# ON PROPERTIES OF MULTI-DIMENSIONAL STATISTICAL TABLES

**Lawrence H. Cox**
**U.S. Environmental Protection Agency**

*Abstract*

Statistical data are often organized in tabular form. Count data are nonnegative integers, and often magnitude data are made to take nonnegative integer values. Two-dimensional tables enjoy mathematical properties on which important statistical methods depend , e.g., for stratified sampling, imputation, disclosure limitation, and sampling and fitting log-linear models to contingency tables. We demonstrate that many of these desirable mathematical properties, and consequently their associated statistical methods, are not extendible to three or higher dimensions. We demonstrate that ill-behaved examples are ubiquitous, abundant and consequently not mathematical anomalies. To address these shortcomings, we provide necessary and sufficient conditions and an empirical test for the existence of an n-dimensional table with prescribed (n-1)-dimensional marginal totals (*feasibility*) and a complete characterization of n-dimensional tables for which the existence of integer-valued entries and associated optima are assured (*integrality*).

Keywords: contingency table, rounding, log-linear models, imputation, iterative proportional fitting, stratified sampling, statistical data base, total unimodularity, mathematical network

## 1. Introduction

Two-dimensional tables are a staple of statistical science. Typically, table entries are nonnegative and often required to be integers, e.g., contingency tables. Important statistical methods depend upon underlying mathematical properties of two-dimensional tables. Modern methods for displaying, retrieving, computing and analyzing statistical data make it 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 all but a well-characterized set of situations.

Section 2 describes eight mathematical properties of two-dimensional tables, and associates to them important methods from statistical science. For each property, Section 3 provides a counterexample to extending it, and consequently associated statistical methods, to n-dimensions, $n \geq 3$. We demonstrate that ill-behaved examples are ubiquitous, abundant and therefore not mathematical anomalies, and moreover are useful to provide insight into differences in mathematical structure between two- and higher-dimensional tables. Detecting infeasibility is important, e.g., in a public access statistical data base query system. We provide necessary and sufficient conditions and an empirical test for feasibility based on iterative proportional fitting. Section 4 provides selected results from integer linear programming theory that are used in Section 5 to establish necessary and sufficient conditions for assuring the existence of integer-valued solutions in higher-dimensions, enabling the extension of important statistical procedures. Section 6 provides discussion.

## 2. Properties of Two-Dimensional Statistical Tables

A *two-dimensional statistical table*, denoted **T(bc)**, b, c $\geq 1$, comprises b row equations, c column equations and a grand total equation:

$$\sum_{j=1}^{c} a_{ij} = a_{i.}, \quad \sum_{i=1}^{b} a_{ij} = a_{.j}, \quad \sum_{i=1}^{b} a_{i.} = \sum_{j=1}^{c} a_{.j} = a_{..} \qquad (1)$$

The typical and challenging case is a *positive table*, $a_{ij} \geq 0$, assumed henceforth. The vector $\mathbf{A} = ((\mathbf{a}_{i.}), (\mathbf{a}_{.j}))^t$ is the vector of *one-dimensional marginal totals* for **T**. The grand total equation assures that the one-dimensional marginal totals are *consistent*. Given **A**, a *feasible* (two-dimensional) table is an assignment of nonnegative *internal entries* $a_{ij}$ satisfying conditions (1). 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.

Given consistent marginal totals for a two-dimensional table, it is always possible to construct a feasible table, viz., the "independence solution" $a_{ij} = a_{i.}a_{.j}/a_{..}$ . Property 1 assures that a two-dimensional joint distribution can always be fit to consistent one-dimensional (viz., (n-1)-dimensional) marginal distributions. This property is crucial in discrete multivariate analysis and imputation. A method important to both is iterative proportional fitting (Deming and Stephan 1940). A recursive method developed to proportionally fit sample observations to known marginal totals (e.g., from a complete enumeration), iterative proportional fitting is more recently used to obtain maximum likelihood estimates of internal entries under a specified log-linear model and its sufficient statistics (certain fixed marginal totals). If starting values exhibit the model, convergence is assured (Bishop et al. 1975, Chapter 3). A feasible starting solution is not required to obtain a feasible fit to consistent marginal totals. As demonstrated in Section 3, convergence depends critically on Property 1, a fact heretofore taken for granted.

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

The *Fréchet upper bound* for an internal entry is the minimum of its two one-dimensional marginal totals. An upper or lower bound on an internal entry is *exact* if it is achieved in at least one feasible table. That the Fréchet upper bound is exact can be seen from a simple *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 has been reduced to zero.

*Property 3*: Fréchet lower bounds in two-dimensional tables are exact.

In a two-dimensional table, the *Fréchet lower bound* for internal entry $a_{ij}$ equals $\max\ \{0,\ a_{i.} + a_{.j} - a_{..}\}$. That this is a lower bound follows from observing that: $a_{i.} + a_{.j} - a_{..} = a_{ij} - \sum_{I \neq i,\ J \neq j} a_{IJ}$ and that the latter sum is nonnegative. Exactness follows by setting all values in the sum to zero. Note that this method and stepping stones provide alternative demonstrations of Property 1.

Properties 2 and 3 are important in statistical disclosure limitation (U.S. Department of Commerce 1994). Certain table entries may represent disclosure, e.g, frequency counts of one or two or an aggregate such as total income dominated by one or two contributors. Disclosure limitation methods modify the values of some or all table entries to thwart unauthorized disclosure of personal or business information. This may be accomplished by rounding, adding random noise (*perturbation*) or suppressing selected entries. Rounding and perturbation are based on a variant of stepping stones (Cox 1987). When entries are suppressed, it is necessary to assure that confidential entries cannot be estimated too narrowly from the released table. This *disclosure audit* amounts to computing exact upper and lower bounds on suppressed values, viz., Properties 2 and 3. Imputation and adjustment of entries also rely on Properties 2 and 3, viz., if internal table entries are to be imputed based on one-dimensional marginal totals and other information, it is desirable to establish the feasible interval for each imputation beforehand. Methods for sampling contingency tables with fixed marginals should also benefit from feasible bounds on internal entries.

*Property 4*: A consistent pair of integer one-dimensional marginal totals assure the existence of a feasible integer two-dimensional table.

Property 4 is assured by Property 1 and the stepping stones algorithm. It also relies upon a mathematical property (total unimodularity) discussed in Section 4. Property 4 is crucial to imputation in two-way contingency tables and to two-way stratified sampling, which requires determining integer sample sizes for each row-by-column category that are additive to predetermined integer sample sizes for row and column categories (Causey et al. 1985).

*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 bxb table **T($b^2$)** with all one-dimensional marginal totals equal to one, there are b! feasible integer tables. For disclosure limitation purposes, the existence of relatively few alternative integer feasible solutions could constitute disclosure.

*Property 6*: Covers and alternating cycles exist for nonzero entries in two-dimensional tables.

