Bounds on Entries in 3-Dimensional Contingency Tables Subject to Given Marginal Totals

Lawrence H. Cox

U.S. National Center for Health Statistics, 6525 Belcrest Road
Hyattsville, MD 20782 USA
lcox@cdc.gov

Abstract: Problems in statistical data security have led to interest in determining exact integer bounds on entries in multi-dimensional contingency tables subject to fixed marginal totals. We investigate the 3-dimensional integer planar transportation problem (3-DIPTP). Heuristic algorithms for bounding entries in 3-DIPTPs have recently appeared. We demonstrate these algorithms are not exact, are based on necessary but not sufficient conditions to solve 3-DIPTP, and that all are insensitive to whether a feasible table exists. We compare the algorithms and demonstrate that one is superior, but not original. We exhibit fractional extremal points and discuss implications for statistical data base query systems.

1 Introduction

A problem of interest in operations research since the 1950s [1] and during the 1960s and 1970s [2) is to establish sufficient conditions for existence of a feasible solution to the *3-dimensional planar transportation problem (3-DPTP)*, viz., to the linear program:

$$\sum_{i=1}^{d_1} n_{ijk} = n_{+jk}, \quad \sum_{j=1}^{d_2} n_{ijk} = n_{i+k}, \quad \sum_{k=1}^{d_3} n_{ijk} = n_{ij+}, \quad n_{ijk} \geq 0 \ , \tag{1}$$

where n_{+jk} , n_{i+k} , $n_{ij+} \geq 0$ are constants, referred to as the *2-dimensional marginal totals*. Attempts on the problem are summarized in [3]. Unfortunately, each has been shown [2, 3] to yield necessary but not sufficient conditions for feasibility. In [4] is given a sufficient condition for multi-dimensional transportation problems based on an iterative nonlinear statistical procedure known as *iterative proportional fitting*. The purpose here is to examine the role of feasibility in the pursuit of exact integral lower and upper bounds on internal entries in a 3-DPTP subject to integer constraints (viz., a *3-DIPTP*) and to describe further research directions. Our notation suggests that internal and marginal entries are integer. Integrality is not required for 3-DPTP (1), nor by the feasibility procedure of [4]. However, henceforth integrality of all entries is assumed as we focus on *contingency tables* - tables of nonnegative integer frequency counts and totals - and on the 3-DIPTP.

J. Domingo-Ferrer (Ed.): Inference Control in Statistical Databases, LNCS 2316, pp. 21 - 33, 2002.

The *consistency conditions* are necessary for feasibility of 3-DIPTP:

$$\sum_{k=1}^{d_3} n_{i+k} = \sum_{j=1}^{d_2} n_{ij+}, \sum_{i=1}^{d_1} n_{ij+} = \sum_{k=1}^{d_3} n_{+jk}, \sum_{i=1}^{d_1} n_{i+k} = \sum_{j=1}^{d_2} n_{+jk}, \quad n_{ijk} \geq 0. \tag{2}$$

The respective values, n_{i++} , n_{+j+} , n_{++k}, are the *1-dimensional marginal totals*. $n_{+++} = \sum_{i=1}^{d_1} n_{i++} = \sum_{j=1}^{d_2} n_{+j+} = \sum_{k=1}^{d_3} n_{++k}$ is the *grand total*. It is customary to represent the *2-dimensional marginal totals* in matrices defined by elements:

$$A_{jk} = \sum_{i=1}^{d_1} n_{ijk}\ ,\ B_{ik} = \sum_{j=1}^{d_2} n_{ijk}\ ,\ C_{ij} = \sum_{k=1}^{d_3} n_{ijk}, \qquad n_{ijk} \geq 0. \tag{3}$$

The *feasibility problem* is the existence of integer solutions to (1) subject to consistency (2) and integrality conditions on the 2-dimensional marginal totals. The *bounding problem* is to determine integer lower and upper bounds on each entry n_{ijk} over contingency tables satisfying (1)-(2). *Exact bounding* determines the interval $[\min\{n_{ijk}\}, \max\{n_{ijk}\}]$, over all integer feasible solutions $n^* = \{n^*_{ijk}\}$ of (1)-(2).

