9-2006

Disclosure Analysis for Two-Way Contingency Tables

Haibing LU

Yingjiu LI
Singapore Management University, yjli@smu.edu.sg

Xintao Wu

DOI: https://doi.org/10.1007/11930242_6

Citation

LU, Haibing; LI, Yingjiu; and Wu, Xintao. Disclosure Analysis for Two-Way Contingency Tables. (2006). *Privacy in Statistical Databases: CENEX-SDC Project International Conference, PSD 2006, Rome, Italy, December 13-15, 2006. Proceedings*. 4302, 57-67. Research Collection School Of Information Systems.
Available at: https://ink.library.smu.edu.sg/sis_research/569

Disclosure Analysis for Two-Way Contingency Tables*

Haibing Lu[1], Yingjiu Li[1], and Xintao Wu[2]

[1] Singapore Management University, 80 Stamford Road, Singapore 178902
[2] University of North Carolina at Charlotte, 9201 Univ City Blvd,
Charlotte, NC 28223
{hblu, yjli}@smu.edu.sg, xwu@uncc.edu

Abstract. Disclosure analysis in two-way contingency tables is important in categorical data analysis. The disclosure analysis concerns whether a data snooper can infer any protected cell values, which contain privacy sensitive information, from available marginal totals (i.e., row sums and column sums) in a two-way contingency table. Previous research has been targeted on this problem from various perspectives. However, there is a lack of systematic definitions on the disclosure of cell values. Also, no previous study has been focused on the distribution of the cells that are subject to various types of disclosure. In this paper, we define four types of possible disclosure based on the exact upper bound and/or the lower bound of each cell that can be computed from the marginal totals. For each type of disclosure, we discover the distribution pattern of the cells subject to disclosure. Based on the distribution patterns discovered, we can speed up the search for all cells subject to disclosure.

1 Introduction

In this paper, we focus on the disclosure problem for two-way contingency tables. The traditional disclosure problem in two-way contingency tables, which has been formulated before (e.g., in [40, 11]), asks whether a data snooper can infer accurate information about any protected cell values given the marginal totals. In this context, the internal cells of a contingency table provide privacy sensitive information, which should be protected, while the marginal totals are the sums of cell values in a row or column, which can be released to the public if they lead to no disclosure of any cell values. This problem has many practical applications such as medical/health statistics, national census, and student records management. In health insurance data, for example, it is important to protect a cell value, which represents how many times a patient undergoes a certain treatment, against being inferred from the marginal totals, which are aggregate statistics on the total number of each treatment being taken or the total number of each patient visiting doctors. For another example, in an agent-stock

* Work of Lu and Li was supported in part by SMU Research Office. Work of Wu was supported in part by USA National Science Foundation Grant IIS-0546027.

J. Domingo-Ferrer and L. Franconi (Eds.): PSD 2006, LNCS 4302, pp. 57–67, 2006.

table, where each cell indicates the volume of a stock in which an agent invests, a commercial secret may be revealed if a snooper infers from the released marginal totals that the agent buys more (or less) than certain amount of the stock.

Previous study on this problem has identified that the disclosure of any cell value depends on the upper bound and lower bound of the cell value which a snooper can derive from the available marginal information (e.g., see [21,22,6]). If the upper bound is the same as the lower bound, the cell value is exposed. Likewise, if the difference between the upper bound and the lower bound is very small, the security of the table is also considered to be compromised [40]. However, there is a lack of systematic definitions on the disclosure of cell values. Also, no previous study has been focused on the distribution of the cells that are subject to various types of disclosure. In this paper, we define four types of possible disclosure based on the exact upper bound and/or the lower bound of each cell that can be computed from the marginal totals. For each type of disclosure, we discover the distribution pattern of the cells subject to disclosure. Based on the distribution patterns discovered, we propose two efficient methods to speed up the search for all cells subject to disclosure.

