modified to account for the PRAM process. Techniques that can be framed in terms of second moments of the data, like regression analysis, can easily be adapted to yield consistent estimation results. However, many more statistical techniques still have to be considered.

REFERENCES

[1] **Dalenius, T.** and **S.P. Reiss** (1982). «Data-swapping: a technique for disclosure control». *Journal of statistical planning and inference,* **6**, 73–85.

[2] **Kooiman, P., L. Willenborg** and **J. Gouweleeuw** (1997). *PRAM: a method for disclosure limitation of microdata*. Report. Department of Statistical Methods, Statistics Netherlands, Voorburg.

[3] **Särndal, C.E., B. Swensson** and **J. Wretman** (1992). *Model assisted survey sampling*. Springer, New York.

[4] **Shannon, C.E.** and **W. Weaver** (1949). *The Mathematical theory of communication*. University of Illinois Press, Urbana.

[5] **Warner, S.L.** (1965). «Randomized Response: a survey technique for eliminating evasive answer bias», *Journal of the American Statistical Association*, **57**, 622–627.

[6] **Willenborg, L.** and **T. De Waal** (1996). *Statistical Disclosure Control in Practice*. Springer, New York.