Statistical Journal of the United Nations ECE 18 (2001) 363–371
IOS Press

Bounds for cell entries in contingency tables induced by fixed marginal totals with applications to disclosure limitation

Adrian Dobra[a] and Stephen E. Fienberg[b]
[a]*Department of Statistics, Carnegie Mellon University, Pittsburgh, PA 15213-3890, USA*
Tel.: +1 412 268 7829; Fax: +1 412 268 7828; E-mail: adobra@stat.cmu.edu
[b]*Department of Statistics and Center for Automated Learning and Discovery, Carnegie Mellon University, Pittsburgh, PA 15213-3890, USA*
Tel.: +1 412 268 2723; Fax: +1 412 268 7828; E-mail: fienberg@stat.cmu.edu

Abstract. We describe new results for upper and lower bounds on the entries in multi-way tables of counts based on a set of released and possibly overlapping marginal tables which have practical importance for assessing disclosure risk. In particular, we present a generalized version of the shuttle algorithm proposed by Buzzigoli and Giusti that is proven to compute sharp integer bounds for an arbitrary set of fixed marginals. The method forms part of a project developing a Web-based query system for statistical databases. Its goal is to allow the use of disclosure limitation methods in response to a series of queries in which the public knowledge of releases is cumulative.

Keywords: Statistical disclosure control, log-linear models, decomposable models, reducible models, integer programming

1. Introduction

The National Institute of Statistical Sciences has recently assembled a team of statistical researchers from multiple universities who, working with statisticians in US statistical agencies, are developing a Web-based query system for statistical databases. Their goal is a system that allows the use of disclosure limitation methods (e.g., see [18,19]) applied sequentially in response to a series of statistical queries in which the public knowledge of releases is cumulative (c.f. a pilot project described in [13]). The idea is to fully automate recent methods for disclosure limitation, intruder behavior (c.f. [9]) and alternative approaches to risk assessment.

Consider a database consisting of a k-way contingency table, for which the queries come in the form of requests for marginal tables. What is intuitively clear from statistical theory is that, as margins are released and cumulated by users, there is increasing information available about the table entries. In response to a new query, the system now examines it in combination with all those previously released margins and decides if the risk of disclosure of individuals in the full unreleased table is too great. Then it might offer one of three responses: (1) yes – release; (2) no – don't release; or perhaps (3) simulate a new table, which is consistent with the previously released margins, and then release the requested margin table from it (c.f. [1]). Because released margins need to be consistent and even simulated, releases become highly constrained.

How might such a system evaluate the risk of disclosure from the release of a new margin? A number of researchers have recently been working on the problem of determining upper and lower bounds on the cells of the cross-classification given a set of margins. This is in one sense an old problem (at least for two-way tables) but it is also deeply linked to recent mathematical statistical developments and has generated a flurry of new research (e.g., see Buzzigoli and Giusti [2], Cox [3], Fienberg [8], and Roehrig et al. [17]). Here we outline some recent results on this problem due to Dobra and Fienberg [6] and Dobra [4,5] and we illustrate our methodology on an example.

2. Technical background

Upper and lower bounds induced by some fixed set of marginals on the cell entries of a contingency table are of great importance in measuring the disclosure risk associated with the release of these marginal totals, e.g., see the various papers in the 1993 and 1998 special issues of The Journal of Official Statistics, as well as the Proceedings of the Statistical Data Protection Conference, Lisbon 1998. The classes of bounds we are concerned with also appear in a number of other contexts such as mass transportation problems. Fréchet originally described bounds on cell counts in cross-classifications of positive counts in terms of cumulative distribution functions (c.d.f. henceforth). If we normalize each entry in a two-dimensional table by dividing it by the grand total, then adding up the appropriate proportions obtained in this way, we end up with the c.d.f. Bonferroni and Hoeffding independently developed related results on bounds. Until recently, the efforts of solving this bound problem have been largely focused on the situation when the fixed marginals are non-overlapping [8], but our interest is in the cases when the margins being fixed are multidimensional and overlapping, in which case consistency constraints have to be imposed [12].

