Survey Methodology, December 1994
Vol. 20, No. 2, pp. 165–169
Statistics Canada

Matrix Masking Methods for Disclosure Limitation in Microdata

LAWRENCE H. COX[1]

ABSTRACT

The statistical literature contains many methods for disclosure limitation in microdata. However, their use by statistical agencies and understanding of their properties and effects has been limited. For purposes of furthering research and use of these methods, and facilitating their evaluation and quality assurance, it would be desirable to formulate them within a single framework. A framework called *matrix masking* - based on ordinary matrix arithmetic - is presented, and explicit matrix mask formulations are given for the principal microdata disclosure limitation methods in current use. This enables improved understanding and implementation of these methods by statistical agencies and other practitioners.

KEY WORDS: Statistical confidentiality; Survey data processing; Mathematical methods.

1. INTRODUCTION

In this Information Age critical activities of society are fuelled by data. Users of statistical data rely especially upon government statistical agencies to collect reliable data and disseminate it in a timely and broadly useful way. Prior to the 1950s, data were released only in printed, tabulated form. Beginning in the 1960s, data at the individual respondent level - *statistical microdata* - began to be released by the U.S. Government.

At present, use of microdata outside statistical agencies for research and policy analysis is often curtailed because appropriate data are not released to users due to confidentiality concerns. For three decades statistical agencies have wrestled with policy and technical issues in microdata release, many of which remain unresolved (Federal Committee on Statistical Methodology 1994). The purpose of this article is to present a class of matrix transformations of microdata intended to help deal with this issue.

Duncan (1990) and Duncan and Pearson (1991) characterized several disclosure limitation methods for microdata - *microdata masks* - by means of matrix addition and multiplication, and named such characterizations "matrix masks." Cox (1991) generalized the concept of matrix masks, and extended the characterization to other microdata masks. The characterization of microdata masks as matrix masks offers conceptual and statistical advantages. Matrix masking provides a simple language to represent, compare and evaluate microdata masking methods. Matrix masking expresses complicated, diverse methods in a form presentable to a wide audience including statisticians and data users, and offers a standard format to develop and optimize the efficiency of transportable microdata masking software.

In this paper, the concept of matrix masks is developed in a mathematically rigorous way. Explicit matrix mask formulations are provided for the principal microdata masking methods in current use, extending those presented in Duncan and Pearson (1991) and Cox (1991). This enables straightforward implementation of these methods in software, and facilitates closer examination and use of microdata masks by statistical agencies. This should lead to improved understanding of the properties of microdata masks and much needed understanding of their effects on data use.

2. MATRIX MASKS

2.1 Definitions

A microdata file containing p attribute values for each of n (respondent-level) data records can be represented as an $n \times p$ matrix X whose entries are denoted x_{ij}. Unless stated otherwise, X contains no missing values. A *matrix mask* (A, B, C) is a transformation of X of the form: $\tilde{X} = AXB + C$, with $A, B \neq 0$, involving ordinary matrix addition and multiplication. As A operates across the rows of X, A is called a *record transforming mask*. B is an *attribute transforming mask*, and C is a *displacing mask* (Duncan and Pearson 1991).

An *elementary matrix mask* of X is a matrix mask of the form AX, XB, or $X + C$. Iterations of (elementary) matrix masks of X are also matrix masks of X. Therefore, a matrix mask of X has the form $\tilde{X} = A\hat{X}B + C$, where either $\hat{X} = X$ or $\hat{X}$ has been obtained from X by application of a sequence of elementary matrix masks. An important advantage of this definition is to enable different statistical disclosure limitation methods to be applied selectively to arbitrary subsets of the records and attributes of X (Section 4).

[1] Lawrence H. Cox, Senior Statistician, U.S. Environmental Protection Agency, AREAL (MD-75), Research Triangle Park, NC 27711, U.S.A.

The matrices A, B, C are not necessarily fixed. For example, a common mask for numeric attributes involves addition of random noise (Tendick 1991), so that C is a random matrix. The matrices A, B, C may depend upon X. For example, to displace X by additive random noise proportional to size, draw the c_{ij} randomly from a normal distribution with mean zero and standard deviation a multiple of $| x_{ij} |$, and set $\tilde{X} = X + C$. Or, with $A = X'$, $M = AX$ is sufficient for ordinary least squares regression (Duncan and Pearson 1991).