The bounding problem is important in statistical data protection. To prevent unauthorized disclosure of confidential subject-level data, it is necessary to thwart narrow estimation of small counts. In lieu of releasing the internal entries of a 3-dimensional contingency table, a national statistical office (NSO) may release only the 2-dimensional marginals. An important question for the NSO is then: how closely can a third party estimate the suppressed internal entries using the published marginal totals? During large-scale data production such as for a national census or survey, the NSO needs to answer this question thousands of times.

Several factors can produce an *infeasible table*, viz., marginal totals satisfying (1)-(2) for which no feasible solution exists, and that infeasible tables are *ubiquitous* and *abundant*, viz., dense in the set of all potential tables [4]. To be useful, bounding methods must be sensitive to infeasibility, otherwise meaningless data and erroneous inferences can result [5].

The advent of public access statistical data base query systems has stimulated recent research by statisticians on the bounding problem. Unfortunately, feasibility and its consequences have been ignored. We highlight and explore this issue, through examination of four papers representing separate approaches to bounding problems. Three of the papers [6-8] were presented at the International Conference on Statistical Data Protection (*SDP'98*), March 25-27, 1998, Lisbon, Portugal, sponsored by the Statistical Office of the European Communities (EUROSTAT). The fourth [9] appeared in *Management Science* and reports current research findings. A fifth, more recent, paper [10] is also discussed. We examine issues raised and offer observations and generalizations.

2 The F-Bounds

Given a 2-dimensional table with consistent sets of column ($\{n_{+j}\}$) and row ($\{n_{i+}\}$) marginal totals, the *nominal upper bound* for n_{ij} equals $\min\{n_{+j}, n_{i+}\}$. The *nominal lower bound* is zero.

It is easy to obtain exact bounds in 2-dimensions. The nominal upper bound is exact, by the *stepping stones algorithm*: set n_{ij} to its nominal upper bound, and subtract this value from the column, row and grand totals. Either the column total or the row total (or both) must become zero: set all entries in the corresponding column (or row, or both) equal to zero and drop this column (or row, or both) from the table. Arbitrarily pick an entry from the remaining table, set it equal to its nominal upper bound, and continue. In a finite number of iterations, a completely specified, consistent 2-dimensional table exhibiting the nominal upper bound for n_{ij} will be reached. Exact lower bounds can be obtained as follows.

As $n_{+j} + n_{i+} - n_{++} = n_{ij} - \sum_{I \neq i,\ J \neq j} n_{IJ}$, then

$n_{ij} \geq \max\{0, n_{+j} + n_{i+} - n_{++}\}$. That this bound is exact follows from observing that $\sum_{I \neq i,\ J \neq j} n_{IJ} = 0$ is feasible if

$n_{+j} + n_{i+} - n_{++} \geq 0$. Therefore, in 2-dimensions, exact bounds are given by:

$$\min\{n_{+j}, n_{i+}\} \geq n_{ij} \geq \max\{0, n_{+j} + n_{i+} - n_{++}\}. \tag{4}$$

These bounds generalize to m-dimensions, viz., each internal entry is contained in precisely m(m-1)/2 2-dimensional tables, each of which yields a candidate lower bound. The maximum of these lower bounds and zero provides a lower bound on the entry. Unlike the 2-dimensional case, in m $\geq$ 3 dimensions these bounds are not necessarily exact [5]. We refer to these bounds as the *F-bounds*. In 3-dimensions, the F-bounds are:

$$\begin{aligned}&\min\{n_{+jk}, n_{i+k}, n_{ij+}\} \geq n_{ijk} \\ &\geq \max\{0, n_{ij+}+n_{i+k}-n_{i++}, n_{ij+}+n_{+jk}-n_{+j+}, n_{i+k}+n_{+jk}-n_{++k}\}.\end{aligned} \tag{5}$$

3 The Bounding Methods of Fienberg and Buzzigoli-Giusti

3.1 The Procedure of Fienberg

In [6], Fienberg does not specify a bounding algorithm precisely, but illustrates an approach via example. The example is a 3x3x2 table of sample counts from the 1990 Decennial Census Public Use Sample [6, Table 1]. For convenience, we present it in the following form. The internal entries are:

Table 1. Fienberg [6, Table 1]