Any contingency table with non-negative integer entries and fixed marginal totals is a lattice point in the convex polytope Q defined by the linear system of equations induced by the released marginals. The constraints given by the values in the released marginals induce upper and lower bounds on the interior cells of the initial table. These bounds or feasibility intervals can be obtained by solving the corresponding linear programming problems. The importance of systematically investigating these linear systems of equations should be readily apparent. If the number of lattice points in Q is below a certain threshold, we have significant evidence that a potential disclosure of the entire dataset might have occurred. Moreover, if the induced upper and lower bounds are too tight or too close to the actual sensitive value in a cell entry, the information associated with the individuals classified in that cell may become public knowledge.

The problem of determining sharp upper and lower bounds for the cell entries subject to some linear constraints expressed in this form is known to be NP-hard [16]. Several approaches have been proposed for computing bounds. However, almost all of them have drawbacks that show the need for alternate solutions. Network models need formal structure to work even for 3-way tables and besides there is no general formulation for higher-way tables. The most natural method for solving linear programming problems is the simplex method. In this case we would have to run the procedure twice for every element in the table and consequently we would ignore the underlying dependencies among the marginals by regarding the maximization/minimization problem associated with some cell as unrelated to the parallel problems associated with the remainder of the cells in the table. Although the simplex method works well for small problems and dimensions, by employing it we would ignore the special structure of the problem because we would consider every table as a linear list of cells. The computational inadequacy of the simplex approach is further augmented by the fact that we may get fractional bounds [3], which are very

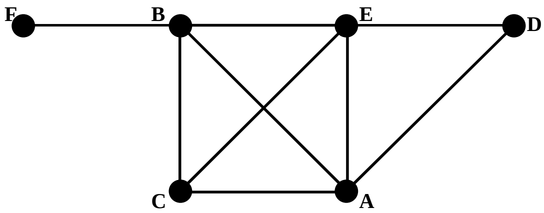

Fig. 1. Independence graph induced by the marginals [BF], [ABCE] and [ADE].

difficult to interpret. To avoid fractional bounds, one would have to make use of integer programming algorithms, but their computational complexity prevent their usage even for problems of modest size. These considerations suggest the need for more specialized, computationally inexpensive algorithms that could fully exploit the special structure of the problem we are dealing with.

Agencies often employ disclosure limitation methods such as cell suppression and data swapping. The object of both methods is to create a replacement table for the genuine unsafe marginal. Both table protection methods preserve a given set of marginal totals that were previously released. Log-linear models (see [1]) are the most usual way of representing and studying contingency tables with fixed marginals, and Fienberg et al. [11] and Fienberg [8] have demonstrated the clear links between log-linear models and disclosure limitation techniques. Throughout this paper, we exploit log-linear models theory to identify special settings amenable to alternative and more efficient techniques for determining sharp bounds. In particular, when the released marginals are the minimal sufficient statistics (MSS henceforth) of a decomposable log-linear model, we are able to express the upper and lower bounds as explicit functions of marginal totals [6]. We extend our results to more general structures for which we can considerably reduce the computational effort required to solve the linear problems.

3. New results on bounds

We visualize the dependency patterns induced by the released marginals by constructing an independence graph for the variables in the underlying cross-classification. Each variable cross-classified in the table is associated with a vertex in this graph. If two variables are not connected, they are conditionally independent given the remainder. Models described solely in terms of such conditional independencies are said to be graphical (e.g., see [14]). The data in Table 1 come from a prospective epidemiological study of 1841 workers in a Czechoslovakian car factory, as part of an investigation of potential risk factors for coronary thrombosis (see [7]). Assume we are provided with three marginal tables [BF], [ABCE], and [ADE] of this 6-way table. These are the marginals corresponding to a graphical model whose independence graph is given in Fig. 1. In order to reach A, D or E starting from F, we have to go through B or C, hence F is independent of A, E, D given B and C. This means that [BC] is a separator of the graph. In the same way, B, C, F are independent of D given A and E, thus [AE] is also a separator of the graph.