The rest of this paper is organized as follows. Section 2 presents the preliminaries for the research of disclosure analysis. Section 3 defines various types of disclosure that are commonly used in practice. Section 4 reveals in a contingency table the distribution patterns of the cells that are subject to different types of disclosure. Based on the distribution patterns discovered, Section 5 investigates how to efficiently detect all cells subject to disclosure. Section 6 reviews the related work. Finally, Section 7 concludes the paper.

2 Preliminaries

A two-way contingency table A with m rows and n columns is denoted by $\{a_{ij} \mid 1 \leq i \leq m, 1 \leq j \leq n\}$, where $a_{ij} \geq 0$. In tradition, the cell values in a contingency table are usually assumed to be nonnegative integers (e.g., counts). We extend this assumption such that the cell values can be nonnegative real numbers. The results given in this paper hold in both integer domain and real domain.

Denote $a_{+j} = \sum_{i=1}^{m} a_{ij}$, $a_{i+} = \sum_{j=1}^{n} a_{ij}$, and $a_{++} = \sum_{ij} a_{ij}$, where a_{+j} and a_{i+} are *marginal totals* and a_{++} is the *grand total*. The marginal totals satisfy $\sum_{j=1}^{n} a_{+j} = \sum_{i=1}^{m} a_{i+} = a_{++}$, which is called the *consistency condition*.

The marginal totals of a two-way contingency table can be released while the cell values are protected. A traditional disclosure analysis question in a two-way contingency table asks [40,11]: *Can any information about protected cells be inferred from the released marginal totals?* The answer to this question depends on the bounds that a snooper can obtain about a protected cell from the marginal totals that are given [21,22,6].

A nonnegative value $\underline{a}_{ij}$ is said to be a *lower bound* of cell value a_{ij} if, for any contingency table $\{a'_{ij}\}$ that has the same marginal totals as A, the inequality $\underline{a}_{ij} \leq a'_{ij}$ holds. A value $\underline{a}_{ij}$ is said to be the *exact lower bound* of a_{ij} if (i) it is a lower bound; and (ii) there exists a contingency table $\{a'_{ij}\}$ such that a) the marginal totals of A' are the same as those of A, and b) $a'_{ij} = \underline{a}_{ij}$. An upper bound or the exact upper bound $\overline{a}_{ij}$ can be defined similarly.

Definition 2.1. (Fréchet bounds) *Given marginal totals $\{a_{+j}\}$ and $\{a_{i+}\}$ of a two-way contingency table A, the Fréchet bounds for any cell value a_{ij} are*

$$\max\{0, a_{i+} + a_{+j} - a_{++}\} \leq a_{ij} \leq \min\{a_{i+}, a_{+j}\}$$

The Fréchet bounds are exact bounds as proven in [13]. Therefore, the Fréchet bounds give a data snooper the “best” estimate of a protected cell from the marginal totals.

3 Disclosure Types

Based on the exact bounds, we define four types of information disclosure in two-way contingency tables: existence disclosure, threshold upward disclosure, threshold downward disclosure, and approximation disclosure.

Definition 3.1. Existence disclosure: *The exact lower bound of a protected cell is positive.*

The concept of existence disclosure can be illustrated using a patient-treatment table. In such table, each cell shows the number of times that a patient undergoes a particular treatment. To protect each patient’s privacy, only the marginal totals are released. However, from the marginal totals, a snooper can easily calculate the exact lower bound of each cell. If an exact lower bound is positive, the snooper may infer that a patient has suffered from certain disease. This type of disclosure is common in privacy protection of statistical data.

Definition 3.2. Threshold upward disclosure: *The exact lower bound of a protected cell is greater than a positive threshold.*

Definition 3.3. Threshold downward disclosure: *The exact upper bound of a protected cell is less than a positive threshold.*

The threshold upward disclosure is similar to the existence disclosure with the difference that the threshold is a positive value rather than zero, while the threshold downward disclosure is a dual to the threshold upward disclosure. In certain applications, knowing that a cell value is positive is not harmful, while knowing that the cell value is greater or less than certain threshold is dangerous. For example, in an agent-stock contingency table, where each cell indicates the volume of certain stock in which an agent invests, it is often trivial if a snooper deduces that an agent invests in certain stock, but a commercial secret may be revealed if the snooper infers that the agent buys more (or less) than certain amount of the stock. These types of disclosure often occur in business and wealth related tables.