2.2 Notation

I denotes the identity matrix. Z denotes the matrix all of whose entries are zero, and J the matrix of all ones. U_{ij} denotes the matrix all of whose entries equal zero, except $u_{ij} = 1$. I is always a square matrix; Z, J and U_{ij} need not be. The U_{ij} matrix, when used as a pre-(post-)multiplier retains the values of only one row (column) of the matrix it multiplies. The dimensions of submatrices may vary between or within individual formulations and will be specified for clarity.

3. REPRESENTATIONS OF DATA MASKS AS ELEMENTARY MATRIX MASKS

3.1 Removing and Selecting Microdata

The most intuitively obvious method for limiting disclosure is to withhold certain microdata from release to data users. Typically, these data are associated with the highest disclosure risk and may require suppressing attributes (columns) or suppressing records (rows) of X prior to release.

Attribute suppression of the k-th attribute can be represented as an attribute transforming mask $\tilde{X} = XB$, where B is the $p \times (p - 1)$ block matrix:

$$\mathbf{Supp}(k) = \begin{bmatrix} I & Z \\ Z & \\ Z & I \end{bmatrix},$$

whose upper I-matrix is of dimension $(k - 1) \times (k - 1)$, whose lower I-matrix is of dimension $(p - k) \times (p - k)$, and whose central Z-matrix is of dimension $1 \times (p - 1)$. An alternative formulation is $\mathrm{Supp}(k) = \sum_{j<k} U_{jj} + \sum_{j>k} U_{j,j-1}$.

Suppression of several attributes can be represented as a product of B-matrices of this form. For example, **Supp(*k*)Supp(*j*)** first suppresses the k-th attribute of X, and then suppresses the j-th attribute of the resulting $n \times (p - 1)$ dimensional matrix X**Supp(*k*)**. The dimensions of **Supp(*k*)** and **Supp(*j*)** are $p \times (p - 1)$ and $(p - 1) \times (p - 2)$.

It is sometimes necessary to delete individual records from X. For example, a respondent may have high identification risk, or a record may be out of scope or spurious. *Record deletion* of the h-th record can be represented as a record transforming mask $\tilde{X} = AX$, where A is an $(n - 1) \times n$ dimensional block matrix identical in structure to **Supp(*h*)**, except: the central Z-matrix of A is of dimension $(n - 1) \times 1$ and the dimensions of the upper and lower I-matrices of A are $(h - 1) \times (h - 1)$ and $(n - h) \times (n - h)$. This A-matrix is denoted **Del(*h*)**. An alternative formulation is $\mathbf{Del}(h) = \sum_{i<h} U_{ii} + \sum_{i>h} U_{i-1,i}$.

Deletion of more than one record is represented as a product of A-matrices **Del(*h*)**. For example, to delete the h-th and i-th records of X, with $i > h$, use **Del(*i* − 1) Del(*h*)**. For $i < h$, use **Del(*i*)Del(*h*)**. The dimensions of **Del(*i* − 1)** and **Del(*h*)** are $(n - 2) \times (n - 1)$ and $(n - 1) \times n$.

The A-matrix that *systematically deletes* every h-th record (for $n = rh$; r an integer) is a block matrix comprising r vertical blocks **Del(*h*)**, each of dimension $(h - 1) \times n$. This generalizes to nonsystematic deletion.

The complement of record deletion is *record sampling*. The A-matrix that systematically samples every h-th record of X, for $n = rh$, is an $r \times n$ matrix whose q-th row is the $1 \times n$ dimensional U-matrix $U_{1,qh}$. More generally, to draw a sample of size s comprising the records of X indexed by the set $S = \{s_v : v = 1, \ldots, s\}$, use the A-matrix **Sam(*X*, *S*)** of dimension $s \times n$, each row of which is a U-matrix U_{1,s_v} of dimension $1 \times n$.

3.2 Aggregating and Grouping Microdata

The risk of a respondent being identified and confidential data disclosed tends to decrease as data are more highly aggregated. *Attribute aggregation* and other microdata masks are based on this principle.