	INCOME					
	High	*Med*	*Low*	*High*	*Med*	*Low*
White	96	72	161	186	127	51
Black	10	7	6	11	7	3
Chinese	1	1	2	0	1	0
	MALE			*FEMALE*		

and the 2-dimensional marginal totals are:

$$\boldsymbol{A} = \begin{pmatrix} 107 & 197 \\ 80 & 135 \\ 169 & 54 \end{pmatrix}, \quad \boldsymbol{B} = \begin{pmatrix} 329 & 364 \\ 23 & 21 \\ 4 & 1 \end{pmatrix}, \quad \boldsymbol{C} = \begin{pmatrix} 282 & 199 & 212 \\ 21 & 14 & 9 \\ 1 & 2 & 2 \end{pmatrix}.$$

In the remainder of this sub-section, we examine the properties of the bound procedure of [6].

Corresponding to internal entry n_{ijk}, there exists a *collapsed* 2x2x2 table:

Table 2. Collapsing a 3-dimensional table around entry $\boldsymbol{n_{ijk}}$

n_{ijk}	$\sum_{J \neq j} n_{iJk}$		$\sum_{K \neq k} n_{ijK}$	$\sum_{J \neq j,\, K \neq k} n_{iJK}$
$\sum_{I \neq i} n_{Ijk}$	$\sum_{I \neq i,\, J \neq j} n_{IJk}$		$\sum_{I \neq i,\, K \neq k} n_{IjK}$	$\sum_{I \neq i,\, J \neq j,\, K \neq k} n_{IJK}$

Entry (2, 2, 2) in the lower right-hand corner is the *complement* of n_{ijk}, denoted $\bar{n}_{ijk}$. Fix (i, j, k). Denote the 2-dimensional marginals of Table 2 to which n_{ijk} contributes by $\bar{n}_{+22}$, $\bar{n}_{2+2}$, $\bar{n}_{22+}$. Observe that:

$$\begin{aligned} & n_{+++} - n_{i++} - n_{+j+} - n_{++k} + n_{ij+} + n_{i+k} + n_{+jk} - \bar{n}_{ijk} \\ & = (n_{+++} - n_{i++} - n_{+j+} + n_{ij+}) + (n_{i+k} + n_{+jk} - n_{++k}) - (\bar{n}_{ijk}) \\ & = (\bar{n}_{22+}) + (n_{ijk} - (\bar{n}_{22+} - \bar{n}_{ijk})) - (\bar{n}_{ijk}) = n_{ijk} . \end{aligned} \tag{6}$$

From (6) results the *2-dimensional Fréchet lower bound* of Fienberg (1999):

$$n_{ijk} \geq \max\{0, n_{+++} - n_{i++} - n_{+j+} - n_{++k} + n_{ij+} + n_{i+k} + n_{+jk} - \min\{\bar{n}_{+22}, \bar{n}_{2+2}, \bar{n}_{22+}\}\} . \tag{7}$$

The nominal (also called *Fréchet*) upper bound on n_{ijk} equals $\min\{n_{+jk}, n_{i+k}, n_{ij+}\}$. The *2-dimensional Fréchet bounds* of [6] are thus:

$$\begin{aligned} &\min\{n_{+jk}, n_{i+k}, n_{ij+}\} \geq n_{ijk} \geq \\ &\max\{0,\ n_{+++}-n_{i++}-n_{+j+}-n_{++k}+n_{ij+}+n_{i+k}+n_{+jk}-\min\{\bar{n}_{+22}, \bar{n}_{2+2}, \bar{n}_{22+}\}\}\ . \end{aligned} \tag{8}$$

Simple algebra reveals that the lower F-bounds of Section 2 and the 2-dimensional lower Fréchet bounds of [6] are identical. The F-bounds are easier to implement. Consequently, we replace (8) by (5).