Definition 3.4. Approximation disclosure: *The difference between the exact lower bound and the exact upper bound of a protected is less than a positive threshold.*

This type of disclosure is defined based on not only the exact lower bound but also the exact upper bound. If the difference between the two exact bounds for a protected cell is small enough, one can estimate the cell's value with a high precision. For example, if one knows that a professor's salary is between 90K and 92K, then the actual salary amount is largely revealed.

Among the four types of disclosure, the definitions of existence disclosure and approximation disclosure summarize the similar concepts discussed in some previous papers (e.g., [40,11,34]). To be more systematic, we extend these concepts to threshold upward disclosure and threshold downward disclosure.

4 Distribution of Cells Subject to Disclosure

In the previous section, we have defined four types of information disclosure in a contingency table. In this section, we study the distribution of the cells that are subject to various types of disclosure. For the first time we discover that the cells subject to disclosure demonstrate some regular patterns.

Theorem 4.1. *Consider existence disclosure or threshold upward disclosure with a fixed threshold in a two-way contingency table. The cells subject to disclosure, if exist, must appear in the same row or column, but not both.*

Proof. Prove by contradiction. Assume there exist two cells $a_{i_1j_1}$ and $a_{i_2j_2}$ subject to existence disclosure and $i_1 \neq i_2, j_1 \neq j_2$. Then

$$a_{i_1+} + a_{+j_1} - a_{++} > 0$$
$$a_{i_2+} + a_{+j_2} - a_{++} > 0$$

These two inequalities lead to $a_{i_1+} + a_{i_2+} - a_{++} - \sum_{j \neq j_1, j_2} a_{+j} > 0$. A contradiction is committed as $a_{i_1+} + a_{i_2+} - a_{++} - \sum_{j \neq j_1, j_2} a_{+j} \leq 0$ must hold (note that $a_{i_1+} + a_{i_2+} - a_{++} \leq 0$). Thus, the theorem is proven for the existence disclosure. Since any cell subject to threshold upward disclosure must also be subject to existence disclosure, the theorem is proven for the threshold upward disclosure. ♢

The above theorem reveals the distribution pattern for the cells that are subject to existence disclosure or threshold upward disclosure. This pattern can be used to limit the search for cells subject to existence disclosure or threshold disclosure, which we will discuss in the next section.

Now consider the distribution of the cells that are subject to threshold downward disclosure or approximation disclosure. The following lemma compares the difference of the exact bounds for any cells that are subject to existence disclosure with that for any cells that are not subject to existence disclosure.

Lemma 4.1. *The difference of the exact bounds for any cell that is subject to existence disclosure is no less than that for any cell that is not subject to existence disclosure in a two-way contingency table.*

Proof. Assume $a_{i_1j_1}$ is subject to existence disclosure. The difference of its exact bounds is

$$min\{a_{i_1+}, a_{+j_1}\} - (a_{i_1+} + a_{+j_1} - a_{++}) = min\{\sum_{i \neq i_1} a_{i+}, \sum_{j \neq j_1} a_{+j}\}$$

Consider any other cell $a_{i_2j_2}$ that is not subject to existence disclosure. Because the exact lower bound of $a_{i_2j_2}$ is zero, the difference of its exact bounds is $min\{a_{i_2+}, a_{+j_2}\}$. To prove the theorem, we need to prove

$$min\{a_{i_2+}, a_{+j_2}\} \leq min\{\sum_{i \neq i_1} a_{i+}, \sum_{j \neq j_1} a_{+j}\}$$

We prove this in three possible cases: (i) $i_2 \neq i_1, j_2 \neq j_1$, (ii) $i_2 \neq i_1, j_2 = j_1$, and (iii) $i_2 = i_1, j_2 \neq j_1$. Clearly, the inequality holds for case (i). In the following, we prove the theorem for case (ii) only. The proof for case (iii) is similar to case (ii).

