Available at www.MATHEMATICSWEB.ORG
POWERED BY SCIENCE DIRECT

Journal of Statistical Planning and Inference 117 (2003) 251–273

journal of statistical planning and inference

www.elsevier.com/locate/jspi

On properties of multi-dimensional statistical tables

Lawrence H. Cox[1]

National Center for Health Statistics, Centers for Disease Control and Prevention, 6525 Belcrest Road Room 915, 20782 Hyattsville, MD, USA

Received 20 June 2000; received in revised form 10 July 2002; accepted 31 July 2002

Abstract

Statistical data often can be conveniently organized in tabular form for display and analysis. Counts are nonnegative integers, and often magnitude data 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 desirable mathematical properties, and consequently their associated statistical methods, are not extendible in all cases 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 characterization of n-dimensional tables with prescribed $(n-1)$-dimensional marginals for which continuous bounds on integer-valued entries exist and are integer (*integrality*).

MSC: 62H17; 62J12

Keywords: Contingency table; Rounding; Log-linear models; Imputation; Iterative proportional fitting; Stratified sampling; Statistical data base; Mathematical network

1. Introduction

Two-dimensional tables are a staple of statistical science. Typically, table entries are nonnegative and often integer-valued, e.g., contingency tables. Important statistical

[1] New Address: Dr. Lawrence H. Cox, Associate Director, National Center for Health Statistics, Office of Research and Development, 3311 Toledo Road, Room 3211, Hyattsville, MD 20782 USA.

E-mail address: lgc9@cdc.gov (L.H. Cox).

0378-3758/02/$ - see front matter
PII: S0378-3758(02)00392-0

methods depend upon underlying mathematical properties of two-dimensional tables with prescribed row and column totals. Modern methods for retrieving, computing and analyzing statistical data enable and encourage extending these methods to three- and higher-dimensional tables. The purpose of this paper is to investigate if and under what conditions extension of these properties to higher dimensions is possible. In higher dimensions, we must specify which marginals will be fixed. We select the generalized row and column totals—the $(n-1)$-dimensional marginals—defined in the next section.

Section 2 describes several mathematical properties of two-dimensional tables, each associated with important methods from statistical science. For each property, a counterexample to extending it and its associated statistical methods to three-, and consequently to n dimensions, $n \geqslant 3$, subject to fixed $(n-1)$-dimensional marginal totals, is provided. The counterexamples presented in Section 2 were developed for their simplicity, and therefore may appear contrived. The question arises as to whether the counterexamples illustrate a problem of importance or are merely mathematical artifacts. This question is answered in Section 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. We examine the importance of being able to detect whether prespecified marginals assure the existence of a feasible n-dimensional table, viz., whether sensible conditional distributions assure the existence of a corresponding joint distribution.

The existence of a compatible joint distribution is important in a variety of applications. An important application is public access statistical data base query systems. The data base accepts data of varying quality possibly from a variety of sources. The extent to which these data, representing separate marginal distributions, can be combined to form a joint distribution is central to the usefulness and quality of the data base. Conversely, marginal distributions that appear to define a joint distribution, but in fact do not, should not be confused with those that do and need to be identified in advance. The question arises as to whether it is possible to identify such failures in advance. This question is answered in Section 3. We provide necessary and sufficient conditions for the existence of a joint distribution compatible with marginal distributions based on fixed $(n-1)$-dimensional marginal totals and an empirical test for feasibility, based on iterative proportional fitting.

Section 4 presents selected results from integer linear programming theory used in Section 5. A final important question is whether there are classes of higher-dimensional tables for which mathematical properties underlying important statistical methods do extend from two to higher dimensions. For extensions of feasibility, this question was answered in Section 3. For extensions of integrality, this question is answered in Section 5. We establish necessary and sufficient conditions for assuring the existence of integer-valued solutions in higher dimensions, subject to fixed $(n-1)$-dimensional marginals, enabling explicit extension of important statistical methods including stratified sampling and controlled rounding. Section 6 provides discussion.

