

# Multi-Dimensional Statistical Tables

Lawrence COX  
*National Center for Health Statistics*  
6525 Belcrest Road, Room 915  
Hyattsville, MD 27082 USA  
[LCOX@CDC.GOV](mailto:LCOX@CDC.GOV)

**Abstract:** Two-dimensional tables with non-negative integer entries, e.g., contingency tables, are a staple of statistical science. Two-dimensional tables enjoy mathematical properties on which important statistical methods depend, e.g., stratified sampling, imputation, statistical data protection, and sampling from and fitting log-linear models to contingency tables. We demonstrate that many desirable mathematical properties and associated statistical methods are not extendible to three or higher dimensions, and further that ill-behaved examples are ubiquitous, abundant and not anomalies. For the most constrained case, viz., an  $n$ -dimensional table with prescribed  $(n-1)$ -dimensional marginal totals, we demonstrate necessary and sufficient conditions and an empirical test for the existence of a feasible table and characterize completely  $n$ -dimensional tables for which the existence of integer-valued entries and integer optima are assured. The importance of these results to statistical disclosure limitation is demonstrated.

**Keywords:** contingency table, log-linear model, iterative proportional fitting, mathematical network

## 1. Introduction

Two-dimensional tables with non-negative integer entries, e.g., contingency tables, are a staple of statistical science. Important statistical methods depend upon mathematical properties of two-dimensional tables, and it is desirable to extend these methods to three and higher-dimensional tables. The purpose of this paper is to demonstrate that doing so may be impossible or unsound in many situations and to characterize a class of tables for which it is possible. Further results and greater detail can be found in [1] and [2].

## 2. Two-Dimensional Statistical Tables

A two-dimensional statistical table  $\mathbf{T}(\mathbf{bc})$  with  $b$  rows and  $c$  columns is represented:

$${}^c_j \mathbf{a}_{ij} \quad \mathbf{a}_{i.} \quad {}^b_{i1} \mathbf{a}_{ij} \quad \mathbf{a}_{.j} \quad {}^b_{i1} \mathbf{a}_{.j} \quad {}^c_{j1} \mathbf{a}_{.j} \quad \mathbf{a}_{..}$$

We restrict attention to positive tables, viz.,  $a_{ij} \geq 0$ .  $\mathbf{A} = ((\mathbf{a}_i), (\mathbf{a}_j))$  is the vector of *one-dimensional marginal totals* for  $\mathbf{T}$ . The grand total equation assures that the one dimensional marginal totals are *consistent*. Given  $\mathbf{A}$ , a *feasible* two-dimensional table is an assignment of nonnegative *internal entries* satisfying the consistency conditions. Note for later reference that the one-dimensional marginal totals in  $n = 2$  dimensions are the “(n-1)-dimensional marginal totals”. These provide our principal focus in n-dimensions. Two dimensional statistical tables enjoy a variety of mathematical properties fundamental to important statistical methods. In this section, we review eight of these properties and associated statistical methods for  $n = 2$ .

*Property 1:* A consistent pair (set) of one-dimensional (viz., (n-1)-dimensional) marginal totals assure the existence of a feasible two-dimensional table.

A feasible table can be constructed either by the independence solution, viz.  $a_{ij} = a_i \cdot a_j / a_{..}$ , or by the stepping stones algorithm (discussed below). Property 1 assures that a two-dimensional joint distribution can always be fit to consistent one-dimensional (viz., (n-1)-dimensional) marginal distributions.

*Property 2:* Fréchet bounds in two-dimensional tables are exact.

The *Fréchet upper bound* for an internal entry  $a_{ij}$  is the minimum of its two onedimensional marginal totals. That the Fréchet upper bound is exact can be seen from the *stepping stones* algorithm: select an internal entry; assume for concreteness that its Fréchet upper bound equals the row total; assign the entry its Fréchet upper bound; set all other entries in the row to zero; subtract the Fréchet upper bound from the column and grand totals; ignoring the row, select another entry and repeat the process; and, stop when the grand total is reduced to zero.

The *Fréchet lower bound* for internal entry  $a_{ij}$  equals  $\max\{0, a_i - a_j, a_j - a_i\}$ . That this is a lower bound can be seen from:  $a_i - a_j = a_{i.} - a_{.j} = a_{ij} - a_{.j} + a_{.i} - a_{ij} = a_{.i} - a_{.j}$ . As the sum is nonnegative, the left-hand side is a lower bound.