Decomposable graphical models have closed form structure and special properties. The expected cell values can be expressed as a function of the fixed marginals. To be more explicit, the maximum likelihood estimates are the product of the marginals divided by the product of the separators. By induction on the number of MSSs, in Dobra and Fienberg [6], we developed generalized Fréchet bounds for decomposable log-linear models with any number of MSSs. These Fréchet bounds are sharp in the

Table 1
Autoworkers data. Source: Edwards and Havranek [7]

F	E	D	C	B	no		yes	
				A	no	yes	no	yes
neg	< 3	< 140	no		44	40	112	67
			yes		129	145	12	23
		$\geqslant 140$	no		35	12	80	33
			yes		109	67	7	9
	$\geqslant 3$	< 140	no		23	32	70	66
			yes		50	80	7	13
		$\geqslant 140$	no		24	25	73	57
			yes		51	63	7	16
pos	< 3	< 140	no		5	7	21	9
			yes		9	17	1	4
		$\geqslant 140$	no		4	3	11	8
			yes		14	17	5	2
	$\geqslant 3$	$\geqslant 140$	no		7	3	14	14
			yes		9	16	2	3
		$\geqslant 140$	no		4	0	13	11
			yes		5	14	4	4

A, smoking; B, strenuous mental work; C, strenuous physical work; D, systolic blood pressure; E, ratio of β and α lipoproteins; F, family anamnesis of coronary heart disease.

sense that they are the tightest possible bounds given the marginals, and, in addition, we can determine feasible tables for which these bounds are attained.

Theorem 1 (*Generalized Fréchet Bounds for Decomposable Models*). Assume that the released set of marginals for a k-way contingency table is the set of MSSs of a decomposable log-linear model. Then the upper bounds for the cell entries in the initial table are the minimum of relevant margins, while the lower bounds are the maximum of zero, or sum of the relevant margins minus the separators.

For example, the upper bounds for the cell entries in Table 1 induced by the marginals [BF], [ABCE], and [ADE] are the minimum of the corresponding entries in the fixed marginals, while the lower bounds are the sum of the same entries minus the sum of the corresponding entries in the marginals associated with the separators of the independence graph, i.e., [B] and [AE]. We give these bounds in Table 2.

When the log-linear model associated with the released set of marginals is not decomposable, it is natural to ask ourselves whether we could reduce the computational effort needed to determine the tightest bounds by employing the same strategy used for decomposable graphs, i.e. decompositions of graphs by means of complete separators. An independence graph that is not necessarily decomposable, but still admits a proper decomposition, is called *reducible* [15]. Once again, we point out the link with maximum likelihood estimation in log-linear models. We define a reducible log-linear model [6] as one for which the corresponding MSSs are marginals that characterize the components of a reducible independence graph. If we can calculate the maximum likelihood estimates for the log-linear models corresponding to every component of a reducible graph G, then we can easily derive explicit formulae for the maximum likelihood estimates in the reducible log-linear model with independence graph G [6]. We state the result, but postpone explaining how to use it for the moment.

Theorem 2 (*Generalized Fréchet Bounds for Reducible Models*). Assume that the released set of marginals is the set of MSSs of a reducible log-linear model. Then the upper bounds for the cell entries in the initial table are the minimum of upper bounds of relevant components, while the lower bounds are the maximum of zero, or sum of the lower bounds of relevant components minus the separators.