In case (ii), let $j_1 = j_2 = j'$. Since $i_1 \neq i_2$, we have $a_{+j'} = a_{i_1j'} + a_{i_2j'} + \sum_{i \neq i_1, i_2} a_{ij'}$ and $a_{i_2+} = a_{i_2j'} + \sum_{j \neq j'} a_{i_2j}$. Because $a_{i_1j'}$ is subject to existence disclosure, we have $a_{i_1+} + a_{+j'} - a_{++} = a_{i_1j'} - \sum_{i \neq i_1, j \neq j'} a_{ij} > 0$; then, we have $a_{i_1j'} > \sum_{i \neq i_1, j \neq j'} a_{ij} \geq \sum_{j \neq j'} a_{i_2j}$. Therefore, we have $a_{+j'} > a_{i_2+}$. Since $a_{i_2j'}$ is not subject to existence inference, the difference of the exact bounds for $a_{i_2j'}$ is a_{i_2+}. To prove the theorem, we need to prove $a_{i_2+} \leq min\{\sum_{i \neq i_1} a_{i+}, \sum_{j \neq j'} a_{+j}\}$.

On the one hand, it is clear $a_{i_2+} \leq \sum_{i \neq i_1} a_{i+}$. On the other hand, since $a_{i_2j'}$ is not subject to existence disclosure, we have $a_{i_2j'} \leq \sum_{i \neq i_2, j \neq j'} a_{ij}$. Adding $\sum_{j \neq j'} a_{i_2j}$ to both sides of this inequality, we have $a_{i_2+} \leq \sum_{j \neq j'} a_{+j}$. The theorem is proven. ♢

From this lemma, one can easily derive the following

Lemma 4.2. *The exact upper bound for any cell that is subject to existence disclosure is no less than that for any cell that is not subject to existence disclosure in a two-way contingency table.*

According to the above lemmas, we have the following theorem regarding the distribution of cells subject to approximation disclosure or threshold downward disclosure.

Theorem 4.2. *Consider approximation disclosure or threshold downward disclosure with a fixed threshold in a two-way contingency table. If a cell is subject to disclosure, the other cells in the same row or column must also be subject to disclosure.*

Proof. First, consider approximation disclosure with a fixed threshold $\tau > 0$. If a cell $a_{i'j'}$ is subject to approximation disclosure, the difference of its exact bounds is less than τ. The theorem is proven in the following two cases.

Case (i): $a_{i'j'}$ is also subject to existence disclosure. From Theorem 4.1, we know that all of the cells that are subject to existence disclosure must be in row i' or column j', but not both. Without loss of generality, we assume that these cells are in row i. Therefore, all of the cells in the column j' except $a_{i'j'}$ are not subject to existence disclosure. According to the Lemma 4.1, the differences of the exact bounds for these cells in column j' are smaller than or equal to the difference of the exact bounds for a_{ij}, which is less than τ. Therefore, all of the cells in column j' are subject to approximation disclosure. The theorem is proven.

Case (ii): $a_{i'j'}$ is not subject to existence disclosure. According to Theorem 4.1, all of the cells in row i' and column j' are not subject to existence disclosure. Since $a_{i'j'}$ is subject to approximation disclosure, we have $min\{a_{i'+}, a_{+j'}\} < \tau$. To prove the theorem, we prove that all of the cells in either row i' or column j' are subject to approximation disclosure. If $min\{a_{i'+}, a_{+j'}\} = a_{i'+} < \tau$, then for any cell $a_{i'j}$ where $j \neq j'$, the difference of the exact bounds for $a_{i'j}$ is $min\{a_{i'+}, a_{+j}\} \leq a_{i'+} < \tau$. Thus, all of the cells in row i' is subject to approximation disclosure. Similarly, if $min\{a_{i'+}, a_{+j'}\} = a_{+j'}$, all of the cells in column j' are subject to approximation disclosure.