Property 2 is important in statistical disclosure limitation. As an illustration, assume that for confidentiality reasons that it is not possible to publish some or all internal cell values. An alternative would be to release only the marginal totals. This requires performing a *disclosure audit*, i.e., verification that exact estimates of internal cells do not constitute unacceptable disclosure risk. Fréchet bounds render this computation trivial. Otherwise, it would often be necessary to solve multiple integer programs. Exactness follows by demonstrating that the sum can be zero. The next two properties follow, respectively, from the stepping stones algorithm and Property 2.

*Property 3:* A consistent pair of integer one-dimensional marginal totals assure the existence of a feasible integer two-dimensional table.

*Property 4:* In a feasible two-dimensional table, consistent integer one-dimensional (viz., (n-1)-dimensional) marginal totals guarantee exact integer lower and upper bounds on internal entries.

*Property 5:* In a two-dimensional table, even if integer one-dimensional marginal totals are small, there often exist many integer feasible tables.

For example, given a  $b \times b$  table  $\mathbf{T}(\mathbf{b}_2)$  with all one-dimensional marginal totals equal to one, there are  $b!$  feasible integer tables. For disclosure limitation purposes, a small number of alternative integer feasible solutions could constitute unacceptable disclosure risk.

*Property 6:* Controlled rounding is assured in two-dimensional tables.

Given an integer rounding base  $B$ , a *controlled rounding* of a table to base  $B$  is a second additive table each of whose entries equals either of the two integer multiples of  $B$  that are numerically adjacent to the original entry. *Zero-restricted* controlled rounding leaves multiples of  $B$  fixed. [3] demonstrates that (zero-restricted) controlled rounding of a two dimensional table is always possible. [4] provides a procedure for unbiased controlled rounding that solves the two-way stratification (a.k.a. *controlled selection*) problem of sampling theory. Controlled rounding is important to iterative proportional fitting [6], for which convergence to integer entries is not assured, thus requiring controlled rounding base  $B = 1$ , and to statistical disclosure limitation and other statistical applications.

### 3. These Properties Fail to Extend to Three- and Higher-Dimensions

In this section, failure of these properties of two-dimensional tables to generalize to three and therefore higher-dimensional tables is illustrated by counterexamples. The figures comprising the examples in this section are read as follows. Each example represents the additive structure of the marginals entries for a *potential* three-dimensional table of nonnegative entries indexed  $(i, j, k)$ ,  $i = 1, \dots, d_1$ ,  $j = 1, \dots, d_2$ ,  $k = 1, \dots, d_3$ . The three dimensional internal entries are to be arranged in the blank boxes. The two-dimensional marginal totals in the  $i$ - and  $j$ -directions appear, respectively, below and to the right of the box. The two-dimensional marginal totals in the  $k$ -direction appear in the box below the dark line. Keep in mind that for  $n = 3$ , the two-dimensional marginal totals are the  $(n-1)$ -dimensional marginal totals, the focus of our interest in  $n$ -dimensions. The three dimensional potential table is formed by “stacking” the two-dimensional tables for successive values of  $k$  on top of the two-dimensional table of  $k$ -directional two-dimensional marginal totals. Remaining entries are the one-dimensional (viz.,  $(n-2)$ -dimensional) marginal totals, located at the lower right beside each box above the line and to the left and right of the box below the line, and the  $(n-3)$ -dimensional marginal total (which in three dimensions is the *grand total*), located at the lower right beside the box below the line. We refer to these structures as “potential three-dimensional tables” because the existence of non-negative internal entries satisfying the additive constraints imposed by the two dimensional marginal totals is not assured, as we now demonstrate.

*Example 1:* Consistency of (n-1)-dimensional marginal totals does not guarantee the existence of a feasible n-dimensional table.