From (6) also follows the *2-dimensional Bonferroni upper bound* of [6]:

$$n_{+++} - n_{i++} - n_{+j+} - n_{++k} + n_{ij+} + n_{i+k} + n_{+jk} \geq n_{ijk}\ . \tag{9}$$

This Bonferroni upper bound is not redundant: if $\max\{\bar{n}_{ijk}\}$ is sufficiently small, it can be sharper than the nominal upper bound. This is illustrated by the n_{111} entry of the 2x2x2 table with marginals:

$$A = \begin{pmatrix} 2 & 8 \\ 9 & 8 \end{pmatrix}\ ,\quad B = \begin{pmatrix} 2 & 9 \\ 9 & 7 \end{pmatrix}\ ,\quad C = \begin{pmatrix} 2 & 9 \\ 8 & 8 \end{pmatrix} \tag{10}$$

$n_{i++} + n_{+j+} + n_{++k} - 2n_{+++} \leq n_{ijk} - \bar{n}_{ijk}$ yields the *1-dimensional Fréchet lower bound* [6]:

$$n_{ijk} \geq n_{i++} + n_{+j+} + n_{++k} - 2n_{+++} \tag{11}$$

This bound is redundant with respect to the lower F-bounds. This is demonstrated as follows:

$$\begin{aligned} &1/3[(n_{ij+}+n_{i+k}-n_{i++}) + (n_{ij+}+n_{+jk}-n_{+j+}) + (n_{i+k}+n_{+jk}-n_{++k})] \\ &\qquad - (n_{i++}+n_{+j+}+n_{++k}-2n_{+++}) \\ &\qquad = 2/3(\bar{n}_{+22} + \bar{n}_{2+2} + \bar{n}_{22+}) \geq 0. \end{aligned} \tag{12}$$

The *Fréchet-Bonferroni bounds* of [6] can be replaced by:

$$\begin{aligned} &\min\{n_{+++}-n_{i++}-n_{+j+}-n_{++k}+n_{ij+}+n_{i+k}+n_{+jk},\ n_{+jk},\ n_{i+k},\ n_{ij+}\} \geq n_{ijk} \geq \\ &\qquad \max\{0, n_{ij+}+n_{i+k}-n_{i++}, n_{ij+}+n_{+jk}-n_{+j+}, n_{i+k}+n_{+jk}-n_{++k}\}\ . \end{aligned} \tag{13}$$

We return to the example [6, Table 1]. Here, the 2-dimensional Bonferroni upper bound (9) is not effective for any entry, and can be ignored. Thus, in this example, the bounds (13) are identical to the F-bounds (5), and should yield identical results. They in fact do not, most likely due to numeric error somewhere in [6]. This discrepancy is need to be kept in mind as we compare computational approaches below, that of Fienberg, using (13), and an alternative using (5).

Fienberg [6] computes the Fréchet bounds, without using the Bonferroni bound (9), resulting in Table 7 of [6]. We applied the F-bounds (5), but in place of his Table 7, obtained sharper bounds. Both sets of bounds are presented in Table 3, as follows. If our bound agrees with [6 , Table 7] we present the bound. If there is disagreement, we include the [6], Table 7 bound in parentheses alongside ours.

Table 3. F-Bounds and Fienberg (Table 7) Fréchet Bounds for Table 1

	INCOME				
MALE			*FEMALE*		
High	*Med*	*Low*	*High*	*Med*	*Low*
(80)85 −107	(53)64 −80	(142)158 −169	175 −197	(113)119 −135	(32)43 −54
0−21	0−14	0−9	0−21	0−14	0−9
0−1	0−2	0−2	0−1	0−1	0−1

Fienberg [6] next applies the Bonferroni upper bound (9) to Table 7, and reports improvement in five cells and a Table 8, the table of exact bounds for the table. We were unable to reproduce these results: the Bonferroni bound provides improvement for none of the entries.

3.2 The Shuttle Algorithm of Buzzigoli-Giusti

In [7], Buzzigoli-Giusti present the iterative *shuttle algorithm*, based on *principles of subadditivity*:

- a sum of lower bounds on entries is a lower bound for the sum of the entries, and
- a sum of upper bounds on entries is an upper bound for the sum of the entries;
- the difference between the value (or an upper bound on the value) of an aggregate and a lower bound on the sum of all but one entry in the aggregate is an upper bound for that entry, and
- the difference between the value (or a lower bound on the value) of an aggregate and an upper bound on the sum of all but one entry in the aggregate is a lower bound for that entry.