The aggregation mask that replaces the first of two attributes (the j-th attribute) by the sum of the two attributes, and deletes the second attribute (the k-th attribute) from X, for $j < k$, can be represented as an attribute transformation $\tilde{X} = XB$, where B is the $p \times (p - 1)$ dimensional block matrix:

$$\mathbf{Agg}(j,k) = \begin{bmatrix} I & Z \\ U_{1j} & \\ Z & I \end{bmatrix}.$$

The upper I-matrix of **Agg(*j*,*k*)** is of dimension $(k - 1) \times (k - 1)$, the lower I-matrix is of dimension $(p - k) \times (p - k)$, and the central U-matrix U_{1j} is of dimension $1 \times (p - 1)$. Alternative formulations are

$$\mathbf{Agg}(j,k) = \mathbf{Supp}(k) + U_{kj}, \quad \text{for } j < k, \text{ and}$$

$$\mathbf{Agg}(j,k) = \mathbf{Supp}(k) + U_{k,j-1}, \quad \text{for } j > k.$$

Aggregation-deletion over more than two attributes can be represented as a product of $\boldsymbol{B}$-matrices of this form. Construct $\boldsymbol{B}_1$ as above to aggregate the first two attributes to a subtotal, replace the first attribute by the subtotal, and delete the second attribute. Proceed iteratively forming $\boldsymbol{B}_2, \ldots, \boldsymbol{B}_{c-1}$ until all summand attributes have been incorporated into the total and deleted. Then $\boldsymbol{B} = \boldsymbol{B}_1 \cdots \boldsymbol{B}_{c-1}$.

An alternative formulation for aggregation of the j-th and k-th attributes, replacement of the j-th attribute, and deletion of the k-th attribute, is given by the $\boldsymbol{B}$-matrix product $\mathbf{Add}(j,k)\mathbf{Supp}(k)$. Aggregation and replacement of the j-th attribute without deleting the k-th attribute can be accomplished using the $p \times p$ dimensional $\boldsymbol{B}$-matrix: $\mathbf{Add}(j,k) = \boldsymbol{I} + \boldsymbol{U}_{kj}$. This generalizes to more summands v by adding more $\boldsymbol{U}_{vj}$. To create a new totals attribute (attribute $p + 1$) from the j-th and k-th attributes without replacing either attribute, form the $p \times (p + 1)$ dimensional $\boldsymbol{B}$-matrix $[\boldsymbol{I} \mid \boldsymbol{U}_{j1} + \boldsymbol{U}_{k1}]$, whose $\boldsymbol{I}$-matrix is of dimension $p \times p$, and whose right-hand submatrix is of dimension $p \times 1$. Aggregating another attribute v amounts to adding additional $\boldsymbol{U}_{v1}$ to the right-hand submatrix.

Grouping categorical data, sometimes referred to as *collapsing categories*, is representable as attribute aggregation. Represent each of the c mutually exclusive categories of a categorical variable by a column of $\boldsymbol{X}$. The absence (presence) of the corresponding trait is represented in each column by 0 (1). Grouping the c attribute categories to form one combined category is simply aggregation across the c attributes, replacing one attribute by the aggregate and deleting the remaining attributes, using $\boldsymbol{B}$-matrices in the manner described above.

It is sometimes desirable to aggregate attribute values across microrecords. For example, if microrecords can be grouped according to some notion of "similarity" (*e.g.*, age or profession, or total value of shipments or size of work force for businesses in a particular industry), then an alternative to releasing high risk microrecords is to release a microdata file whose records are *microaggregates* or *microaverages* of subsets of the original records.

Record aggregation can be performed in several ways. A typical case is to replace all summands by the corresponding totals. Assume that the records to be microaggregated are arranged consecutively, and denote the respective sizes of the record groups by $n_1, n_2, \ldots, n_s$, where $n = n_1 + n_2 + \ldots + n_s$. Microaggregation can be accomplished using a diagonal block $\boldsymbol{A}$-matrix of dimension $n \times n$. The main diagonal of $\boldsymbol{A}$ is comprised of an ordered block of square $\boldsymbol{J}$-matrices of dimension $n_v \times n_v$, $v = 1, \ldots, s$; the remaining entries of $\boldsymbol{A}$ are zero. Under microaggregation (microaveraging), original values are replaced by microaggregates (microaverages) in each record of the aggregation group. Alternatively, in each group one record may be replaced by the microaggregated record while the other records are deleted. This may be accomplished using $\boldsymbol{J}$-matrices of dimension $1 \times n_v$, in which case the dimension of $\boldsymbol{A}$ is $s \times n$. To construct microaverages in lieu of microaggregates, each $\boldsymbol{J}$-matrix is replaced by its corresponding $(\mathbf{1}/n_v)\boldsymbol{J}$.