In n-dimensions, the (n-1)-dimensional marginal totals are defined by holding n-1 indices fixed and summing over the remaining index. They organize naturally into n sets, each corresponding to one summation index. If these sets of (n-1)-dimensional marginal totals admit a feasible n-dimensional table, then any pair of them must admit a feasible two dimensional table, and this pair must therefore obey the consistency condition. There are  $n(n-1)/2$  such pairs. If each pair of sets of (n-1)-dimensional marginal totals is mutually consistent, then a *consistent set of (n-1)-dimensional marginal totals* is said to exist. Consistent sets of lower-dimensional marginal totals can be defined similarly, but are not of concern here as consistency of (n-1)-dimensional marginals assures consistency of lowerdimensional marginals. Presence of a consistent set of (n-1)-dimensional marginal totals is thus a necessary condition for the existence of a feasible n-dimensional table. It is not, however, sufficient, as illustrated by Example 1a. This potential three-dimensional table is infeasible, as can be demonstrated with moderate effort (exercise for the reader). Therefore, consistent (n-1)-dimensional marginal totals do not guarantee the existence of a feasible table in three and higher dimensions and consequently for  $n > 3$  marginal distributions do not guarantee existence of an underlying n-dimensional joint distribution.

|       |   |
|-------|---|
|       | 3 |
|       | 1 |
|       | 1 |
| 3 1 1 | 5 |

|       |   |
|-------|---|
|       | 1 |
|       | 3 |
|       | 1 |
| 3 1 1 | 5 |

|       |   |
|-------|---|
|       | 1 |
|       | 1 |
|       | 3 |
| 1 1 3 | 5 |

|       |    |
|-------|----|
| 1 1 3 | 5  |
| 1 3 1 | 5  |
| 3 1 1 | 5  |
| 5 5 5 | 15 |

**Example 1a:** Infeasible 3-D Table with Apparent Fréchet Bounds

Several statistical procedures proven in two dimensions are *insensitive* to infeasibility in  $n > 3$  dimensions, viz., they will produce a final result regardless of whether or not a feasible table exists. Often, these methods seek to generalize a proven twodimensional procedure to higher dimensions. For example, generalizations of Fréchet bounds are insensitive to infeasibility. This is illustrated in Example 1a: all Fréchet bounds are computable, and in addition are *apparent bounds*, viz., do not suffer the inconsistency of a lower bound exceeding the corresponding upper bound. However, no underlying three-dimensional table

exists. Such putative bounds are certainly misleading. By virtue of Example 1a, any method based on a finite number of additions, subtractions and multiplications on consistent sets of (n-1)-dimensional marginal totals must be insensitive to infeasibility. [7], [8] and [9] offer methods for bounding internal entries in threedimensional tables for disclosure limitation purposes. All are insensitive to Example 1a. The first two methods employ these arithmetic operations in conjunction with iteration of subadditive relations. They are provably inoptimal and insensitive ([2]). [9] is provably optimal but insensitive, as follows.

The authors [9] consider the following problem. Marginal totals are available for two *views* of a three-dimensional table (viz., two of three planes of (n-1)-dimensional marginal totals), but not the third. This can arise in the context of a *statistical data base query system* if, e.g., as the authors suggest, the underlying table consists of counts of patient-by-doctor-by-treatment occurrences. The three-dimensional data and the two-dimensional patient bytreatment view are assumed confidential, but the doctor-by-treatment and the patient-bydoctor views are not and are released. The problem for the data intruder is to obtain optimal lower and upper bounds on the missing patient-by-treatment view. The authors offer a network optimization algorithm for this purpose.

Consider the  $i = 4$  and  $j = 4$  views from Example 1a, and assume that the objective is to estimate the third view (viz., the vertical plane of marginal totals,  $k = 4$ ), presumed missing, using [9]. The bounds illustrated at the bottom of Example 1b result. These bounds appear reasonable, viz., are apparent, despite the fact that no table can exist.