2. Important properties of two-dimensional statistical tables that fail in higher dimensions

2.1. Terminology and notation

A *two-dimensional statistical table*, denoted $T(d_1 d_2)$, $d_1, d_2 \geqslant 2$, comprises d_2 column equations, d_1 row equations and a grand total equation, denoted, respectively

$$\sum_{i=1}^{d_1} a_{ij} = a_{.j}, \quad \sum_{j=1}^{d_2} a_{ij} = a_{i.}, \quad \sum_{i=1}^{d_1} a_{i.} = \sum_{j=1}^{d_2} a_{.j} = a_{..}. \tag{1}$$

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

An n-dimensional table is denoted $T(d_1 d_2 \ldots d_n)$, $d_k \geqslant 2$. The $(n-1)$-*dimensional marginal totals* are defined by holding $n-1$ indices fixed and summing over the remaining index. The set of marginal totals formed by summation over the qth variable is $A_q = \{a_{i_1,..+,..i_n}\}$. A *feasible n-dimensional table* admits at least one nonnegative solution that sums correctly to each of the prescribed $(n-1)$-dimensional marginal totals. The number of $(n-1)$-dimensional marginals is large: $\sum_{k=1}^{n} \prod_{q=k} d_q$ (see Section 2.2).

The remainder subsections of this section are concerned with the failure of several properties of two-dimensional tables to generalize completely to three-dimensional tables, and consequently to higher-dimensional tables. Failures are demonstrated in the form of counterexamples.

The figures containing the counterexamples are to be read as follows. Each figure presents a *potential* three-dimensional table of nonnegative entries subject to the two-dimensional (viz., $(n-1)$-dimensional) marginal totals shown. We say "potential" because failure of the property may prevent the existence of an actual table. The (potential) internal entries of the table are indexed (i,j,k), $i = 1,\ldots,d_1$, $j = 1,\ldots,d_2$, $k = 1,\ldots,d_3$. The potential three-dimensional internal entries are to be arranged in the blank boxes. The two-dimensional marginal totals defined by summation along the i- and j-directions for fixed k appear, respectively, below and to the right of the corresponding box. The two-dimensional marginal totals in the k-direction appear in the box below the dark line. The potential three-dimensional 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. For clarity, the one-dimensional (viz., $(n-2)$-dimensional) marginal totals are presented at the lower right beside each box above the line and below and to the right of the box below the line. The $(n-3)$-dimensional marginal total (the grand total) is located at the lower right next to the box below the line. With this notation and conventions, we are prepared to examine in higher dimensions some important properties enjoyed by two-dimensional tables.

2.2. Feasibility

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 always exists given 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 feasibility (Property 1).

Example 1. Consistent $(n-1)$-dimensional marginal totals do not assure existence of a feasible n-dimensional table.