3.3 Scrambling Record Order

A microdata file $\boldsymbol{X}$ being prepared for public use is typically derived from a larger data file (*e.g.*, by sampling) or from a more detailed file (*e.g.*, by removal of directly identifying information such as name, address, and social security number). The larger file is often maintained in a prescribed sort order, such as by geography or social security number, and $\boldsymbol{X}$ is apt to inherit this ordering. To reduce disclosure risk, the order of the microrecords of $\boldsymbol{X}$ must be *scrambled*. Record scrambling can be accomplished using a stochastic $\boldsymbol{A}$-matrix. Given a reordering of the rows (records) of $\boldsymbol{X}$ (*i.e.*, a permutation $\boldsymbol{P}$ of the row numbers $\{\mathbf{1}, \ldots, \boldsymbol{n}\}$), then for $P(i) = h$, set the i-th row of $\boldsymbol{A}$ equal to the $\boldsymbol{U}$-matrix $\boldsymbol{U}_{1h}$ of dimension $1 \times n$. $\boldsymbol{A}$ is denoted $\mathbf{Reo}(\boldsymbol{P})$. An alternative formulation is $\mathbf{Reo}(\boldsymbol{P}) = \sum_{i=1}^{n} \boldsymbol{U}_{i,p(i)}$.

3.4 Rounding and Perturbing Microdata

Data rounding is used by statistical agencies for several purposes, including disclosure limitation. Integer variables such as age or years worked, or number of children, presented exactly, could be used in combination with other information to identify respondents (Bethlehem, Keller and Pannekoek 1990). *Conventional rounding* (*e.g.*, base 5, remainders 0, 1, 2 are rounded down; remainders of 3, 4 are rounded up), does not preserve additivity to totals, and *controlled rounding*, designed to preserve additivity to totals in one and two way tabulations, may be preferred (Cox and Ernst 1982). Methods are also available for *unbiased controlled rounding* in one- or two-way tables (Cox 1987).

Data perturbation limits disclosure by introducing slight changes to microdata values. Additive perturbation amounts to adding appropriate perturbation values to original values. Additive perturbation values are often drawn randomly from a distribution with mean zero and variance small relative to that of the data. Nonrandom perturbation is also used.

Rounding and additive perturbation can be represented as displacing masks. For each value $\boldsymbol{x}_{ij}$, the displacement $\boldsymbol{c}_{ij}$ to $\boldsymbol{x}_{ij}$ is computed according to the rounding or perturbation algorithm, with $c_{ij} = 0$ for those values not subject to change. Then, $\tilde{\boldsymbol{X}} = \boldsymbol{X} + \boldsymbol{C}$ is the matrix of rounded (perturbed) values.

3.5 Attribute Topcoding

Attribute topcoding is a method by which, given a predetermined (large) value $\boldsymbol{T}_j$ of the j-th attribute, all values $x_{ij} > T_j$ are replaced by $\boldsymbol{T}_j$. Given $\boldsymbol{x}_{ij} = f_{ij}\boldsymbol{T}_j + r_{ij}$,

for f_{ij} the integer quotient, and r_{ij} the remainder, $0 \leq r_{ij} < T_j$, compute $t_{ij} = (\mathbf{Max}\{r_{ij}, (T_j + 1)^{f_{ij}} - 1\}) \bmod (T_j + 1)$. To topcode X, use the displacing mask $\mathbf{Tco}(X) = (t_{ij} - x_{ij})$.

4. REPRESENTATIONS OF DATA MASKS AS MATRIX MASKS

4.1 Selecting and Modifying Attribute-Record Combinations

The formulations of the preceding section, based on elementary matrix masks, are applied to the entire microdata file X, and do not enable selective masking of arbitrary subsets of records (rows) and/or attributes (columns) of X. The ability to selectively manipulate microdata values within subsets of X (*i.e.*, to apply data masks selectively to submatrices of X) is important for disclosure limitation purposes. This can be accomplished by combining elementary matrix masks that enable *subset selection* along rows and columns, or both, in X with elementary matrix masks as presented previously. This is accomplished in three stages.