Then consider the threshold downward disclosure with a fixed threshold. The theorem can be proven similarly as in the case of approximation disclosure. The only difference is that one needs to replace the phrase "approximation disclosure" with "threshold downward disclosure", "the difference of the exact bounds" with "the exact upper bound", and "lemma 4.1" with "lemma 4.2" in the proof. ♢

Note that the distribution pattern for the cells that are subject to approximation disclosure or threshold downward disclosure is different from that for the cells that are subject to existence disclosure or threshold upward disclosure. The former pattern is a single row or column, but not both, while the latter must "fill" some rows or columns.

5 Disclosure Detection

An important task in contingency table protection is to detect all cells that are subject to disclosure before one can eliminate such disclosure using some disclosure limitation method. We consider disclosure detection in this section, while disclosure limitation will be summarized in the related work section.

A naive approach to disclosure detection is to check all cells one by one. To check whether a cell is subject to disclosure, one needs to compute its Fréchet lower bound (two plus/minus operations and one comparison operation) and/or Fréchet upper bound (one comparison operation), depending on what type of disclosure is of concern. This naive approach requires checking all mn cells in an $m \times n$ contingency table.

We improve this naive approach by reducing its time complexity from $O(mn)$ to $O(m+n)$. Such an improvement is meaningful in practice especially for some information organizations (e.g., statistical offices) which routinely process a large number of sizable contingency tables.

First, consider the existence disclosure and the threshold upward disclosure. According to Theorem 4.1, the cells subject to disclosure must exist in a single row or column, but not both. Based on this distribution pattern, we propose the following

Procedure 1. (Disclosure detection for existence disclosure or threshold upward disclosure)

1. Discover all i' and j' such that $a_{i'+} = \max_i\{a_{i+}\}$ and $a_{+j'} = \max_j\{a_{+j}\}$; proceed to step (2) if $a_{i'j'}$ is subject to disclosure; otherwise, output no cell subject to disclosure.
2. Check all cells in row i'. If no cell is subject to disclosure, continue checking all cells in column j'. Output all cells subject to disclosure that are discovered in both step (1) and step (2).

If there exists at least one cell subject to existence disclosure or threshold upward disclosure, $a_{i'j'}$ must be one of such cells since the exact upper bound of any other cell is less than or equal to the exact upper bound of $a_{i'j'}$. According to this fact and the distribution pattern, it is easy to know that this procedure outputs all and only the cells that are subject to existence disclosure or threshold upward disclosure.

Second, consider the threshold downward disclosure and the approximation disclosure. According to Theorem 4.2, the cells subject to disclosure must "fill" some rows or columns. Based on this distribution pattern, we propose the following

Procedure 2. (Disclosure detection for threshold downward disclosure)

1. Discover all i' and j' such that $a_{i'+} < \tau$ and $a_{+j'} < \tau$.
2. Output all cells in the discovered rows i' and columns j' to be subject to disclosure.

For threshold downward disclosure with threshold τ, a cell $a_{i'j'}$ is subject to disclosure if and only if its marginal total $a_{i'+}$ or $a_{+j'}$ is less than τ. According to this fact and the distribution pattern, it is easy to know that the above procedure outputs all and only the cells that are subject to threshold downward disclosure.

For approximation disclosure with threshold τ, one can classify those cells that are subject to disclosure into two categories: (i) cells that are subject to threshold downward disclosure with threshold τ, and (ii) cells that are not subject to threshold downward disclosure with threshold τ. It is clear that the cells in category (ii) must be subject to existence disclosure (and approximation disclosure). Procedure 2 can be used to discover all and only the cells in category (i), while procedure 1 can be easily extended to discover all and only the cells in category (ii). The union of the cells discovered in categories (i) and (ii) is the set of cells subject to approximation disclosure.