The shuttle algorithm begins with nominal lower and upper bounds. For each entry and its 2-dimensional marginal totals, the sum of current upper bounds of all other entries contained in the 2-dimensional marginal is subtracted from the marginal. This is a candidate lower bound for the entry. If the candidate improves the current lower bound, it replaces it. This is followed by an analogous procedure using sums of lower bounds and potentially improved upper bounds. This two-step procedure is repeated until all bounds are stationary. The authors fail to note but it is evident that *stationarity* is reached in a finite number of iterations because the marginals are integer.

3.3 Comparative Analysis of Fienberg , Shuttle, and F-Bounding Methods

We compare the procedure of Fienberg ([6]), the shuttle algorithm and the F-bounds. As previously observed, the bounds of [6] can be reduced to the F-bounds plus the 2-dimensional Bonferroni upper bound, viz., (13). The shuttle algorithm produces bounds at least as sharp as the F-bounds, for two reasons. First, the iterative shuttle procedure enables improved lower bounds to improve the nominal and subsequent upper bounds. Second, lower F-bounds can be no sharper than those produced during the first set of steps of the shuttle algorithm. To see this, for concreteness consider the candidate lower F-bound $n_{i+k} + n_{+jk} - n_{++k}$ for n_{ijk}. One of the three candidate shuttle lower bounds for n_{ijk} equals $n_{i+k} - \sum_{J \neq j} n^{U}_{iJk}$, where n^{U}_{iJk} denotes the current upper bound for n_{iJk}.

Thus, $n^{u}_{iJk} \leq n_{+Jk}$ and therefore $\sum_{J \neq j} n^{U}_{+Jk} \leq \sum_{J \neq j} n_{+Jk} = n_{++k} - n_{+jk}$.

Consequently, the shuttle candidate lower bound is greater than or equal to

$n_{i+k} - (n_{++k} - n_{+jk}) = n_{i+k} + n_{+jk} - n_{++k}$, so the shuttle candidate is at least as sharp as the F-candidate. If the shuttle algorithm is employed, all but the nominal lower Fienberg (1999) and F-bounds (namely, 0) are redundant.

Buzzigoli-Giusti [7] illustrate the 3-dimensional shuttle algorithm for the case of a 2x2x2 table. It is not clear, for the general case of a $d_1 x d_2 x d_3$ table, if they intend to utilize the collapsing procedure of Table 2, but in what follows we assume that they do. Consider the 2-dimensional Bonferroni upper bounds (9). From (6), the Bonferroni upper bound for n_{ijk} equals $n_{ijk} + \overline{n}_{ijk}$. Consider the right-hand 2-dimensional table in Table 2. Apply the standard 2-dimensional lower F-bound to the entry in the upper left-hand corner:

$$\sum_{K \neq k} n_{ijK} \geq \sum_{K \neq k} n_{ijK} - \overline{n}_{ijk} = ((n_{i++} - n_{i+k}) + (n_{+j+} - n_{+jk}) - (n_{+++} - n_{++k})) \;. \tag{14}$$

As previously observed, the shuttle algorithm will compute this bound during step 1and, if it is positive, replace the nominal lower bound with it, or with something sharper. During step 2, the shuttle algorithm will use this lower bound (or something sharper) to improve the upper bound on n_{ijk}, as follows:

$$\begin{aligned} n_{ijk} = n_{ij+} - \sum_{k \neq K} n_{ijK} &\leq n_{ij+} - ((n_{i++} - n_{i+k}) + (n_{+j+} - n_{+jk}) - (n_{+++} - n_{++k})) \\ &= n_{+++} - n_{i++} - n_{+j+} - n_{++k} + n_{ij+} + n_{i+k} + n_{+jk} \;. \end{aligned} \tag{15}$$

Thus, the Fienberg [6] 2-dimensional Bonferroni upper bound is redundant relative to the shuttle algorithm. Consequently, if the shuttle and collapsing methodologies are applied in combination, it suffices to begin with the nominal bounds and run the shuttle algorithm to convergence. Application of this approach (steps 1-2-1) to Table 1 yields Table 4 of exact bounds:

Table 4. Exact bounds for Table 1 from the Shuttle Algorithm

INCOME					
MALE *High*	MALE *Med*	MALE *Low*	FEMALE *High*	FEMALE *Med*	FEMALE *Low*
85–107	64–79	158–168	175–197	120–135	44–54
0–21	0–14	0–9	0–21	0–14	0–9
0–1	1–2	1–2	0–1	0–1	0–1

3.4 Limitations of All Three Procedures

Although the shuttle algorithm produced exact bounds for this example, the shuttle algorithm, and consequently the Fienberg ([6]) procedure and the F-bounds, are inexact, as follows. Examples 7b,c of [4] are 3-DIPTP exhibiting one or more non-integer continuous exact bounds. Because it is based on iterative improvement of integer upper bounds, in these situations the shuttle algorithm can come no closer than one unit larger than the exact integer bound, and therefore is incapable of achieving the exact integer bound.

The shuttle algorithm is not new: it was introduced by Schell [1] towards purported sufficient conditions on the 2-dimensional marginals for feasibility of 3-DPTP. By means of (16), Moravek-Vlach [2] show that the *Schell conditions* are necessary, but not sufficient, for the existence of a solution to the 3-DPTP. This counterexample is applicable here. Each 1-dimensional Fréchet lower bounds is negative, thus not effective. Each Bonferroni upper bound is too large to be sharp. No lower F-bound is positive. Iteration of the shuttle produces no improvement. Therefore, each procedure yields nominal lower (0) and upper (1) bounds for each entry. Each procedure converges. Consequently, all three procedures produce seemingly correct bounds when in fact no table exists. A simpler counterexample, (Example 2 of [5]), given by (17), appears in the next section.

$$A = \begin{pmatrix} 1 & 4 \\ 1 & 4 \\ 1 & 4 \\ 4 & 1 \\ 4 & 1 \\ 4 & 1 \\ 4 & 1 \\ 3 & 2 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 7 \\ 2 & 6 \\ 7 & 1 \\ 6 & 2 \\ 6 & 2 \end{pmatrix}, \quad C = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}. \tag{16}$$

4 The Roehrig and Chowdhury Network Models

Roehrig et al. (1999) [8] and Chowdhury et al. (1999) [9] offer network models for computing exact bounds on internal entries in 3-dimensional tables. Network models are extremely convenient and efficient, and most important enjoy the *integrality property*, viz., integer constraints (viz., 2-dimensional marginal totals) assure integer optima. Network models provide a natural mechanism and language in which to express 2-dimensional tables, but most generalizations beyond 2-dimensions are apt to fail.

Roehrig et al. [8] construct a network model for 2x2x2 tables and claim that it can be generalized to all 3-dimensional tables. This must be false. Cox ([4]) shows that the class of 3-dimensional tables (of size $d_1 x d_2 x d_3$) representable as a network is the set of tables for which $d_i < 3$ for at least one index *i*. This is true because, if all $d_i \geq 3$, it is possible to construct a 3-dimensional table with integer marginals whose corresponding polytope has non-integer vertices ([4]), which contradicts the integrality property.

Chowdhury et al. [9] address the following problem related to 3-DIPTP, also appearing in [8]. Suppose that the NSO releases only two of the three sets of 2-dimensional marginal totals (**A** and **B**), but not the third set (**C**) or the internal entries n_{ijk}. Is it possible to obtain exact lower and upper bounds for the remaining marginals (**C**)? The authors construct d_3 independent networks corresponding to the 2-dimensional tables defined by the appropriate (**A**, **B**) pairs and provide a procedure for obtaining exact bounds.

Unfortunately, this problem is quite simple and can be solved directly without recourse to networks or other mathematical formalities. In particular, the F-bounds of Section 2 suffice, as follows. Observe that, absent the **C**-constraints, the minimum (maximum) feasible value of a **C**-marginal total $\mathbf{C}_{ij}$ equals the sum of the minimum (maximum) values for the corresponding internal entries n_{ijk}. As the n_{ijk} are subject only to 2-dimensional constraints within their respective k-planes, then exact bounds for each n_{ijk} are precisely its 2-dimensional lower and upper F-bounds. These can be computed at sight and added along the k-dimension thus producing the corresponding **C**-bounds without recourse to networks or other formulations.