Table 2
Bounds for autoworkers data given the marginals [BF], [ABCE], [ADE]

F	E	D	C	B	no		yes	
				A	no	yes	no	yes
neg	< 3	< 140	no		[0,88]	[0,62]	[0,224]	[0,117]
			yes		[0,261]	[0,246]	[0,25]	[0,38]
		$\geqslant 140$	no		[0,88]	[0,62]	[0,224]	[0,117]
			yes		[0,261]	[0,151]	[0,25]	[0,38]
	$\geqslant 3$	< 140	no		[0,58]	[0,60]	[0,170]	[0,148]
			yes		[0,115]	[0,173]	[0,20]	[0,36]
		$\geqslant 140$	no		[0,58]	[0,60]	[0,170]	[0,148]
			yes		[0,115]	[0,173]	[0,20]	[0,36]
pos	< 3	< 140	no		[0,88]	[0,62]	[0,126]	[0,117]
			yes		[0,134]	[0,134]	[0,25]	[0,38]
		$\geqslant 140$	no		[0,88]	[0,62]	[0,126]	[0,117]
			yes		[0,134]	[0,134]	[0,25]	[0,38]
	$\geqslant 3$	< 140	no		[0,58]	[0,60]	[0,126]	[0,126]
			yes		[0,115]	[0,134]	[0,20]	[0,36]
		$\geqslant 140$	no		[0,58]	[0,60]	[0,126]	[0,126]
			yes		[0,115]	[0,134]	[0,20]	[0,36]

4. A new algorithm

When the independence graph corresponding to a set of released marginals is not reducible, the Fréchet-like inequalities presented in the preceding section do not produce sharp bounds, and we must employ an iterative procedure. The main drawback of the existing iterative procedures such as the simplex method is that they offer no guarantee that the resulting bounds will be the best integer bounds. Again, the statistical literature on contingency tables helps us. Up until now, we were able to visualize the dependencies induced among the variables cross-classified in a table of counts by the set of fixed marginals by constructing the related independence graph. Nevertheless, if all (k-1)-dimensional marginal tables are given, the corresponding independence graph is complete, hence the line of reasoning we followed so far is ineffective in this setting. The log-linear model of no (k-1)-order interaction is not graphical, and the only way we could compute the maximum likelihood estimates associated with it is through some iterative method such as the iterative fitting procedure [1]. By exploiting the intrinsic conditional independence relationships, we were able to considerably reduce or even completely eliminate the need of employing an iterative procedure in the situations we studied before, but in this particular context there are no such relationships to exploit. Buzzigoli and Giusti [2] proposed what they called the shuttle algorithm for computing the upper and lower bounds induced by the (k-1)-way marginals on the cell entries of a k-way table. Their procedure alternates iteratively between upper and lower bounds, but does not always converge to the sharpest bounds possible [3].

Following Fienberg [8], we note that, if the table is dichotomous, the log-linear model of no kth-order interaction has only one degree of freedom, thus we can uniquely express the counts in any cell as a function of one single fixed cell alone. By imposing the non-negativity constraints for every cell in our contingency table, we are then able to derive sharp upper and lower bounds. Therefore there is no need for employing iterative methods in this case. It turns out that dichotomous tables are the key to derive sharp bounds for an arbitrary k-way table.

In order to capture the underlying dependencies induced among the cell counts in a k-way table, we consider the set S of all possible dichotomous tables obtained by collapsing the original k-way table, not only across variables, but also across categories within variables. The categories associated with a variable are divided in two groups; hence we replace every variable with a dichotomous one, and end up with a dichotomous k-way table. We let T be the set containing the cells of all dichotomizations S of the original table, formed by collapsing the initial table in every possible way. Therefore, if we fix a set of marginals, we are able to state the bounds problem in a new equivalent form: "Find the bounds T^U and T^L for the cells T given that we know the values of some cells $T_0 \subset T$".