6 Related Work

The problem of protecting sensitive data (e.g., privacy related information) against disclosure from nonsensitive data (e.g., aggregations) has long been a focus in statistical database research [1, 19, 46, 24, 26]. The proposed techniques can be roughly classified into restriction-based and perturbation-based. The restriction-based techniques limit the disclosure of privacy information by posing restrictions on queries [5, 45, 44], including the number of values aggregated in each query [19], the common values aggregated in different queries [20], and the rank of a matrix representing answered queries [10]. Other restriction-based techniques include partition [9,39], microaggregation [25,46], suppression and generalization [14, 13, 31, 43], and k-anonymity privacy protection [37, 41, 47]. The perturbation-based techniques protect/distort sensitive private data by adding random noises without affecting the use of data significantly. The random noises can be added to data structures [38], query answers [4], or source data [42,2,3,35,8,36]. Recently, however, people have discovered that the original sensitive data can be estimated accurately from the perturbed data [32,30], indicating that the perturbation-based techniques should be examined carefully in practice so as to protect sensitive data effectively.

For protecting contingency tables, people have developed various techniques including cell suppression, controlled rounding, and controlled tabular adjustment. Cell suppression is applied to suppress any sensitive cells as well as other appropriately selected cells so as to prevent inference to sensitive cells from marginal totals [14,17,28,29]. The challenge is to provide sufficient protection while minimizing the amount of information loss due to suppression [27].

Controlled rounding is another disclosure limitation method which rounds each cell value in a contingency table to adjacent integer multiples of a positive integer base [16,15,7]. It requires that the sum of the rounded values for any row or column be equal to the rounded value of the corresponding marginal total. The controlled round can be customized for limiting various types of disclosure.

Controlled tabular adjustment (or synthetic substitution) [18] uses threshold rules to determine how cells should be modified. It replaces a sensitive cell value by a “safe” value (e.g., either zero or a threshold value) and uses linear programming to make small adjustments to other cells so as to restore the tabular structure. Similar to the controlled rounding method, this method requires that some cell values be modified, thus introducing errors to the protected data.

Our study on the disclosure analysis is complementary to the previous study on disclosure limitation. To apply any disclosure limitation method, one needs to first discover all cells that are subject to disclosure. Rather than applying a naive brute-force approach, we investigate the distribution patterns for the cells subject to disclosure and, based on the patterns, propose efficient methods to speedup the searching process significantly.

Parallel to the development of data protection techniques for two-way tables, an active line of research deals with protecting multiway contingency tables or “cubes.” It has been known that the Fréchet bounds, after being extended to high-dimensional space, may not necessarily be the exact bounds [13]. Recent studies have been focused on estimating the exact bounds [13, 12, 6, 33] or giving the exact bounds in some special cases [21, 22, 23]. Once the exact bounds are given, our definitions on various types of disclosure can be easily extended to multiway contingency tables. The challenge is that the distribution patterns discovered in two-way contingency tables may not hold in high dimensions. Therefore, it deserves further study on multiway contingency tables.

7 Conclusion

The major contribution of this paper can be summarized as follows. Firstly, we defined four types of disclosure for evaluating the disclosure of cell values in contingency tables. Secondly, for each type of disclosure, we discovered the distribution patterns for the cells subject to disclosure in a two-way contingency table. The discovery of the distribution patterns is important as it enables us to speed up the search for all cells subject to disclosure. In the future, we plan to extend our study to multiway contingency tables. The major challenge in multiway contingency tables is that the Fréchet bounds may not be exact bounds in general. Some recent efforts have been made to approach the exact bounds beyond two-dimensions [21, 22, 33].