The Chowdhury et al. [9] method is insensitive to infeasibility, again demonstrated by example (16): all Fréchet lower and nominal upper bounds computed in the two 2-dimensional tables defined by k = 1 and k = 2 contain the corresponding $\mathbf{C}_{ij}$ (equal to 1 in all cases), but there is no underlying table as the problem is infeasible. Insensitivity is also demonstrated by Example 2 of [5], also presented at the 1998 conference, viz.,

$$\boldsymbol{A} = \boldsymbol{B} = \boldsymbol{C} = \begin{pmatrix} 1 & 1 & 3 \\ 3 & 1 & 1 \\ 1 & 3 & 1 \end{pmatrix} . \tag{17}$$

5 Fractional Extremal Points

Theorem 4.3 of [4] demonstrates that the only multi-dimensional integer planar transportation problems (*m-DIPTP*) for which integer extremal points are assured are those of size 2^{m-2}xbxc. In these situations, linear programming methods can be relied upon to produce exact integer bounds on entries, and to do so in a computationally efficient manner even for problems of large dimension or size. The situation for other *integer problems of transportation type*, viz., m-dimensional contingency tables subject to k-dimensional marginal totals, k = 0, 1, ..., m-1, is quite different: until a direct connection can be demonstrated between exact continuous bounds on entries obtainable from linear programming and exact integer bounds on entries, linear programming will remain an unreliable tool for solving multi-dimensional problems of transportation type. Integer programming is not a viable option for large dimensions or size or repetitive use. Methods that exploit unique structure of certain subclasses of tables are then appealing, though possibly of limited applicability.

Dobra-Fienberg [10] present one such method, based on notions from mathematical statistics and graph theory. Given an m-dimensional integer problem of transportation type and specified marginal totals, if these marginals form a *set of sufficient statistics* for a specialized log-linear model known as a *decomposable graphical model*, then the model is feasible and exact integer bounds can be obtained from straightforward formulae. These formulae yield, in essence, the F- and Bonferroni bounds. The reader is referred to [10] for details, [11] for details on log-linear models, and [12] for development of graphical models.

The m-dimensional planar transportation problem considered here, m > 2, corresponds to the *no m-factor effect* log-linear model, which is not graphical, and consequently the Dobra-Fienberg method [10] does not apply to problems considered here.

The choice here and perhaps elsewhere of the 3- or m-DIPTP as the initial focus of study for bounding problems is motivated by the following observations. If for reasons of confidentiality the NSO cannot release the full m-dimensional tabulations (viz., the m-dimensional marginal totals), then its next-best strategy is to release the (m-1)-dimensional marginal totals, corresponding to the m-DIPTP. If it is not possible to release all of these marginals, perhaps the release strategy of Chowdhury et al. [9] should be investigated. Alternatively, release of the (m-2)-dimensional marginals might be considered. This strategy for release is based on the principle of releasing the most specific information possible without violating confidentiality. Dobra-Fienberg offers a different approach, a class of marginal totals, perhaps of varying dimension, that can be released while assuring confidentiality via easy computation of exact bounds on suppressed internal entries.

A formula driven bounding method is valuable for large problems and for repetitive, large scale use. Consider the m-dimensional integer problem of transportation type specified by its 1-dimensional marginal totals. In the statistics literature, this is known as the *complete independence* log-linear model [11]. This model, and in particular the 3-dimensional complete independence model, is graphical and decomposable. Thus, exact bounding can be achieved using Dobra-Fienberg.

Such problems can exhibit non-integer extremal points. For example, consider the 3x3x3 complete independence model with 1-dimensional marginal totals given by the vector:

$$\Big(\sum_{i,\,j} n_{ijk}\Big) \;=\; \Big(\sum_{i,\,k} n_{ijk}\Big) \;=\; \Big(\sum_{j,\,k} n_{ijk}\Big) \;=\; (2 \quad 1 \quad 2)\ . \tag{18}$$

Even though all continuous exact bounds on internal entries in (18) are integer, one extremal point at which n_{312} is maximized (at 1) contains four non-integer entries and another contains six. Bounding using linear programming would only demonstrate that $n_{312} = 1$ is the continuous, not the integer, maximum if either of these extremal points were exhibited. A strict network formulation is not possible because networks exhibit only integer extremal points (although use of networks with side constraints is under investigation). Integer programming is undesirable for large or repetitive applications. Direct methods such as Dobra-Fienberg may be required. A drawback of Dobra-Fienberg is that it applies only in specialized cases.