Given a two-dimensional table with at least two nonzero internal entries in each row or column, it is possible to construct one or more *covers* containing any selected nonzero entry, i.e., a set of nonzero entries such that, for each entry in the set, at least one additional entry from its row and likewise one from its column are also in the set. Minimal covers form *alternating cycles*, i.e., two entries from each covered row and column are present and each is assigned a different sign (+/-). Alternating cycles are necessarily of even length. Alternating cycles are always available in two-dimensions and enable *balanced adjustment* (a.k.a. *controlled perturbation*) of internal entries subject to preserving additivity to marginal totals, e.g., for rounding (Cox et al. 1986). In n-dimensions, a cover requires at least one additional entry from each of the n generalized "rows" containing the entry. Generalizations of (alternating) cycles to $n \geq 3$ dimensions, when they exist, are called *circuits*.

Often one or more entries in a statistical table is fixed to some value. As this is equivalent to subtracting the entry from the table and fixing the remaining value to zero, it is referred to as *zero-restriction*. Zero-restrictions arise in several statistical settings, e.g., structural (sometimes, sampling) zeroes, values already rounded, or values publicly known or which should not be changed (viz., as mandated by law). Feasible zero-restrictions can be preserved in two-dimensions. Zero-restricted alternating cycles enable balanced adjustment by alternatively adding/subtracting a positive quantity from selected entries. Balanced adjustment is essential to disclosure limitation in two-dimensional tables, to unbiased controlled rounding (see Property 8), and to methods for imputing/adjusting table entries. See Cox et al. (1986) for details.

Feasible zero-restrictions do not pose obstacles for such methods in two-dimensional tables. For example, consider the alternating cycle in Figure 1. If, for example, the marginal entries are all multiples of some integer rounding base, then it is possible to adjust the entries marked +/- so that the resulting table is both additive and rounded to the same base. See Cox (1987) for the complete method.

<<<<<<<<<<<Figure 1 HERE>>>>>>>>>

*Property 7*: 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 7 is assured by Properties 1-4. In the absence of Property 7, an exact continuous bound on an integer entry typically does not pinpoint the exact integer bound (e.g., an exact continuous bound is not necessarily numerically adjacent to an exact integer bound). Indeed, unsupported by other methods, continuous bounding methods, e.g., linear programming, can be insensitive to whether or not integer solutions even exist.

*Property 8*: Controlled rounding is assured in two-dimensional tables.

Given an integer rounding base B, a *controlled rounding* of a statistical 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. Cox and Ernst (1982) demonstrate that (zero-restricted) controlled rounding of a two-dimensional table is always possible. Cox (1987) 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, 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 (Causey et al. 1985).

**3. Failure of The Properties to Extend to Three or More Dimensions and Effects on Statistical Methods**

3.1 <u>Three-Dimensional Counterexamples to Extendibility of The Properties</u>

In this section, failure of the preceding eight 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, ..., $d_1$, j = 1, ..., $d_2$, k = 1, ..., $d_3$. 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.

*Examples 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 defined by the index over which summation is performed (for n = 2, these are the sets of row and column totals). If these sets of (n-1)-dimensional marginal totals do in fact 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 of Section 2. 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 lower-dimensional 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 Examples 1a, b: consistent (n-1)-dimensional marginal totals do not guarantee the existence of a feasible table in three and higher dimensions. Consequently, in $n \geq 3$ dimensions, consistent sets of (n-1)-dimensional marginal distributions do not guarantee existence of an underlying n-dimensional joint distribution.

<<<<<<<<<<Example 1a HERE>>>>>>>>>>

Despite the fact that Property 1 fails in $n \geq 3$ dimensions, important statistical methods depend, sometimes subtly, on the assumption that Property 1 holds. Iterative proportional fitting is one such method. Algorithms are available for iterative proportional fitting in n $\geq$ 3 dimensions (Bishop et al. 1975, Sections 3.5.1-2), as follows. Beginning with a log-linear model, the minimum number of configurations of sufficient statistics for this model, and a set of starting values for internal entries exhibiting the structure of the model, a sequence of proportional fits (one to each configuration) is made. This process is iterated and, in two-dimensions, it converges. However, as shown below, convergence in higher-dimensions depends critically on the existence of a feasible table. From the general theory, for any log-linear model, all sufficient statistics can be derived from the (n-1)-dimensional marginal totals and $a^{(0)}_{i_1,\ldots,i_q,\ldots i_n} = 1$ can be used for starting values.

As an illustration, consider the "no n-factor effect" model, for which the usual n sets of (n-1)-dimensional marginal totals $A_q = \{a_{i_1,\ldots+,\ldots i_n}\}$ are sufficient statistics (summation over the q[th] index only). Use $a^{(0)}_{i_1,\ldots,i_q,\ldots i_n} = 1$ as
starting values. At its first iteration, the iterative proportional fitting algorithm sets:

$$a^{(1)}_{i_1,\ldots,i_n} = \left( \frac{a^{(0)}_{i_1,\ldots,i_n}}{\sum_i a^{(0)}_{i,i_2,\ldots,i_n}} \right) a_{+,i_2,\ldots,i_n}.$$

Iteration is on each index$\{i_1,\ldots,i_n\}$ (inner loop), each dimension q (mod n) = 1,..,n - 1, 0 (middle loop), and each cycle m = 1,......, (outer loop), as follows:

$$a^{((m-1)n+q)}_{i_1,\ldots i_q,\ldots,i_n} = (a^{((m-1)n+q-1)}/a^{(m)}_{i_1,\ldots,+,\ldots,i_n})a_{i_1,\ldots,+,\ldots,i_n} \qquad (2)$$

where $a_{i_1,\ldots,+,\ldots,i_n} \in A_q$ are the fixed marginal totals and $a^{(m)}_{i_1,\ldots,+,\ldots,i_n}$ are the (n-1)-dimensional sums of current values (cycle m, iteration q), both summations across dimension q (only).

Theorem 3.5-1 of Bishop et al. (1975) assures convergence of this algorithm. However, this theorem depends subtly on Property 1, viz., that the marginals were constructed from original feasible values assumed to exist, even though the fitting problem is motivated and the algorithm is well-defined only in terms of the sufficient statistics, which in all cases equal or are derivable from the (n-1)-dimensional marginals.

For example, if n-dimensional entries are not available but the (n-1)-dimensional marginals are, then it is possible and reasonable to attempt to attempt to fit a log-linear model directly to these marginals. However, if, unbeknownst to the analyst, the marginals have been corrupted, e.g., by rounding, then a feasible table may not exist, and the iterative proportional fitting algorithm can fail to converge, viz., to produce a final result. This is illustrated in Example 1b. An attempt to fit the no three-factor effect model (or any log-linear model) fails, viz., in lieu of convergence, the n = 3 subsequences of proportional estimates for the (1, 1, 1) entry converge to three different values: 10, 13.0278 and 16.9722. We return to this issue in Section 3.3.