It is not hard to see that the upper and lower bounds for the cells in T are interlinked, i.e., bounds for some cells in T induce bounds for some other cells in T. The dependencies induced by the fixed marginals among the cells in T can be expressed as "two-cell" dependencies defined as follows. Let t_1 and t_2 in T such that their join t_{12} is still in T. Then upper and lower bounds for the cells t_1 and t_2 translate into upper and lower bounds for their join t_{12}:

$$t_1^L + t_2^L \leqslant t_{12} = t_1 + t_2 \leqslant t_1^U + t_2^U.$$

Similarly, bounds for t_2 and t_{12} translate into bounds for t_1:

$$t_{12}^L - t_2^U \leqslant t_1 = t_{12} - t_2 \leqslant t_{12}^U - t_2^L.$$

We iterate thorough these "two-cell" dependencies until the upper bounds for the cells in T no longer decrease and the lower bounds no longer increase. The careful reader will notice that this is actually a generalized version of the Buzzigoli-Giusti shuttle algorithm. We "know" a cell if the current upper bound is equal to the current lower bound. As we attempt to adjust the bounds so that the two-cell dependencies are simultaneously satisfied, the feasibility interval for every cell will shrink, hence the set of "known" cells T_0 will get larger. Unfortunately, the bounds we end up with are not necessarily sharp, except in: (i) the decomposable case, and (ii) the case of a dichotomous k-way table with all (k-1)-way marginals fixed. To be more explicit, if the marginals we fix are the MSSs of a decomposable log-linear model, the bounds calculated by the generalized shuttle algorithm will coincide with the bounds obtained by making use of Theorem 1, whereas in case (ii), the generalized shuttle algorithm will successfully determine the best integer bounds by expressing any cell as a function of any other cell, and then imposing the non-negativity conditions on these constraints. There may be other situations when the generalized shuttle algorithm will converge to the best integer bounds, but further research is needed to identify them.

For the general k-way bound problem, we need to "correct" the resulting bounds by constructing feasible integer tables for which those bounds are actually attained. We explore the space Q by repeatedly assigning values to the cells in the original table. We do not perform an exhaustive search of Q since we immediately adjust the upper and lower bounds for the remaining cells in T once we picked a value for a cell entry, and consequently the values we attempt to assign to a particular cell are chosen from the current feasibility interval associated with that entry.

A combination of assigned values is inconsistent if, given that we know t_1, t_2 and t_{12} such that the join of t_1, t_2 is t_{12}, we have $t_1 + t_2 \neq t_{12}$. Once we encountered an inconsistency, we stop and attempt to assign new values to the cells we previously fixed. We successfully determined a feasible table if we managed to pick a value for all the cells in the initial table. The algorithm we described will therefore be able to identify any combination of marginals that does not correspond to an integer feasible table, i.e., it will be able to highlight the cases when the convex polytope defined by the fixed marginals does not contain any lattice points. Further technical details can be found in Dobra [5].

Table 3
Marginal [ABCE] of autoworkers data and bounds for this marginal given all two-dimensional totals

E	C	B	no		no		B	no		yes	
		A	no	yes	no	yes	A	no	yes	no	yes
< 3	No		88	62	224	117		[0,206]	[0,167]	[0,404]	[0,312]
											∴ [0,312.67]
	Yes		261	246	25	38		[0,421]	[0,463]	[0,119]	[0,119]
$\geqslant 3$	No		58	60	170	148		[0,181]	[0,167]	[0,363]	[0,339]
	Yes		115	173	20	36		[0,314]	[0,344]	[0,119]	[0,119]

Table 4
Marginal [AED] of autoworkers data and bounds for this marginal given all two-dimensional totals

E	D	A	no	yes	A	no	yes
< 3	no		333	312		[182,515]	[130,463]
	yes		265	151		[83,416]	[0,333]
$\geqslant 3$	no		182	227		[0,333]	[76,409]
	yes		181	190		[30,363]	[8,341]

5. Example reconsidered

To clarify the concepts and results presented so far, we again make use of autoworkers data in Table 1. This time, however, we assume that the fixed marginals are [BF], [BC], [BE], [AB], [AC], [AE], [CE], [DE], [AD]. Note that the independence graph associated with this set of marginals is the same in Fig. 1 since the log-linear model whose MSSs correspond to those marginals is not graphical. Hence every component of the independence graph is not necessarily associated with a single minimal sufficient statistic, but possibly with two or more MSSs.