References

1. N. R. Adam and J. C. Wortmann. Security-control methods for statistical databases: a comparative study. *ACM Computing Surveys*, 21(4):515–556, 1989.
2. Dakshi Agrawal and Charu C. Aggarwal. On the design and quantification of privacy preserving data mining algorithms. In *PODS*, 2001.
3. Rakesh Agrawal and Ramakrishnan Srikant. Privacy-preserving data mining. In *SIGMOD Conference*, pages 439–450, 2000.
4. Leland L. Beck. A security mechanism for statistical databases. *ACM Trans. Database Syst.*, 5(3):316–338, 1980.
5. Alexander Brodsky, Csilla Farkas, and Sushil Jajodia. Secure databases: Constraints, inference channels, and monitoring disclosures. *IEEE Trans. Knowl. Data Eng.*, 12(6):900–919, 2000.
6. L. Buzzigoli and A. Giusti. An algorithm to calculate the lower and upper bounds of the elements of an array given its marginals. In *Proceedings of the conference for statistical data protection*, pages 131–147, 1999.
7. B. D. Causey, L. H. Cox, and L. R. Ernst. Applications of transportation theory to statistical problems. *Journal of the American Statistical Association*, 80:903–909, 1985.
8. Keke Chen and Ling Liu. Privacy preserving data classification with rotation perturbation. In *ICDM*, pages 589–592, 2005.
9. Francis Y. L. Chin and Gultekin Özsoyoglu. Statistical database design. *ACM Trans. Database Syst.*, 6(1):113–139, 1981.
10. Francis Y. L. Chin and Gultekin Özsoyoglu. Auditing and inference control in statistical databases. *IEEE Trans. Software Eng.*, 8(6):574–582, 1982.
11. S. Chowdhury, G. Duncan, R. Krishnan, S. Roehrig, and S. Mukherjee. Disclosure detection in multivariate categorical databases: auditing confidentiality protection through two new matrix operators. *Management Sciences*, 45:1710–1723, 1999.
12. L. Cox. Bounding entries in 3-dimensional contingency tables. In *SDC: From Theory to Practice*, 2001. http://vneumann.etse.urv.es/amrads/papers/coxlux.pdf.

13. L. Cox. On properties of multi-dimensional statistical tables. *Journal of Statistical Planning and Inference*, 117(2):251–273, 2003.
14. L. H. Cox. Suppression methodology and statistical disclosure control. *Journal of American Statistical Association*, 75:377–385, 1980.
15. L. H. Cox. A constructive procedure for unbiased controlled rounding. *Journal of the American Statistical Association*, 82:520–524, 1987.
16. L. H. Cox and J. A. George. Controlled rounding for tables with subtotals. *Annuals of operations research*, 20(1-4):141–157, 1989.
17. Lawrence H. Cox. Network models for complementary cell suppression. *Journal of the American Statistical Association*, 90:1453–1462, 1995.
18. R. A. Dandekar and L. H. Cox. Synthetic tabular data: An alternative to complementary cell suppression. Manuscript available from URL http://mysite.verizon.net/vze7w8vk/.
19. D. E. Denning and J. Schlorer. Inference controls for statistical databases. *IEEE Computer*, 16(7):69–82, 1983.
20. David P. Dobkin, Anita K. Jones, and Richard J. Lipton. Secure databases: Protection against user influence. *ACM Trans. Database Syst.*, 4(1):97–106, 1979.
21. A. Dobra and S. E. Fienberg. Bounds for cell entries in contingency tables given fixed marginal totals and decomposable graphs. *Proceedings of the National Academy of Sciences of the United States of America*, 97(22):11885–11892, 2000.
22. A. Dobra and S. E. Fienberg. Bounds for cell entries in contingency tables induced by fixed marginal totals with applications to disclosure limitation. *Statistical journal of the united states*, 18:363–371, 2001.
23. A. Dobra, A. Karr, and A. Sanil. Preserving confidentiality of high-dimensional tabulated data: Statistical and computational issues. *Statistics and Computing*, 13:363–370, 2003.
24. Josep Domingo-Ferrer. Advances in inference control in statistical databases: An overview. In *Inference Control in Statistical Databases*, pages 1–7, 2002.
25. Josep Domingo-Ferrer and Josep Maria Mateo-Sanz. Practical data-oriented microaggregation for statistical disclosure control. *IEEE Trans. Knowl. Data Eng.*, 14(1):189–201, 2002.
26. Csilla Farkas and Sushil Jajodia. The inference problem: A survey. *SIGKDD Explorations*, 4(2):6–11, 2002.
27. M. Fischetti and J. Salazar. Solving the cell suppression problem on tabular data with linear constraints. *Management sciences*, 47(7):1008–1027, 2001.
28. M. Fischetti and J. J. Salazar. Solving the cell suppression problem on tabular data with linear constraints. *Management Sciences*, 47:1008–1026, 2000.
29. M. Fischetti and J. J. Salazar. Partial cell suppression: a new methodology for statistical disclosure control. *Statistics and Computing*, 13:13–21, 2003.
30. Zhengli Huang, Wenliang Du, and Biao Chen. Deriving private information from randomized data. In *SIGMOD Conference*, pages 37–48, 2005.
31. Vijay S. Iyengar. Transforming data to satisfy privacy constraints. In *KDD*, pages 279–288, 2002.
32. Hillol Kargupta, Souptik Datta, Qi Wang, and Krishnamoorthy Sivakumar. On the privacy preserving properties of random data perturbation techniques. In *ICDM*, pages 99–106, 2003.
33. Yingjiu Li, Haibing Lu, and Robert H. Deng. Practical inference control for data cubes (extended abstract). In *IEEE Symposium on Security and Privacy*, 2006.
34. Yingjiu Li, Lingyu Wang, and Sushil Jajodia. Preventing interval-based inference by random data perturbation. In *Privacy Enhancing Technologies*, pages 160–170, 2002.