<<<<<<<<<<Example 1b HERE>>>>>>>>>>

Several statistical procedures proven in two dimensions are *insensitive* to infeasibility in n $\geq$ 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 two-dimensional procedure to higher dimensions. For example, generalizations of Fréchet bounds (see Examples 2, 3 below) are insensitive to infeasibility. Such bounds, defined by a finite set of arithmetic operations on the (n-1)-dimensional marginals, are always computable. However, if the marginals define an infeasible table, as in Examples 1a, b, meaningless bounds are produced, carrying with them the implication that actual feasible values residing between these bounds exist, when in fact there are no feasible values. Such putative bounds are certainly misleading. By virtue of Examples 1a,b, 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. Buzzigoli and Giusti (1999), Fienberg (1999) and Chowdhury et al. (1999) offer methods for bounding internal entries in three-dimensional tables for disclosure limitation purposes. The first two methods employ these arithmetic operations in conjunction with iteration of subadditive relations. They are provably inoptimal and insensitive (Cox 2000). Chowdhury et al. (1999) is provably optimal but also provably insensitive, as follows.

Chowdhury et al. (1999) 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 if, e.g.., as the authors suggest, the underlying table consists of counts of patient-by-doctor-by-treatment occurrences. The three-dimensional data are confidential, as are the patient-by-treatment marginal totals, but assume that the doctor-by-treatment and the patient-by-doctor views are nonconfidential and released. The problem for the data user is then to obtain optimal lower and upper bounds on the missing patient-by-treatment marginals, for which the authors provide a network optimization algorithm.

Consider the i = 4 and j = 4 views from Example 4a, and assume that the objective is to estimate the third view (viz., the vertical plane of marginal totals, k = 4), presumed missing, using Chowdhury et al. (1999). Example 1c results. The Chowdhury bounds are exact if the three views correspond to a feasible three-dimensional table. However, the Chowdhury method is insensitive to feasibility, as demonstrated in Example 1c which displays the Chowdhury estimates for the k = 4 view of Example 4a, when in fact no table exists.

<<<<<<<<<<Example 1c HERE>>>>>>>>>>

An argument that can be raised surrounding these and examples to follow is that infeasible tables do not arise in statistics and therefore are of no practical interest. Clearly, given feasible internal entries, a table with marginals is defined and all of the statistical methods described in this paper, and many others, can be performed properly. However, complete feasible tables are not always available. Marginal totals are frequently not fixed, derived directly by addition from internal entries, free of error, or precisely known. For example, each set of (n-1)-dimensional marginal totals for a presumed 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 take flawed and incomplete data and attempt to combine it for their purposes. All of this can lead to infeasibility. If infeasibility were a rare occurrence, perhaps the problem could for practical purposes be ignored. However, as demonstrated in Section 3.2, infeasibility is anything but rare in $n \geq 3$ dimensions.

*Example 2*: Fréchet upper bounds are not exact in $n \geq 3$ dimensions.

In n-dimensions, the Fréchet upper bound of an internal entry is the minimum of the n (n-1)-dimensional marginal totals to which the entry contributes. In Example 2, the (3, 3, 1) entry (lower right corner of 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 corresponding to these marginal totals.

<<<<<<<<<<Example 2 HERE>>>>>>>>>>

*Example 3*: Fréchet lower bounds are not exact in n $\geq$ 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 3 HERE>>>>>>>>>>

*Examples 4*: Fréchet consistency does not guarantee the existence of a feasible n-dimensional table.

Although the (n-1)-dimensional marginal totals of Example 1a are consistent, there is an inconsistency among some of the two- and one-dimensional marginal totals, as follows. Consider the (2, 1, 1) entry of Example 1a. Its Fréchet upper bound equals 0, while its Fréchet lower bound equals 1. Thus, a feasible table cannot exist.

Call a table *Fréchet consistent* if the (n-1)-dimensional marginal totals are consistent and if also all Fréchet lower bounds are less than or equal to their corresponding Fréchet upper bounds. It is tempting to conjecture that Fréchet consistency guarantees feasibility. Unfortunately, Example 4a demonstrates otherwise.

<<<<<<<<<<Example 4a HERE>>>>>>>>>>

Examples 1 and 4a are important to disclosure limitation in public use statistical data base query systems. Even though all original tables are feasible, it is possible that infeasibility can result through iterative, independent adjustment of entries, such as for display (viz., rounding) or confidentiality purposes. This possibility is increased if the data base is updated dynamically or if it releases population estimates derived from sample observations. For example, consider the feasible table in Example 4b. Zero-restricted, additive integer rounding (B = 1) of the totals entries of this example could lead to a feasible table (e.g., Example 2) or an infeasible one (e.g., Example 4a).

<<<<<<<<<<Example 4b HERE>>>>>>>>>>

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

The feasible solutions to Example 5 can be parameterized by three linearly constrained continuous variables, whereas the number of feasible integer solutions is only four.

<<<<<<<<<<Example 5 HERE>>>>>>>>>>

*Example 6*: Covers may exist, but alternating circuits may not, in $n \geq 3$ dimensions.

<<<<<<<<<<Example 6 HERE>>>>>>>>>>

Example 6 illustrates a 3x3x3 table with 11 of its 27 entries zero-restricted. The remaining 16 entries, denoted “*”, are assumed to be positive. Whereas this pattern of * entries provides a cover for these 16 values, this cover is not a circuit (viz., it is impossible to assign alternating “+” and “-“ signs to the * entries). Therefore, irrespective of the values of a consistent set of (n-1)-dimensional marginal totals, it is not possible to modify these values in a controlled manner. In particular, controlled perturbation, such as for disclosure limitation (Cox et al. 1986) is impossible. Any algorithm seeking to do so must be heuristic and must fail.

*Examples 7*: Generalized circuits of odd length occur in $n \geq 3$ dimensions. So do fractional optima.

Example 7a exhibits a 3x3x3 table **T** whose entries designated “0" are zero-restricted, while those designated “+” or “-“ are set to positive values. Whereas **T** does not admit an alternating circuit among its nonzero entries, it does admit a unique *generalized circuit* of nonzero entries of odd length (17) centered around the (3, 2, 1) entry (the **+** in boldface). The circuit is not alternating, but it enables adjustment of the table by moving a positive quantity into the (3, 2, 1) entry through the nonzero entries of **T**, as follows. Alternatively add/subtract quantity q to/from internal entries of **T** in this manner: 111+, 121-, 122+, 112-, 212+, 213-, 223+, 323-, 313+, 311-, **321+**, 221-, 231+, 232-, 332+, 322-, **321+**, 331- (with obvious abuse of notation). This results in addition of quantity 2q to the (3, 2, 1) entry. By reversing the roles of “+” and “-“, it is possible to move a quantity 2p out of this entry.

<<<<<<<<<<Example 7a HERE>>>>>>>>>>

Example 7b is an elaboration of Example 7a, with marginal entries included. Three internal entries, marked “0", are zero-restricted. The maximum (continuous) value of each of the eight other entries marked “+” in Example 7b (viz., excluding the (3, 2, 1) entry) is ½. Similarly, the minimum value of each of the eight entries marked “-“ is ½. Thus, only the (3, 2, 1) entry has integer optima. This can be verified via linear programming or directly, as follows. Set the (3, 2, 1) entry to zero and all other marked entries to ½. This corresponds to a feasible solution **x*** (viz., a feasible table). As the value of the (3, 2, 1) entry cannot decrease, then neither can any of the entries marked “-“. Similarly, as its value cannot exceed one then the values of the entries marked “+” cannot exceed ½. **x*** is an extreme point of the polytope of feasible tables. There is precisely one other extreme point, viz., corresponding to setting the (3, 2, 1) entry to one, and this extreme point is integer.