The independence graph in Fig. 1 decomposes in three components, [BF], [ABCE], and [ADE], and two separators, [B] and [AE]. The first component, [BF], is assumed fixed, hence there is nothing to be done. The other two components are not fixed, however, and we need to compute upper and lower bounds for each of them. By making use of the generalized shuttle algorithm, we computed bounds for the cell entries in the marginal [ABCE] given the marginals [BC], [BE],[AB], [AC], [AE], [CE] (see Table 3). We did the same for the marginal [ADE] given the marginals [AE], [DE], [AD] (see Table 4).

If we were to compute bounds for the marginal [ABCE] by employing the simplex method, we would obtain a fractional upper bound. In Table 3, this fractional upper bound is indicated by the symbol "∴". Fractional bounds are exactly the situations when the simplex approach fails to correctly solve the integer programming problem of interest. Although one might argue that these situations rarely occur, we have no way to know beforehand when this phenomenon will take place. Since we have upper and lower bounds for each of the components of a reducible graph, Theorem 2 now allows us to piece together the bounds for the components [BF], [ABCE] and [ADE] to obtain sharp integer bounds for the original 6-way table – see Table 5. Again, we note that the simplex method would have failed to compute sharp integer bounds for two cells in the table.

We emphasize that Theorem 2 is a sound technique for replacing the original problem, namely, computing bounds for a 6-way table, by two smaller ones, i.e., computing bounds for a 4-way and a 3-way table. The computational effort required for implementing and using Theorem 2 is ignorable, and thus exploiting it in this fashion could lead to appreciable computational savings.

Table 5
Bounds for the autoworkers data given the margins [BF], [BC], [BE],[AB], [AC], [AE], [CE], [DE], [AD]

F	E	D	C	B: no		B: yes	
			A	no	yes	no	yes
Neg	< 3	< 140	no	[0,206]	[0,167]	[0,404]	[0,312]
							$\therefore$ [0,312.67]
			yes	[0,421]	[0,463]	[0,119]	[0,119]
		$\geqslant 140$	no	[0,206]	[0,167]	[0,404]	[0,312]
							$\therefore$ [0,312.67]
			yes	[0,416]	[0,333]	[0,119]	[0,119]
	$\geqslant 3$	< 140	no	[0,181]	[0,167]	[0,333]	[0,339]
			yes	[0,314]	[0,344]	[0,119]	[0,119]
		$\geqslant 140$	no	[0,181]	[0,167]	[0,363]	[0,339]
			yes	[0,314]	[0,341]	[0,119]	[0,119]
Pos	< 3	< 140	no	[0,134]	[0,134]	[0,126]	[0,126]
			yes	[0,134]	[0,134]	[0,119]	[0,119]
		$\geqslant 140$	no	[0,134]	[0,134]	[0,126]	[0,126]
			yes	[0,134]	[0,134]	[0,119]	[0,119]
	$\geqslant 3$	< 140	no	[0,134]	[0,134]	[0,126]	[0,126]
			yes	[0,134]	[0,134]	[0,119]	[0,119]
		$\geqslant 140$	no	[0,134]	[0,134]	[0,126]	[0,126]
			yes	[0,134]	[0,134]	[0,119]	[0,119]

6. Conclusions

In this paper we have shown how log-linear model statistical theory can help identify situations when explicit formulas exist for computing the best integer bounds on the entries of a cross-classification of arbitrary dimension given a set of marginal totals. When such formulas do not exist, we illustrated how to derive similar formulas for reducing the computational effort. In addition, we explained how log-linear models provide the basis for correcting the shuttle algorithm originally proposed by Buzzigoli and Giusti, and transform it into a general procedure for computing sharp integer bounds given any set of marginals. The generalized shuttle algorithm described here simultaneously computes sharp integer bounds for the cells in T by fully exploiting the structure of the bounds problem for multi-way contingency tables and, in addition, it can update the bounds, as more marginals are being released.