|   |   |              |   |   |   |   |   |   |   |  |   |   |   |   |   |   |   |   |   |   |
|---|---|--------------|---|---|---|---|---|---|---|--|---|---|---|---|---|---|---|---|---|---|
| <table style="border-collapse: collapse; width: 100px; height: 100px;"> <tr><td style="padding: 2px 10px;">3</td><td style="padding: 2px 10px;">1</td><td style="padding: 2px 10px;">1</td></tr> <tr><td style="padding: 2px 10px;">1</td><td style="padding: 2px 10px;">3</td><td style="padding: 2px 10px;">1</td></tr> <tr><td style="padding: 2px 10px;">1</td><td style="padding: 2px 10px;">1</td><td style="padding: 2px 10px;">3</td></tr> </table> | 3 | 1            | 1 | 1 | 3 | 1 | 1 | 1 | 3 |  | <table style="border-collapse: collapse; width: 100px; height: 100px;"> <tr><td style="padding: 2px 10px;">3</td><td style="padding: 2px 10px;">1</td><td style="padding: 2px 10px;">1</td></tr> <tr><td style="padding: 2px 10px;">1</td><td style="padding: 2px 10px;">3</td><td style="padding: 2px 10px;">1</td></tr> <tr><td style="padding: 2px 10px;">1</td><td style="padding: 2px 10px;">1</td><td style="padding: 2px 10px;">3</td></tr> </table> | 3 | 1 | 1 | 1 | 3 | 1 | 1 | 1 | 3 |
| 3   | 1 | 1            |   |   |   |   |   |   |   |  |   |   |   |   |   |   |   |   |   |   |
| 1   | 3 | 1            |   |   |   |   |   |   |   |  |   |   |   |   |   |   |   |   |   |   |
| 1   | 1 | 3            |   |   |   |   |   |   |   |  |   |   |   |   |   |   |   |   |   |   |
| 3   | 1 | 1            |   |   |   |   |   |   |   |  |   |   |   |   |   |   |   |   |   |   |
| 1   | 3 | 1            |   |   |   |   |   |   |   |  |   |   |   |   |   |   |   |   |   |   |
| 1   | 1 | 3            |   |   |   |   |   |   |   |  |   |   |   |   |   |   |   |   |   |   |
| $i = 4$ View  |   | $j = 4$ View |   |   |   |   |   |   |   |  |   |   |   |   |   |   |   |   |   |   |

|   |   |   |   |   |   |
|---|---|---|---|---|---|
| 1 | 5 | 0 | 3 | 0 | 3 |
| 0 | 3 | 1 | 5 | 0 | 3 |
| 0 | 3 | 0 | 3 | 1 | 5 |

**Example 1b:** Bounds from [9] on  $k = 4$  (Vertical) View of Example 1

An argument can be raised surrounding these examples that infeasible tables do not arise in statistics and therefore are of no practical interest. Clearly, given feasible internal entries, a table with consistent marginals results and all of the statistical methods described in this paper, and others, can be performed. However, complete feasible tables are not always available. Marginal totals are not always derived directly by addition from internal entries, free of error, unaltered, or known with precision. For example, each set of (n-1) dimensional marginal totals for a potential n-dimensional table might be based on estimates from different sample surveys. After adjusting the estimated marginals to achieve consistency, it is reasonable to attempt to fit internal entries to them, with or without a starting sample. Owing to Property 1, all of this can be done in two-dimensions without explicit concern for feasibility. However, it can fail in three or higher dimensions, and the failure can go undetected, so that an investigator can be analyzing and drawing conclusions

from a table that in fact does not exist.

There are other possible situations where marginal totals are consistent but infeasible. A great deal of statistical data are estimates derived from samples. Estimates and counts are subject to error. Counts may have been perturbed for disclosure limitation purposes, or subjected to rounding or imputation. While there are ways to control such operations in two-dimensions (Property 8), such methods can fail in higher dimensions. The advent of statistical data base query systems brings the need to combine data in various ways to produce estimates. Dynamic data bases or data bases subject to certain disclosure limitation procedures allow different answers to the same query. Analysts have and will continue to use and combine flawed and incomplete data.

|     |     |   |     |
|-----|-----|---|-----|
| 0.5 | 1   | 1 | 2.5 |
| 1   | 0.5 | 0 | 1.5 |
| 1   | 0   | 0 | 1   |
| 2.5 | 1.5 | 1 | 5   |

|   |     |     |     |
|---|-----|-----|-----|
| 0 | 0.5 | 0.5 | 1   |
| 1 | 0.5 | 1   | 2.5 |
| 0 | 1.5 | 0   | 1.5 |
| 1 | 2.5 | 1.5 | 5   |

|     |     |     |     |
|-----|-----|-----|-----|
| 0.5 | 0   | 1   | 1.5 |
| 0.5 | 0   | 0.5 | 1   |
| 0   | 1.5 | 0   | 2.5 |
| 1.5 | 1   | 2.5 | 5   |

|     |     |     |    |
|-----|-----|-----|----|
| 1   | 1.5 | 2.5 | 5  |
| 2.5 | 1   | 1.5 | 5  |
| 1.5 | 2.5 | 1   | 5  |
| 5   | 5   | 5   | 15 |

**Example 1c:** Feasible 3-D Table

All of this can lead to infeasibility. If infeasibility were rare, the problem might be ignored. However, as demonstrated in Section 4, infeasibility is far from rare in  $n > 3$  dimensions. The potential to create infeasibility from original data is illustrated as follows. Consider Example 1c. This is a feasible table. Henceforth, ignore the internal values, viz., pretend they are unknown. Round the 2-dimensional marginal totals to integers. This can be done in several ways. One way would produce Example 2, which is feasible. Another, equally plausible, rounding would produce Example 1a, which is infeasible.