<<<<<<<<<<Example 7b HERE>>>>>>>>>>

Linear programming reveals that the maximum continuous value of the (3, 3, 1) entry in Example 7c is ½. This is the only fractional optimum, and is achieved without explicit zero-restrictions. Roehrig (1999) presented a similar, perhaps the first, such counterexample; Example 7a first appeared in Cox (1999). These examples illustrate 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 (U.S. Department of Commerce 1994, Recommendation 11).

<<<<<<<<<<Example 7c HERE>>>>>>>>>>

*Example 8*: Controlled rounding is not always possible in n ≥ 3 dimensions.

Ernst (1989) provides a counterexample to three-dimensional controlled rounding.

3.2 Counterexamples are Abundant

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 5x5x5 table, place Example 1a in the upper-front-left of the 5x5x5 space, and place Example 4a 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 Section 5 for exceptions). Similar constructions are possible for the other examples.

$\mathbf{T}(d_1, \ldots, d_n; A)$ denotes an n-dimensional statistical table of *size* $(d_1, \ldots, d_n)$. 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 $MX = A$, 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 $T = T(d_1, \ldots, d_n)$ as a *generic* n-dimensional table, viz., $T(d_1, \ldots, d_n; A)$ for arbitrary **A**. Shorthand such as $\mathbf{T(2^n)}$ or **T(bc)** is used when the meaning is clear, and $mT(d_1, \ldots, d_n; A) = T(d_1, \ldots, d_n; mA)$. Eliminate all linearly dependent rows from **M**. Let $\mathbf{M_B}$ be a nonsingular submatrix of **M** of maximal rank. Reorder the columns (variables) of **M** and **A** so that $M = (M_B, M_N)$ and $A = (A_B, A_N)$. Then the *basic solution* of $MX = A$ corresponding to $M_B$ is $x = (M_B^{-1} A_B, 0)$.

We have shown that counterexamples to feasibility $\mathbf{T(d_1, \ldots, d_n; A)}$ exist in dimension n ≥ 3 whenever $d_j > 3$ for three or more distinct values of j. 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 3.1***: Let $T = T(d_1, \ldots, d_n)$ denote a generic table satisfying: n ≥ 3 and $d_j > 3$ for at least three distinct values of j. Let $T_f = T(d_1, \ldots, d_n; A_f)$ denote a feasible table and $T_i = T(d_1, \ldots, d_n; A_i)$ denote an infeasible table. Then there exists an integer *p* such that $T_f + mT_i$ is infeasible for all m ≥ p.

*Proof*: To say that $\mathbf{T_i}$ is infeasible is equivalent to saying that every basic solution of $MX = A_i$ contains at least one negative entry. As the set of basic solutions of $MX = A_f$ is bounded, then there exists an integer p such that, whenever m ≥ p, every basic solution of $MX = A_f + mA_i$ contains at least one negative entry. *Q.E.D.*

Theorem 3.1 demonstrates that infeasibility in higher dimensions is pervasive, and not a mathematical anomaly. Moreover, infeasibility is not easily recognized, viz., whereas the presence of zeroes and small values in our examples were critical to their construction and made these examples easy to recognize (e.g., Example 1a), adding a feasible table to a sufficiently large multiple of an infeasible table produces a less discernible but nonetheless infeasible table. As an illustration, consider infeasible Example 1b. It is formed by adding a feasible table to three times Example 1a and multiplying the result by 10. Infeasibility here is much more difficult to detect than in 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, and that the methods of Buzzigoli and Giusti (1999) and Fienberg (1999) are insensitive. As observed by Roehrig (1999), because these two procedures proceed towards upper bounds from outside the feasible region, viz., from larger to smaller values (from smaller to larger for lower bounds), then they must terminate at or before reaching the integer part of the optimal upper bound. If the continuous upper bound is noninteger (e.g., Examples 7b, c), then necessarily these algorithms must fail to identify the optimal integer upper bound. Similarly, heuristic algorithms for controlled data perturbation (e.g., Duncan and Fienberg 1999), 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, e.g, Fréchet consistency. A statistical approach might yield insight into compatibility conditions between an n-dimensional joint distribution and its marginals. Fortunately, there is a statistical method that can be used to detect infeasibility.

3.3 A Test for Feasibility

*Theorem 3.2 (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*: (Only if): This is Theorem 3.5-1 of Bishop et al. (1975).

(If): The sufficient statistics corresponding to starting value of all ones are the (n-1)-dimensional marginal totals. From (2), observe that for each m and q:

$$\sum_{i_q} a^{((m-1)n+q)}_{i_1,\ldots,i_q,\ldots i_n} = a^{((m-1)n+q)}_{i_1,\ldots,+,\ldots i_n} = a_{i_1,\ldots,+,\ldots i_n} \in A_q \tag{3}$$

viz., at the end of each middle loop of the iteration (indexed by q), the current values must add correctly to the original marginal totals in at least one direction (viz., that with index = q). This result holds for all values of m, and hence as $m \to \infty$. Consequently, if the iterative proportional fitting algorithm converges for all internal entries, then there must exist M* such that given $\varepsilon > 0$, whenever M ≥ M*, we have:

$\left| \sum_{i_q} a^{((M-1)n+q)}_{i_1,\ldots,i_q,\ldots i_n} - a_{i_1,\ldots,+,\ldots,i_n} \right| < \varepsilon$ for all q = 1,..., n. But this is precisely the statement that the convergence limits define a feasible table. *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 (as in Example 1b), 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 **T**, define $a^{(0)}_{i_1,\ldots,i_q,\ldots i_n} = 0$, and define $a^{(0)}_{i_1,\ldots,i_q,\ldots i_n} = 1$ otherwise. (Structural zeroes include 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 **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 **T** converges for each internal entry of **T** that is not a structural zero.

## 4. Preliminaries on Integer Linear Programming

*Linear programming* is concerned with minimization of a linear objective function of continuous variables subject to linear constraints on these variables. Linear programming is well-established theoretically and usually can be accomplished in reasonable computational time, even for large problems. The opposite can be said about *integer linear programming*, viz., minimizing a linear objective function of integer variables subject to linear constraints on these variables: such problems are typically difficult or computationally infeasible to accomplish, even for moderate size problems, and considerably less theory is available. In this section, we summarize concepts and results from integer optimization needed to establish the results of the next section. The interested reader is referred to standard texts such as Nemhauser and Wolsey (1988) (and in particular their Chapter III.1 Integral Polyhedra).

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. One aim of this paper is to develop that connection.

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 corresponds to a two-dimensional statistical table, as follows.

The network ***N*(T)** corresponding to table **T(bc)** is illustrated in Figure 2 (for b = c = 3). Note that arc flows are designated by variables $x_{ij}$ (these appear as x_ij on Figure 2), and that $x_{ij} = a_{ij}$ is one, but in general not the only, *feasible solution* to the network.