6 Discussion

In this paper we have examined the problem of determining exact integer bounds for entries in 3-dimensional integer planar transportation problems. This research was motivated by previous papers presenting heuristic approaches to similar problems that failed in some way to meet simple criteria for performance or reasonableness. We examined these and other approaches analytically and demonstrated one's superiority. We demonstrated that this method is imperfect and a reformulation of a method from the operations literature of the 1950s. We demonstrated that these methods are insensitive to infeasibility and can produce meaningless results otherwise undetected. We demonstrated that a method purported to generalize from 2x2x2 tables to all 3-dimensional tables could not possibly do so. We demonstrated that a problem posed and solved using networks in a *Management Science* paper can be solved by simple, direct means without recourse to mathematical programming. We illustrated the relationship between computing integer exact bounds, the presence of non-integer extremal points and the applicability of mathematical programming formulations such as networks.

NSOs must rely on automated statistical methods for operations including estimation, tabulation, quality assurance, imputation, rounding and disclosure limitation. Algorithms for these methods must converge to meaningful quantities. In particular, these procedures should not report meaningless, misleading results such as seemingly correct bounds on entries when no feasible values exist. These risks are multiplied in statistical data base settings where data from different sources often are combined. Methods examined here for bounds on suppressed internal entries in 3-dimensional contingency tables fail this requirement because they are heuristic and based on necessary, but not sufficient, conditions for the existence of a solution to the 3-DPTP. In addition, most of these methods fail to ensure exact bounds, and are incapable of identifying if and when they do in fact produce an exact bound. Nothing is gained by extending these methods to higher dimensions.

Disclaimer

Opinions expressed are solely those of the author and are not intended to represent policy or practices of the National Center for Health Statistics, Centers for Disease Control and Prevention, or any other organization.

References

1. Schell, E. Distribution of a product over several properties. Proceedings, 2nd Symposium on Linear Programming. Washington, DC (1955) 615-642
2. Moravek, J. and Vlach, M. (1967). On necessary conditions for the existence of the solution to the multi-index transportation problem. Operations Research 15, 542-545
3. Vlach, M. Conditions for the existence of solutions of the three-dimensional planar transportation problem. Discrete Applied Mathematics 13 (1986) 61-78
4. Cox, L. (2000). On properties of multi-dimensional statistical tables. Manuscript (April 2000) 29 pp.

5. Cox, L. Invited Talk: Some remarks on research directions in statistical data protection. Statistical Data Protection: Proceedings of the Conference.EUROSTAT, Luxemburg (1999)163-176
6. Fienberg, S. Fréchet and Bonferroni bounds for multi-way tables of counts with applications to disclosure limitation. Statistical Data Protection: Proceedings of the Conference. EUROSTAT, Luxembourg (1999) 115-129
7. Buzzigoli, L. and Giusti, A. An algorithm to calculate the lower and upper bounds of the elements of an array given its marginals. Statistical Data Protection: Proceedings of the Conference. EUROSTAT, Luxemburg (1999) 131-147
8. Roehrig, S., Padman, R., Duncan, G., and Krishman, R. Disclosure detection in multiple linked categorical datafiles: A unified network approach. Statistical Data Protection: Proceedings of the Conference. EUROSTAT, Luxembourg (1999) 149-162
9. Chowdhury, S., Duncan, G., Krishnan, R., Roehrig, S., and Mukherjee, S. Disclosure detection in multivariate categorical databases: Auditing confidentiality protection through two new matrix operators. Management Science 45 (1999) 1710-1723
10. Dobra, A. and S. Fienber, S.. Bounds for cell entries in contingency table given marginal totals and decomposable graphs. Proceedings of theNational Academy of Sciences 97 (2000) 11185-11192
11. Bishop, Y., Fienberg, S., and Holland, P. Discrete Multivariate Analysis: Theory and Practice. Cambridge, MA: M.I.T. Press (1975)
12. Lauritzen, S. Graphical Models. Oxford: Clarendon Press (1996)