35. Kun Liu, Hillol Kargupta, and Jessica Ryan. Random projection-based multiplicative data perturbation for privacy preserving distributed data mining. *IEEE Trans. Knowl. Data Eng.*, 18(1):92–106, 2006.
36. K. Muralidhar and R. Sarathy. A general aditive data perturbation method for database security. *Management Sciences*, 45:1399–1415, 2002.
37. P. Samarati and L. Sweeney. *Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppression.* Technical report, SRI International. 1998.
38. Jan Schlörer. Security of statistical databases: Multidimensional transformation. *ACM Trans. Database Syst.*, 6(1):95–112, 1981.
39. Jan Schlörer. Information loss in partitioned statistical databases. *Comput. J.*, 26(3):218–223, 1983.
40. Bernd Sturmfels. Week 1: Two-way contingency tables, 2003. John von Neumann Lectures 2003 at the Technical University München. http://www-m10.mathematik.tu-muenchen.de/neumann/lecturenotes/neumann_week1.pdf.
41. L. Sweeney. Achieving k-anonymity privacy protection using generalization and suppression. *International Journal on Uncertainty, Fuzziness and Knowledge-based Systems*, 10(5):571–588, 2002.
42. J. F. Traub, Yechiam Yemini, and Henryk Wozniakowski. The statistical security of a statistical database. *ACM Trans. Database Syst.*, 9(4):672–679, 1984.
43. Ke Wang, Philip S. Yu, and Sourav Chakraborty. Bottom-up generalization: A data mining solution to privacy protection. In *ICDM*, pages 249–256, 2004.
44. Lingyu Wang, Sushil Jajodia, and Duminda Wijesekera. Securing olap data cubes against privacy breaches. In *IEEE Symposium on Security and Privacy*, pages 161–175, 2004.
45. Lingyu Wang, Yingjiu Li, Duminda Wijesekera, and Sushil Jajodia. Precisely answering multi-dimensional range queries without privacy breaches. In *ESORICS*, pages 100–115, 2003.
46. L. Willenborg and T. de Walal. *Statistical Disclosure Control in Practice.* Springer Verlag, 1996.
47. Chao Yao, Xiaoyang Sean Wang, and Sushil Jajodia. Checking for k-anonymity violation by views. In *VLDB*, pages 910–921, 2005.