*Example 2:* Fréchet bounds are not exact in  $n > 3$  dimensions.

In  $n$ -dimensions, the *Fréchet upper bound* of an internal entry is the minimum of the  $(n-1)$ -dimensional marginal totals to which the entry contributes. In Example 2, the (3, 3, 1) entry (lower right, first box) has Fréchet upper bound equal to one. However, this entry achieves a unique value of zero, as there is precisely one feasible table.

|       |   |
|-------|---|
|       | 3 |
|       | 1 |
|       | 1 |
| 3 1 1 | 5 |

|       |   |
|-------|---|
|       | 1 |
|       | 3 |
|       | 1 |
| 1 3 1 | 5 |

|       |   |
|-------|---|
|       | 1 |
|       | 1 |
|       | 3 |
| 1 1 3 | 5 |

|   |   |   |    |
|---|---|---|----|
| 1 | 2 | 2 | 5  |
| 2 | 1 | 2 | 5  |
| 2 | 2 | 1 | 5  |
| 5 | 5 | 5 | 15 |

**Example 2:** Feasible 3-D Table with Inexact Fréchet Upper Bound

|     |   |
|-----|---|
|     | 2 |
|     | 1 |
|     | 0 |
| 2 1 | 3 |

|   |   |   |
|---|---|---|
| 3 | 1 | 0 |
| 1 | 1 | 1 |
| 0 | 1 | 1 |
| 1 | 1 | 2 |

|     |   |
|-----|---|
|     | 2 |
|     | 0 |
|     | 0 |
| 1 1 | 2 |

|   |   |   |
|---|---|---|
| 3 | 1 | 4 |
| 1 | 1 | 2 |
| 0 | 1 | 1 |
| 4 | 3 | 7 |

**Example 3:** Feasible (Unique) 3-D Table with Inexact Fréchet Lower Bound

|         |   |
|---------|---|
|         | 1 |
|         | 1 |
|         | 1 |
|         | 1 |
| 1 1 1 1 | 4 |

|         |   |
|---------|---|
|         | 1 |
|         | 1 |
|         | 1 |
|         | 0 |
| 1 1 1 0 | 3 |

|         |   |
|---------|---|
|         | 0 |
|         | 1 |
|         | 1 |
|         | 0 |
| 0 0 1 1 | 2 |

|         |   |
|---------|---|
|         | 0 |
|         | 0 |
|         | 1 |
|         | 1 |
| 1 0 0 1 | 2 |

|   |   |   |   |   |
|---|---|---|---|---|
| 1 | 0 | 1 | 0 | 2 |
| 0 | 1 | 1 | 1 | 3 |
| 1 | 1 | 1 | 1 | 4 |
| 1 | 0 | 0 | 1 | 2 |
| 3 | 2 | 3 | 3 | 1 |
|   |   |   |   | 1 |

**Example 4: Fractional Optimum In the Absence of Zero-Restrictions**

*Example 3:* Fréchet lower bounds are not exact in  $n > 3$  dimensions.

In  $n$ -dimensions, the *Fréchet lower bound* equals the maximum of zero and the  $n(n-1)/2$  possible two-dimensional Fréchet lower bounds. In Example 3, all three two-dimensional Fréchet lower bounds for the (1, 1, 1) entry equal one, so its Fréchet lower bound equals one. However, this entry has a unique value of two.

*Example 4:* Exact continuous lower and upper bounds can fail to be integer.