At the first stage, apply the ignoring mask $\mathbf{Ign}(Q, R) = AXB$, where A is the $n \times n$ dimensional matrix $A = \sum_{i \in Q} U_{ii}$, and B is the $p \times p$ dimensional matrix $B = \sum_{j \in R} U_{jj}$. A leaves the values in the selected rows Q of X unchanged, and replaces all other values by zeroes; B has similar effect on the columns R. At the second stage, apply the appropriate mask or combination of masks M of Section 3 to $\mathbf{Ign}(Q, R)$ to effect the desired changes, yielding $\tilde{X} = M(\mathbf{Ign}(Q, R))$. As M is designed to change only the selected values, then all ignored values - which $\mathbf{Ign}(Q, R)$ replaced by zero - remain zero after applying M. To preserve the dimensions of $\hat{X}$, deletion operations are modified to replace values to be deleted by zero. Finally, restore the ignored original values of X by means of

$$\tilde{X} = M(\mathbf{Ign}(Q, R)) + X - \mathbf{Ign}(Q, R).$$

4.2 Blurring

When the operation M is microaveraging, the formulation of Section 4.1 provides a matrix mask for the data mask *blurring* of Strudler, Oh and Scheuren (1986).

4.3 Data Swapping

Data swapping is a method whereby selected data values are exchanged between selected sets of records, in a manner that ensures that certain one, two and higher-way tabulations remain unchanged (Dalenius and Reiss 1982). Setting $M = \mathbf{Reo}(P)$, where the swapping rule is given by a permutation P of the affected records, Section 4.1 yields a matrix mask for data swapping.

5. CONCLUDING COMMENTS

A formulation based on matrix algebra for representing the principal statistical disclosure limitation methods for microdata has been developed. Computational issues, such as for large files, are not addressed. However, the partitioning methods of Section 4.1 could be used to reduce effective computational size when working with extremely large files.

Matrix masks offer a comprehensive framework in which statistical agencies can develop, evaluate and use reliable microdata disclosure limitation software. Such software could be shared among agencies. Exploration of the uses of matrix masks by U.S. statistical agencies has been encouraged by an expert panel (Federal Committee on Statistical Methodology 1994, p. 82). The potential effect of the widespread use of matrix masks would be to standardize the microdata disclosure limitation methods available for use by agencies, while expanding each agency's options to evaluate and apply these methods.

ACKNOWLEDGEMENTS

The author is indebted to Professor George T. Duncan, Carnegie Mellon University, for introducing the concept of matrix masks and for collaborations leading to an earlier version of this paper, and to Sumitra Mukherjee, Duncan's doctoral student, for his critical reading and for developing some of the alternative formulations presented here. Preliminary research on this topic was supported in part by National Science Foundation Grant SES 91-10512. The views expressed are those of the author and are not intended to represent the policies or practices of the U.S. Environmental Protection Agency.

REFERENCES

BETHLEHEM, J.G., KELLER, W.J., and PANNEKOEK, J. (1990). Disclosure control of microdata. *Journal of the American Statistical Association*, 85, 38-45.

COX, L. (1987). A constructive procedure for unbiased controlled rounding. *Journal of the American Statistical Association*, 82, 398, 520-524.

COX, L. (1991). Comment (on Duncan, G.T. and R.W. Pearson 1991, below), *Statistical Science*, 6, 232-234.

COX, L., and ERNST, L. (1982). Controlled rounding. *INFOR*, 20, 423-432.

DALENIUS, T., and REISS, S. (1982). Data swapping: A technique for disclosure control. *Journal of Statistical Planning and Inference*, 6, 73-85.

DUNCAN, G.T. (1990). Inferential disclosure-limited microdata dissemination. *Proceedings of the Survey Research Section, American Statistical Association*, 440-445.

DUNCAN, G.T., and LAMBERT, D. (1989). The risk of disclosure for microdata. *Journal of Business and Economic Statistics*, 7, 207-217.

DUNCAN, G.T., and PEARSON, R.W. (1991). Enhancing access to microdata while protecting confidentiality: Prospects for the future. *Statistical Science*, 6, 219-239.

FEDERAL COMMITTEE ON STATISTICAL METHODOLOGY (1994). Report on disclosure limitation methodology. Statistical Policy Working Paper 22, Office of Management and Budget, Washington, DC.

STRUDLER, M., OH, L., and SCHEUREN, F. (1986). Protection of taxpayer confidentiality with respect to the tax model. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 375-381.

TENDICK, P. (1991). Optimal noise addition for preserving confidentiality in multivariate data. *Journal of Statistical Planning and Inference*, 27, 341-353.