<<<<<<<<<<Figure 2 HERE>>>>>>>>>>>

## 5. Integer Optima in n-Dimensions

Properties 2-4 and 6-8 depend on the total unimodularity of two-dimensional statistical tables. The efficiency of associated computations, viz., for computing exact bounds (Gusfield 1988), owes to structure of the associated network (Kennington and Helgason 1980). The examples of Section 3 demonstrate that these properties can be lost in n $\geq$ 3 dimensions. In this section, we investigate circumstances under which they will be preserved.

**TxT** denotes a generic (n+1)-dimensional statistical table of the form $\boldsymbol{TxT(d_1, \ldots, d_n, 2)}$. Feasible (n+1)-dimensional tables are nonnegative solutions of:

$$\begin{pmatrix} M_1 & 0 \\ 0 & M_2 \\ I & I \end{pmatrix} (X_1, X_2) = \begin{pmatrix} A_1 \\ A_2 \\ A_3 \end{pmatrix}$$ , for {0, 1}-matrices $\mathbf{M}_1$, $\mathbf{M}_2$ and integer matrices $\mathbf{A}_1$, $\mathbf{A}_2$, $\mathbf{A}_3$.

We say that a generic n-dimensional table **T** is *totally integer* if **M** is totally unimodular, and **T** is *network* if **M** is network. From the general theory, **T** is totally integer if and only if for all integer **A** the solutions of $\mathbf{MX} \leq \mathbf{A}$ are integer (Nemhauser and Wolsey 1988, Chapter III.1).

*Theorem 4.1*: **T** is totally integer if and only if **TxT** is totally integer.

*Proof*: (If): Trivial.

(Only if): Assume that **T** is totally integer. Let $\mathbf{x^*} = (\mathbf{x^*}_{1j}, \mathbf{x^*}_{2j})$ be an extreme point of:

$$\begin{pmatrix} M_1 & 0 \\ 0 & M_2 \\ I & I \end{pmatrix} (X_1,\ X_2) = \begin{pmatrix} A_1 \\ A_2 \\ A_3 \end{pmatrix}, \text{ for } M_1,\ M_2 \in \{0,\ 1\}, \text{ integer } \mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3 \geq \mathbf{0}, \text{ and } X_1,\ X_2 \geq 0 \tag{4}$$

Observe that **x*** achieves the maximum for the linear cost function:

$$c(x) = \sum_{x_{1j}>0} x_{1j} + \sum_{x_{2j}>0} x_{2j} = \sum_{x_{1j}>0} x_{1j} + \sum_{x_{1j}<a_{3j}} (a_{3j} - x_{1j}) \tag{5}$$

The latter optimization is over only the $\mathbf{X}_1$-constraints corresponding to one copy $\mathbf{T}_1$ of **T** (viz., over $M_1X_1 = A_1$ ) subject to additional integer capacity constraints given by: $X_1 \leq A_3$. As capacity constraints do not affect the total unimodularity of $\mathbf{M}_1$, then $\mathbf{x^*}_1$ is an integer point of the polyhedron defined by: $M_1X_1 = A_1,\ X_1 \leq A_3$. As $\mathbf{A}_3$ is integer, then $\mathbf{x^*}_2 = \mathbf{A}_3 - \mathbf{x^*}_1$ is integer, and hence **x*** is integer. Therefore, **TxT** is totally integer. *Q.E.D.*

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

*Theorem 4.2*: **T** is network if and only if **TxT** is network.

*Proof*: (If) Trivial.

(Only if) Assume that **T** is network. Consider Figures 3.

<<<<<<<<<Figures 3 HERE>>>>>>>>>>>

A network representation ***N*(TxT)** of **TxT** is obtained by applying the method illustrated in Figures 3 successively for each n-dimensional index j, as follows. Prior to introducing the (n+1)st set of (n-1)-dimensional totals $X_1 + X_2 = A_3$, two networks $\mathbf{N(T_1)}$ and $\mathbf{N(T_2)}$ identical in node-arc structure exist, viz., one for each copy of **T**. Portions of these networks containing the arc corresponding the variable $\mathbf{x_j}$ is illustrated in Figure 3a. The two networks differ only in the individual node requirements, which are determined, respectively, by the (n-1)-dimensional totals $\mathbf{A}_1$ and $\mathbf{A}_2$. Figure 3b illustrates the state of network ***N*(TxT)** at arc *j* just after introducing the (n+1)st set of (n-1)-dimensional totals. Note that there are no longer j-arcs to nodes $O_{2j}$ and $I_{2j}$: once the alterations of Figure 3b are made for all j-indexes, these nodes can be removed from the network because they represent linearly dependent equations. *Q.E.D.*

*Corollary 4.2.1*: $\mathbf{T(2^n)}$, $\mathbf{T(2^n b)}$ and $\mathbf{T(2^n bc)}$, b, c $\geq$ 3, n $\geq$ 0, are network.

*Example 9*: Let **M** denote the coefficient matrix corresponding to the table in Example 7a, b. **M** is obtained by deleting all columns corresponding to entries marked "0" from the coefficient matrix $\mathbf{M}_{333}$ of a full 3x3x3 table. A feasible solution **x*** for the table is to set the (3, 2, 1) entry to 1 and all other nonzero entries to ½.

Let **A** denote the column vector of (n-1)-dimensional totals corresponding to **x***, viz., **A** =

(1,1,1,1,1,1,0,1,1,1,1,1,1,1,1,1,1,0,1,1,0,1,1,1,1,1,1). As the (3, 2, 1) entry cannot exceed 1, then no other + entry can exceed ½, and thus **M** is not totally unimodular. As **x*** maximizes each + entry, it is an extreme point of **MX** = **A**.

*Theorem 4.3*: **T** is totally integer, and network, if and only if **T** = $\mathbf{T(2^n)}$, $\mathbf{T(2^n b)}$ or $\mathbf{T(2^n bc)}$; b, c $\geq$ 3, n $\geq$ 0.

*Proof*: (If) Corollary 4.2.1 provides the proof.

(Only if) Example 7b provides the proof: **T** of any other size must contain $\mathbf{T(3^3)}$, which by Example 7b fails to be totally integer, and therefore **T** must fail as well. Q.E.D.

**6. Discussion**

We have shown that it incorrect to assume that mathematical properties that hold for two-dimensional tables necessarily hold in higher dimensions, and consequently that statistical methods and algorithms for higher-dimensional problems based on such assumptions are prone to fail in practice. Unanticipated failures can produce incorrect results and cause serious operational problems in large-scale data processing and analysis environments such as national censuses and surveys. 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, not present in two dimensions, in higher dimensions are ubiquitous, numerous and therefore not mathematical anomalies that simply can 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. To address these shortcomings, we have identified necessary and sufficient conditions on the (n-1)-dimensional marginal distributions that ensure 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 for which integrality is assured.

Extension of properties and algorithms for two-dimensional tables to more complex structures needs to be attempted with caution. While it is appropriate for investigators and practitioners to seek to build upon that which is well-understood and familiar, it is important that differences in underlying mathematical structure between familiar and new situations first be understood. Understanding higher-dimensional structure directly benefits applications including imputation, disclosure limitation, and survey sampling. An important potential application is sampling from multi-dimensional distributions (contingency tables) with known conditional distributions (marginals). For sufficient statistics including n consistent sets of (n-1)-dimensional marginal totals, empirical distributions can be constructed using methods from algebraic geometry (Diaconis and Sturmfels 1998). However, such methods are extremely demanding computationally, viz., even for 3x3x3 tables. The ability to detect feasibility and to compute exact integer bounds on internal entries efficiently should provide computational, and perhaps theoretical, assistance.