Example 4 illustrates a three-dimensional table with integer two-dimensional marginal totals for which at least one entry has a noninteger optimum: linear programming reveals that the maximum continuous value of the (3, 3, 1) entry in Example 4 is  $\frac{1}{2}$ . This is the only fractional optimum. This example illustrates that, in  $n > 3$  dimensions, linear programming does not always provide exact bounds on missing or adjusted integer internal entries. Nevertheless, its use for such purposes is standard practice, e.g., in official statistics, for auditing data suppressed for disclosure limitation purposes.

|   |   |   |   |
|---|---|---|---|
|   |   |   | 3 |
|   |   |   | 1 |
|   |   |   | 1 |
| 3 | 1 | 1 | 5 |

|   |   |   |   |
|---|---|---|---|
|   |   |   | 1 |
|   |   |   | 3 |
|   |   |   | 1 |
| 1 | 3 | 1 | 5 |

|   |   |   |   |
|---|---|---|---|
|   |   |   | 1 |
|   |   |   | 1 |
|   |   |   | 3 |
| 1 | 1 | 3 | 5 |

|   |   |   |    |
|---|---|---|----|
| 2 | 1 | 2 | 5  |
| 1 | 2 | 2 | 5  |
| 2 | 2 | 1 | 5  |
| 5 | 5 | 5 | 15 |

**Example 5: Feasible 3-D Table with 3 df But only 4 Integer Solutions**

*Example 5:* Few integer feasible solutions may exist.

*Example 6:* Controlled rounding is not always possible in  $n > 3$  dimensions.

[5] provides a counterexample to three-dimensional controlled rounding.

**4. The Structure of Three-Dimensional Tables**

Examples 1 and 4 enable creation of consistent but infeasible three-dimensional tables **T(abc)** of arbitrary size  $a, b, c > 3$  by combining these examples with blocks of zeroes. For example, to create an infeasible  $5 \times 5 \times 5$  table, place  $3 \times 3 \times 3$  Example 1a in the upper-front-left of the  $5 \times 5 \times 5$  space, and place a similar  $2 \times 2 \times 2$  example (not shown) in the lower-back-right. Fill in the remainder of the block with zeroes. Similarly, infeasible three-dimensional tables can be combined as blocks along a diagonal

to create a four-dimensional infeasible table, and so on, thus demonstrating the existence of infeasible tables with consistent (n-1)-dimensional marginal totals of arbitrary dimension  $n > 3$  and nearly arbitrary size (see below for exceptions). Similar constructions are possible for the other examples.  $T(d_1, \dots, d_n; A)$  denotes an n-dimensional statistical table of size  $(d_1, \dots, d_n)$ . Thus,  $T$  comprises  $\prod_{j=1}^n d_j$  nonnegative internal entries,  $d_j > 1$ , constrained by n sets of (n-1)-dimensional aggregation equations  $MXA$ , for  $M$  the  $\{0, 1\}$  aggregation matrix of a generic n-dimensional table of this size and  $A$  a vector of consistent (n-1)-dimensional totals. We are interested in properties of all tables of a particular dimension and size, viz., in the properties of  $M$ . We refer to  $T(d_1, \dots, d_n)$  as a *generic* n-dimensional table, viz.  $T(d_1, \dots, d_n; A)$ , for arbitrary  $A$ . Shorthand such as  $T(2_n)$  or  $T(bc)$  is used when the meaning is clear, and  $mT(d_1, \dots, d_n; A)$  is  $T(d_1, \dots, d_n; mA)$ . Eliminate all linearly dependent rows from  $M$ . Let  $M_B$  be a nonsingular submatrix of  $M$  of maximal rank. Reorder the columns (variables) of  $M$  and  $A$  so that and. Then  $M' (M_B, M_N) A' (A_B, A_N)$  the *basic solution*  $MX' A$  of corresponding to is  $M_B x' (M_B A_B, 0)$

We have shown that counterexamples to feasibility  $T(d_1, \dots, d_n; A)$  exist in dimension  $n > 3$ . Counterexamples are more the rule than the exception: given a feasible table of such dimension and size, there exists a corresponding, countably infinite set of infeasible tables.

**Theorem 1:** Let  $T(d_1, \dots, d_n)$  denote a generic table for  $n > 3$ . Let  $T(d_1, \dots, d_n; A_i)$  denote a feasible table and denote an  $T(d_1, \dots, d_n; A_i)$  infeasible table. There exists an integer  $p$  such that  $T(d_1, \dots, d_n; mA_i)$  is infeasible for all  $m > p$ .  
*Proof:* Omitted. See [1]. *Q.E.D.*

Theorem 1 demonstrates that infeasibility in higher dimensions is pervasive, and not a mathematical anomaly. Moreover, infeasibility is typically not easy to recognize, e.g., Example 1a.