Preparation of this paper was supported in part by the National Science Foundation under Grant EIA-9876619 to the National Institute of Statistical Sciences.

References

[1] Y.M.M. Bishop, S.E. Fienberg and P.W. Holland, *Discrete Multivariate Analysis: Theory and Practice,* M.I.T. Press, Cambridge, MA, 1975.

[2] L. Buzzigoli and A. Giusti, An algorithm to calculate the lower and upper bounds of the elements of an array given its marginals, in *Statistical Data Protection* (*SDP'98*) *Proceedings,* Eurostat, Luxembourg, 1999, pp. 131–147.

[3] L.H. Cox, Some remarks on research directions in statistical data protection, in *Statistical Data Protection* (*SDP'98*) *Proceedings,* Eurostat, Luxembourg, 1999, pp. 163–176.

[4] A. Dobra, Measuring the disclosure risk for multi-way tables with fixed marginals corresponding to decomposable log-linear models, Tech. Rep., Department of Statistics, Carnegie Mellon University, 2000.

[5] A. Dobra, Computing bounds for entries in contingency tables given a set of fixed marginals, Tech. Rep., Department of Statistics, Carnegie Mellon University, 2000.

[6] A. Dobra and S.E. Fienberg, Bounds for cell entries in contingency tables given marginal totals and decomposable graphs, *Proceedings of the National Academy of Sciences* **97** (2000), 11885–11892.

[7] D.E. Edwards and T. Havranek, A fast procedure for model search in multidimensional contingency tables, *Biometrika* **72** (1985), 339–351.
[8] S.E. Fienberg, Fréchet and Bonferroni bounds for multi-way tables of counts with applications to disclosure limitation, in *Statistical Data Protection* (*SDP'98*) *Proceedings,* Eurostat, Luxembourg, 1999, pp. 115–129.
[9] S.E. Fienberg, E.U. Makov and A.P. Sanil, A Bayesian approach to data disclosure: optimal intruder behavior for continuous data, *Journal of Official Statistics* **13** (1997), 75–89.
[10] S.E. Fienberg, E.U. Makov, M.M. Meyer and R.J. Steele, Computing the exact distribution for a multi-way contingency table conditional on its marginal totals, in *Data Analysis from Statistical Foundations: Papers in Honor of D.A.S. Fraser,* A. Saleh, ed., Nova Science Publishing, 2001.
[11] S.E. Fienberg, E.U. Makov and R.J. Steele, Disclosure limitation using perturbation and related methods for categorical data, *Journal of Official Statistics* **14** (1998), 485–502.
[12] H. Joe, *Multivariate Models and Dependence Concepts,* Chapman & Hall, New York, 1997.
[13] S. Keller-McNulty and E.A. Unger, A database system prototype for remote access to information based on confidential data, *Journal of Official Statistics* **14** (1998), 347–360.
[14] S.L. Lauritzen, *Graphical Models,* Clarendon Press, Oxford, 1996.
[15] H.G. Leimer, Optimal decomposition by clique separators, *Discrete Mathematics* **113** (1993), 99–123.
[16] S.F. Roehrig, Auditing disclosure in multi-way tables with cell suppression: simplex and shuttle solutions, in *American Statistical Association* (*ASA'99*) *Proceedings,* 1999.
[17] S.F. Roehrig, R. Padman, G.T. Duncan and R. Krishnan, Disclosure detection in multiple linked categorical datafiles: a unified network approach, in *Statistical Data Protection* (*SDP'98*) *Proceedings,* Eurostat, Luxembourg, 1999, pp. 149–162.
[18] L. Willenborg and T. de Waal, *Statistical Disclosure Control in Practice,* (Vol. 111), Lecture Notes in Statistics, Springer-Verlag, New York, 1996.
[19] L. Willenborg and T. de Waal, *Elements of Statistical Disclosure Control,* (Vol. 155), Lecture Notes in Statistics, Springer-Verlag, New York, 2000.