This paper is intended to contribute to improved understanding of the structure of multi-dimensional statistical tables. Theorem 4.3, which distinguishes between higher-dimensional structures that behave like two-dimensional tables and those that do not, is useful for this purpose. Investigators offering heuristic algorithms for higher-dimensional tables tend to examine an algorithm's properties only on standard, simple examples, e.g., $\mathbf{T(2^n)}$ or **T(2bc)**. However, because, by Theorem 4.3, in essence these cases do not differ from the two-dimensional case and can be solved efficiently using networks, these examinations are likely to be misleading and disguise fundamentally different and complex properties of the higher-dimensional problem.

In the preceding sections, (n-1)-dimensional marginal totals **A** were treated as constants. In general, operations such as rounding and structures such as circuits involve both internal and marginal entries. No generality is lost, however, by treating marginals as constants: one simply increases the size of the table by one in each dimension and incorporates a variable representing all or a part of the marginal. The resulting table has constant marginals. This procedure is referred to as *folding-in* the marginal totals. See Cox et al. (1986) for details.

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:

$$\begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix}$$

**Disclaimer**

The information in this article was funded wholly or in part by the United States Environmental Protection Agency. It has been subjected to Agency review and approved for publication. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

**References**

Bishop, Y.M.M., S.E Fienberg and P.W. Holland (1975). **Discrete Multivariate Analysis: Theory and Practice**. Cambridge, MA: The MIT Press.

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.

Causey, B.D., L.H. Cox and L.R. Ernst (1985). Applications of transportation theory to statistical problems. *Journal of the American Statistical Association* **80**, 903-909.

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.

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

_____ (1999). Some remarks on research directions in statistical data protection. **Statistical Data Protection: Proceedings of the Conference**. Luxembourg: EUROSTAT. 163-176.

_____ and L.R. Ernst (1982). Controlled rounding. *INFOR* **20**, 423-432.

_____, J. Fagan, B. Greenberg and R. Hemmig (1986). Research at the Census Bureau into disclosure avoidance techniques for tabular data. *Proceedings of the Survey Research Methods Section, American Statistical Association*. Alexandria, VA. 388-393.

_____ (2000). Bounding entries in 3-dimensional transportation arrays. Manuscript. Available from author.

Deming, W.E. and F.F. Stephan (1940). On a least squares adjustment of a sampled frequency table with the expected marginals are known. *Annals of Mathematical Statistics* **11**, 427-444.

Diaconis, P. and B. Sturmfels (1998). Algebraic algorithms for sampling from conditional distributions. *Annals of Statistics* **26**, 363-397.

Duncan, G.T. and S.E. Fienberg (1999). Obtaining information while preserving privacy: a Markov perturbation method for tabular data. **Statistical Data Protection: Proceedings of the Conference**. Luxembourg: EUROSTAT. 351-362.

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

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.

Gusfield, D. (1988). A graph theoretic approach to statistical data security. *SIAM Journal of Computing* **17**, 552-571.

Kennington, J. and R. Helgason (1980). **Algorithms for Network Programming**. New York: John Wiley & Sons.

Nemhauser, G. and L. Wolsey (1988). **Integer and Combinatorial Optimization**. New York: John Wiley & Sons.

Roehrig, S.F. (1999). Auditing disclosure in multi-way tables with cell suppression: simplex and shuttle solutions. Manuscript, The Heinz School of Public Policy and Management, Carnegie Mellon University, 21 pp.

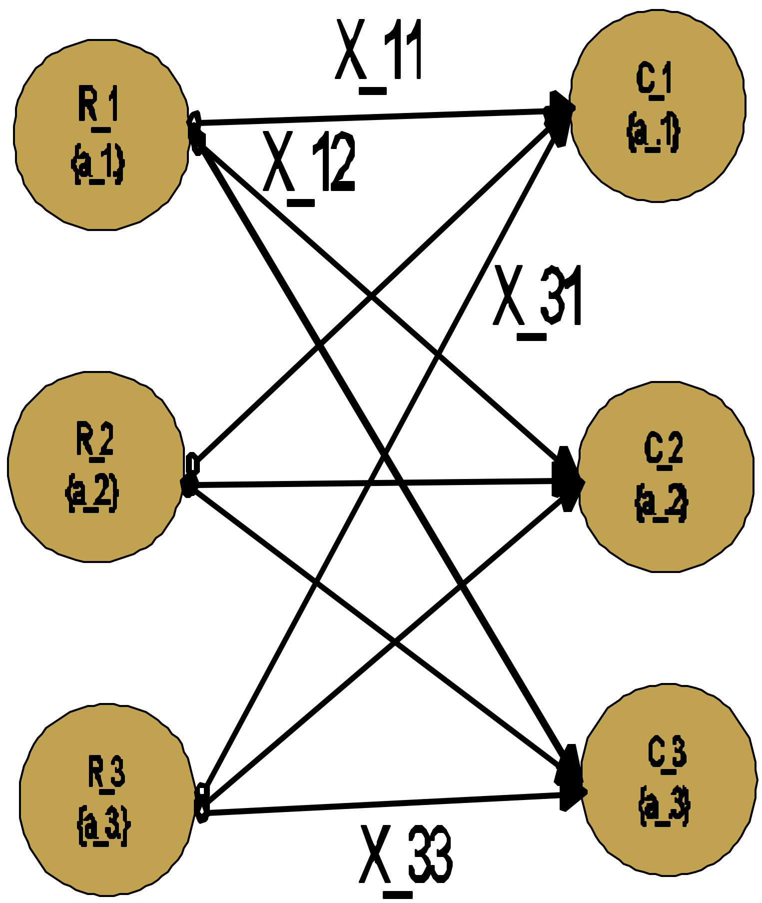
U.S. Department of Commerce (1994). **Statistical Policy Working Paper 22: Report on Statistical Disclosure Limitation Methodology**. Washington, DC: U.S. Department of Commerce (NTIS Document Sales, PB94-165305).