We previously remarked that heuristic arithmetic algorithms based on sets of consistent (n-1)-dimensional marginal totals to bound internal entries are insensitive to feasibility. Typically such procedures proceed towards upper bounds from outside the feasible region, viz., from larger to smaller values (from smaller to larger for lower bounds). Such procedures must terminate at or before reaching the integer part of the optimal upper bound. If the continuous upper bound is noninteger (e.g., Examples 4), then necessarily these algorithms must fail to identify the optimal integer upper bound. Similarly, heuristic algorithms for controlled data perturbation controlled rounding, and other applications in n-dimensional tables are likely to encounter infeasibility at unpredictable times and in unpredictable ways.

Feasibility can be tested using linear programming, but without insight into what conditions on the (n-1)-dimensional marginals ensure feasibility. Research on the feasibility of the three-dimensional transportation problem produced only necessary conditions. A statistical approach might yield insight into compatibility conditions between an ndimensional joint distribution and its marginals. Fortunately, there is a statistical method that can be used to detect infeasibility.

*Theorem 2 (Feasibility Test):* A table is feasible if and only if, starting with all ones, iterative proportional fitting with respect to the  $n$  sets of  $(n-1)$ -dimensional marginal totals converges for each internal entry.

*Proof:* Omitted. See [1]. *Q.E.D.*

This result is theoretical, but owing to rapid convergence of the iterative proportional fitting algorithm, it is a practical tool for detecting feasibility and producing a feasible solution. It can be difficult to prove analytically that a particular sequence does not converge. However, if divergence manifests itself as  $n$  subsequences converging to two or more distinct limits, then it should be possible to detect infeasibility with equal confidence and ease.

The result extends to the case of structural zeroes, as follows. For each structural zero of  $\mathbf{T}$ , define  $a_{i_1, \dots, i_q, \dots, i_n}^{(0)} = 0$ , and define  $a_{i_1, \dots, i_n}^{(0)}$  otherwise. (Structural zeroes include  $\mathbf{a}^{(0)}$  any entry at least one of whose  $(n-1)$ -dimensional marginal totals equals zero.) Then the arguments above prove:

*Corollary:* A table with structural zeroes  $\mathbf{T}$  is feasible if and only if, starting with  $\mathbf{a}^{(0)}$  as above, iterative proportional fitting with respect to the  $(n-1)$ -dimensional marginal totals of  $\mathbf{T}$  converges for each internal entry of  $\mathbf{T}$  that is not a structural zero.

Our final set of results requires some preliminaries from linear programming theory. One class of integer linear programs that is well-studied is based on linear constraint systems exhibiting totally unimodular structure. Namely, a matrix is *totally unimodular* if all of its square submatrices have determinant  $-1$ ,  $0$ , or  $+1$ . This condition obviates the need for integer division in the computation of matrix inverses, and hence guarantees integer solutions to feasible integer problems with totally unimodular systems of constraints.

Clearly, all entries of a totally unimodular matrix must equal  $-1$ ,  $0$ , or  $+1$ , and immediately a connection is apparent between totally unimodular matrices and systems of *aggregation equations* that define one-, two- and higher-dimensional tables and other structures familiar to statistical science.

A particular, but ubiquitous, form of totally unimodular matrix is that associated with a *network flow* problem. A network  $N$  consists of objects called *nodes* (denoted by circles or dots) and other objects called *directed arcs* (denoted by directed line segments) between ordered pairs of nodes. The first connection with aggregation is to consider each arc to be a variable and each node to be an aggregation equation defined by the condition that the sum of flow along arcs directed out of a node equals the sum of flow along arcs directed into the node plus a balancing constant known as the *node requirement*. The simplest, but also suitably general, form for a network is a *bipartite* network. A bipartite network  $N(\mathbf{T})$  corresponds to a two-dimensional statistical table  $\mathbf{T}(\mathbf{bc})$ .

$\mathbf{T}\mathbf{x}\mathbf{T}$  denotes a generic  $(n+1)$ -dimensional statistical table of the form  $\mathbf{T}\mathbf{x}\mathbf{T}(\mathbf{d}_1, \dots, \mathbf{d}_n, \mathbf{2})$ . Feasible  $(n+1)$ -dimensional tables are nonnegative solutions of:

$$\begin{bmatrix} \mathbf{M}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{M}_2 \\ \mathbf{I} & \mathbf{I} \end{bmatrix} (\mathbf{X}_1, \mathbf{X}_2) \begin{bmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \\ \mathbf{A}_3 \end{bmatrix}$$

for  $\{0, 1\}$ -matrices  $\mathbf{M}_1, \mathbf{M}_2$ , integer matrices  $\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3$ .