Original: December 1999
Current: June 2000

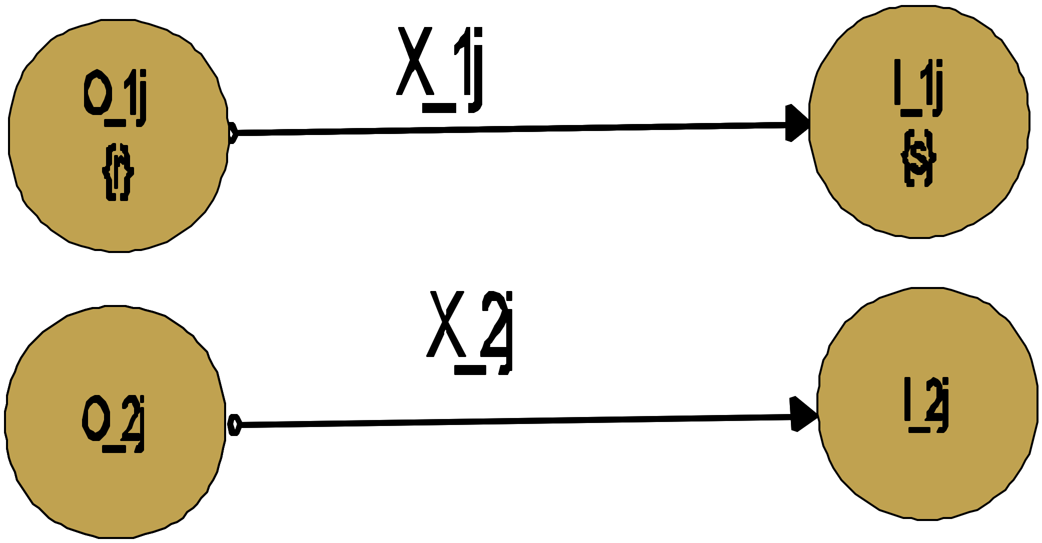
| | | | |
|---|---|---|---|
| 0 | % | 0 | & |
| % | & | 0 | 0 |
| & | 0 | % | 0 |
| 0 | 0 | & | % |

**Figure 1: An Alternating Cycle**

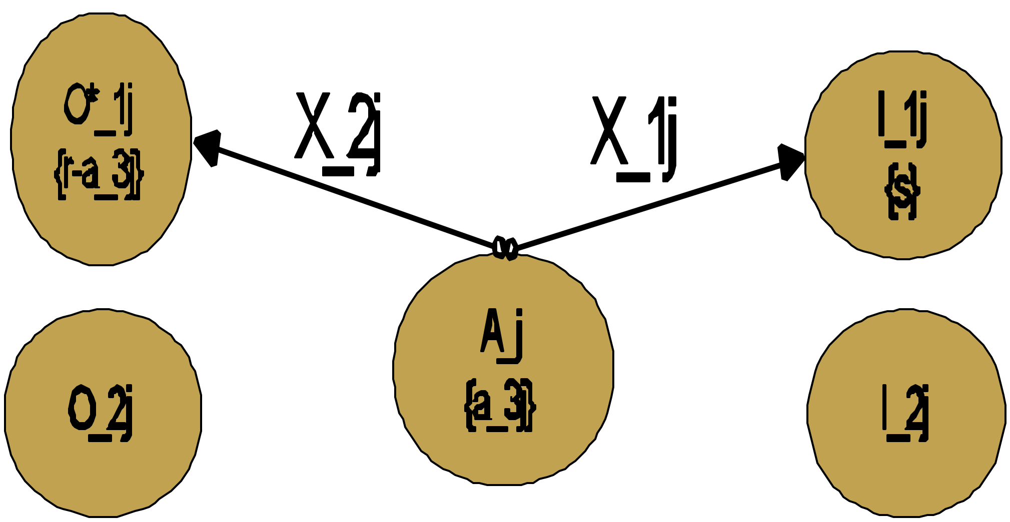
# Figure 2: Network for a 3x3 Two-Dimensional Table



## Figure 3a: Original Networks




## Figure 3b: Creating N(TxT)



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

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

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

**Example 1a: Consistent But Infeasible 3-D Table**

| | | |
|---|---|---|
| | | 10 |
| | | 30 |
| 30 | 10 | 40 |

| | | |
|---|---|---|
| | | 30 |
| | | 10 |
| 10 | 30 | 40 |

| | | |
|---|---|---|
| 30 | 10 | 40 |
| 10 | 30 | 40 |
| 40 | 40 | 80 |

**Example 1b: Consistent But Infeasible 3-D Table**

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

**i = 4 View**

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

**j = 4 View**

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

**Example 1c: Chowdhury et al. (1999) Bounds on k = 4 (Vertical) View of Example 4a**

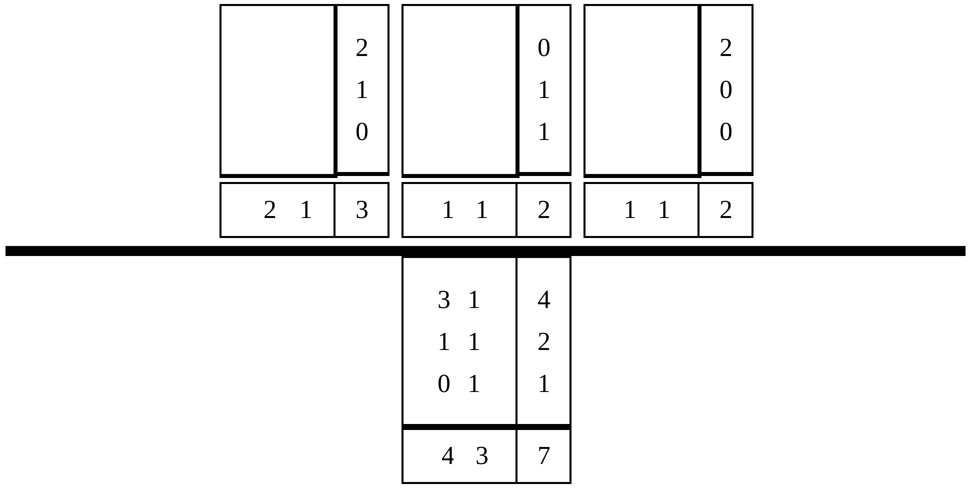
| | | | |
|---|---|---|---|
| | | | 3 |
| | | | 1 |
| | | | 1 |
| 3 | 1 | 1 | 5 |

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

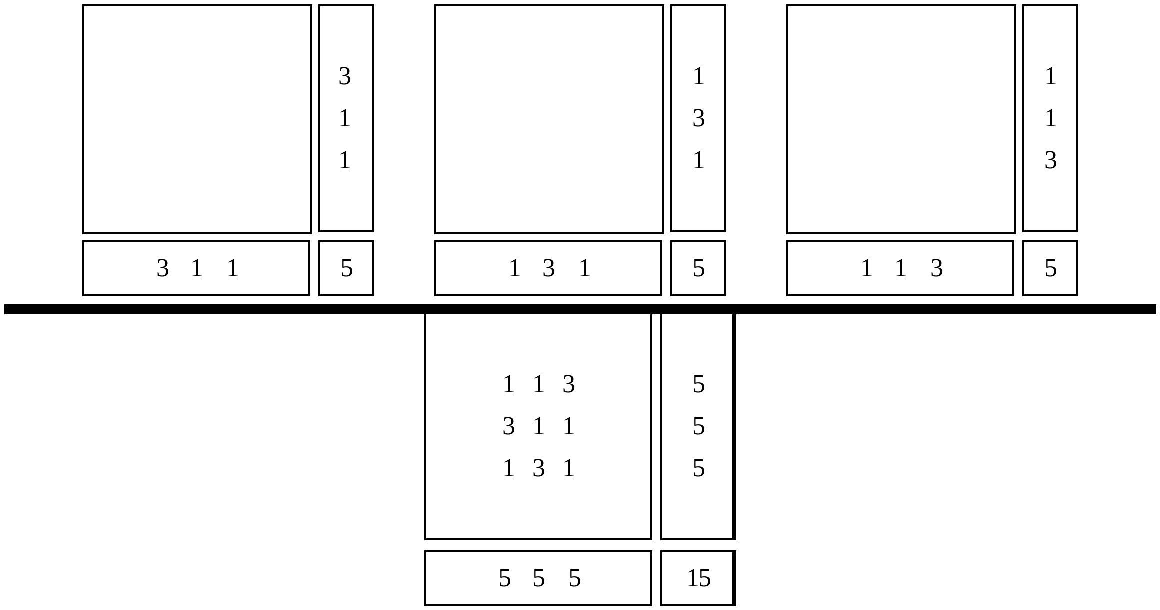
| | | | |
|---|---|---|---|
| | | | 1 |
| | | | 1 |
| | | | 3 |
| 1 | 1 | 3 | 5 |

| | | | |
|---|---|---|---|
| 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**




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




**Example 4a: Infeasible Fréchet Consistent 3-D Table**

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

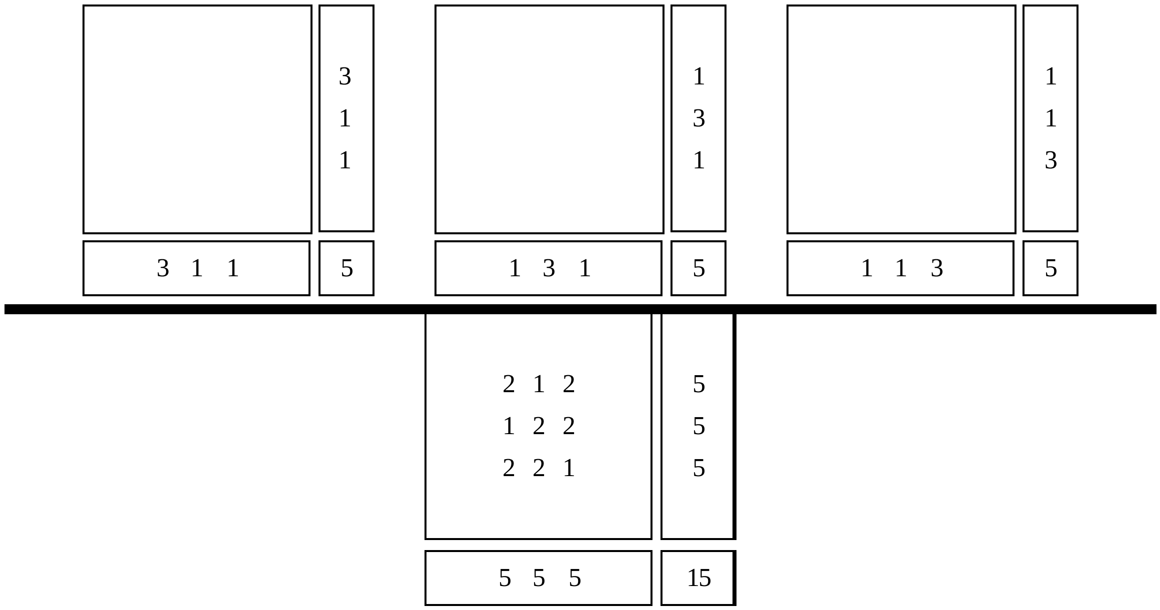
| | | | |
|---|---|---|---|
| 0 | 0.5 | 0.5 | 1 |
| 1 | 0.5 | 1 | 2.5 |
| 0 | 1.5 | 0 | 1.5 |
| 1 | 2.5 | 1.5 | 5 |

| | | | |
|---|---|---|---|
| 0.5 | 0 | 1 | 1.5 |
| 0.5 | 0 | 0.5 | 1 |
| 0.5 | 1 | 1 | 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 4b: Feasible Fréchet Consistent 3-D Table**




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

| | | |
|---|---|---|
| 0 | ( | ( |
| ( | 0 | ( |
| ( | ( | 0 |

| | | |
|---|---|---|
| ( | ( | 0 |
| ( | 0 | ( |
| 0 | ( | ( |

| | | |
|---|---|---|
| ( | 0 | ( |
| 0 | 0 | 0 |
| ( | 0 | ( |

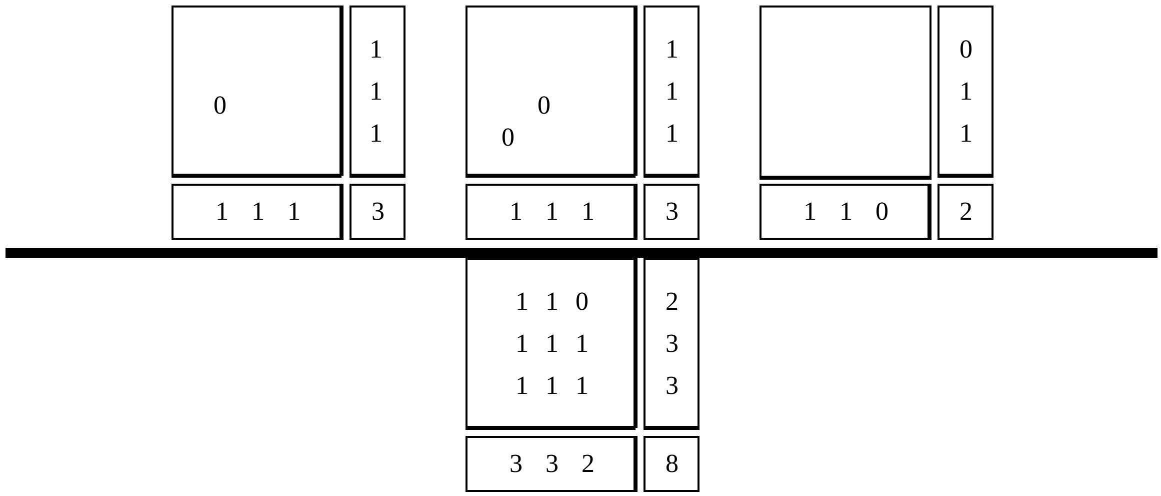
**Example 6: 3x3x3 Table With a Unique Cover But No Circuit**

| | | |
|---|---|---|
| % | & | 0 |
| 0 | & | % |
| & | **%** | & |

| | | |
|---|---|---|
| & | % | 0 |
| % | 0 | & |
| 0 | & | % |

| | | |
|---|---|---|
| 0 | 0 | 0 |
| & | % | 0 |
| % | & | 0 |

**Example 7a: A Unique Odd Circuit**




**Example 7b: Fractional Optima Under Zero-Restrictions**

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | 1 | | 1 | | 0 | | 0 |
| | 1 | | 1 | | 1 | | 0 |
| | 1 | | 1 | | 1 | | 1 |
| | 1 | | 0 | | 0 | | 1 |
| 1 1 1 1 | 4 | 1 1 1 0 | 3 | 0 0 1 1 | 2 | 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 | 11 |

**Example 7c: Fractional Optimum In the Absence of Zero-Restrictions**