We say that a generic  $n$ -dimensional table  $\mathbf{T}$  is *totally integer* if  $\mathbf{M}$  is totally unimodular, and  $\mathbf{T}$  is *network* if  $\mathbf{M}$  is network. From the general theory,  $\mathbf{T}$  is totally integer if and only if for all integer  $\mathbf{A}$  the solutions of  $\mathbf{M}\mathbf{X} < \mathbf{A}$  are integer.

*Theorem 3:*  $\mathbf{T}$  is totally integer if and only if  $\mathbf{T}\mathbf{x}\mathbf{T}$  is totally integer.

*Proof:* Omitted. See [1]. *Q.E.D.*

A stronger result, holds, enabling efficient optimal estimation of missing integer values.

*Theorem 5:*  $\mathbf{T}$  is network if and only if  $\mathbf{T}\mathbf{x}\mathbf{T}$  is network.

*Proof:* Omitted. See [1]. *Q.E.D.*

*Corollary:*  $\mathbf{T}(2_n), \mathbf{T}(2_n\mathbf{b})$  and  $\mathbf{T}(2_n\mathbf{bc})$ ,  $b, c > 3, n > 0$ , are network.

*Theorem 6:*  $\mathbf{T}$  is totally integer, and network, if and only if  $\mathbf{T} = \mathbf{T}(2_n), \mathbf{T}(2_n\mathbf{b})$  or  $\mathbf{T}(2_n\mathbf{bc})$ ;  $b, c > 3, n > 0$ .

*Proof:* Omitted. See [1]. *Q.E.D.*

## 5. Discussion

We have shown that it incorrect to assume that mathematical properties that hold for twodimensional tables necessarily hold in higher dimensions: algorithms for higherdimensional problems based on such assumptions are prone to fail. Unanticipated failures can produce incorrect results and operational problems in large-scale data processing and analysis environments. They can cause irreconcilable inconsistencies in statistical data base query systems, particularly dynamic systems. Failures can go undetected.

We have shown that infeasibility and failure of integrality are ubiquitous in higher dimensions and cannot be ignored. Because data items are often subjected to statistical adjustment, imputation, rounding, etc., independently and due to the need to constantly merge or create new data, there is every likelihood that infeasible tables can be created or integrality lost in complex data base environments. We have identified necessary and sufficient conditions on the  $(n-1)$ -dimensional marginal distributions ensuring the existence of a feasible  $n$ -dimensional joint distribution, and presented an empirical test to detect infeasibility. We have characterized completely those higher-dimensional tables where integrality is assured.

A natural next direction for these inquiries are linked two-dimensional tables, linked

higher-dimensional tables, and more general structures. This promises to be challenging, e.g., this simple example of linked one-dimensional tables fails total unimodularity:

$$\begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} \quad \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} \quad \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}$$

## References

- [1] Cox, L.H. (2000a). On properties of multi-dimensional statistical tables. Submitted.
- [2] Cox, L.H. (2000b). Bounding entries in 3-dimensional transportation arrays. Submitted.
- [3] Cox, L.H. and L.R. Ernst (1982). Controlled rounding. *INFOR* **20**, 423-432.
- [4] Cox, L.H. (1987). A constructive procedure for unbiased controlled rounding. *Journal of the American Statistical Association* **82**, 520-524.
- [5] Ernst, L.R. (1989). Further applications of linear programming to sampling problems. Technical Report—Census/SRD/RR-89-05. Washington, DC: Statistical Research Division, U.S. Census Bureau, 33pp. Available: <http://www.census.gov/srd/papers/pdf/rr89-05.pdf>
- [6] Deming, W.E. and F.F. Stephan (1940). On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *Annals of Mathematical Statistics* **11**, 427-444.
- [7] Buzzigoli, L. and A. Giusti (1999). 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**. Luxembourg: EUROSTAT. 131-147.
- [8] Fienberg, S.E. (1999). Fréchet and Bonferroni bounds for multi-way tables of counts with applications to disclosure limitation. **Statistical Data Protection: Proceedings of the Conference**. Luxembourg: EUROSTAT. 115-129.
- [9] Chowdhury, S.D., G.T. Duncan, R. Krishnan, S.F. Roehrig and S. Mukherjee (1999). Disclosure detection in multivariate categorical databases: auditing confidentiality protection through two new matrix operations. *Management Science* **45**, 1710-1723.