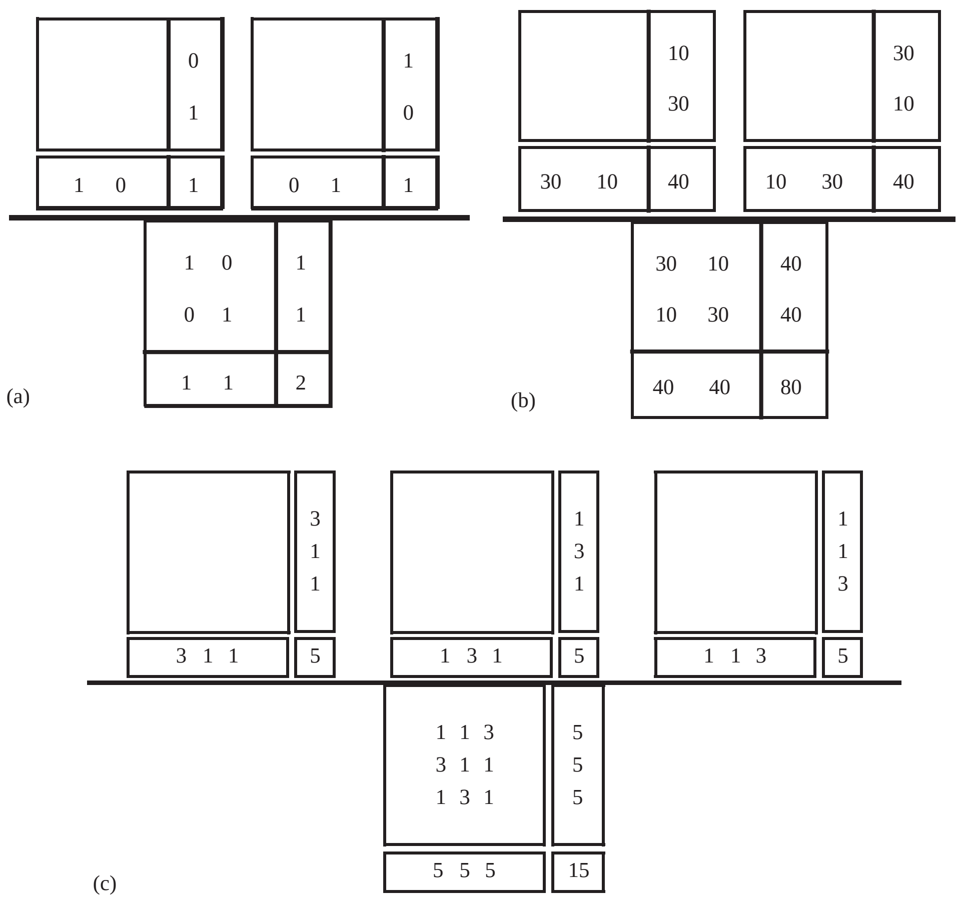
In n dimensions, if the $(n-1)$-dimensional marginal totals admit a feasible n-dimensional table, then any pair of them agreeing on $n-2$ indices must comprise the column and row totals of a feasible two-dimensional subtable of the original n-dimensional table. These subtables are identified as follows. Each pair $a_I \in A_k$, $a_J \in A_l$, $k \neq l$ of $(n-1)$-dimensional marginal totals that agree on all indices other than k and l form a unique two-dimensional subtable of the n-dimensional table. These subtables may be organized into $\binom{n}{2}$ sets, corresponding to pairs of indices (k, l). The set corresponding to pair (k, l) contains precisely $\prod_{q \neq k,l} d_q$ two-dimensional subtables of the original table, resulting in $\sum_{(k,l)} \prod_{q \neq k,l} d_q$ two-dimensional subtables.

If the full table is feasible, then the $(n-1)$-dimensional marginal totals must obey the *consistency condition*: any pair of sets of $(n-1)$-dimensional marginals defining a two-dimensional table must add to a mutual value, namely, the grand total of the two-dimensional table they define. If this condition holds for two-dimensional subtable, a *consistent set of* $(n-1)$-*dimensional marginal totals* is said to exist. A consistent set of $(n-1)$-dimensional marginal totals is a necessary condition for the existence of a feasible n-dimensional table. It is not, however, sufficient, as illustrated by Examples 1a–c. Consequently, in $n \geqslant 3$ dimensions, a consistent set of $(n-1)$-dimensional marginal distributions does not guarantee existence of an underlying n-dimensional joint distribution. Any method that seeks to impute internal entries from marginal entries under these conditions will fail. Any method that seeks to establish lower and upper bounds on internal entries under these conditions will also fail, and may mislead the analyst by producing seeming meaningful results (see Cox, 2002).

Example 1a is for clarity: it is easy to see that, subject to the prescribed two-dimensional marginals, a feasible table cannot exist. Examples 1b and c illustrate that infeasibility is not always easy to spot. Indeed: Example 1b equals 10 times the sum of a feasible table and three times Example 1a. This sort of construction is useful and is examined in Section 3.

Consider Examples 1c and d: the former is infeasible and the latter is feasible (indeed, has a unique feasible solution). It would be difficult to impossible to draw these distinctions only on inspection. On further examination, Example 1b illustrates a second point: the $n=3$ subsequences of iterative proportional fitting estimates for the potential $(1,1,1)$ entry converge to three different values: 10, 13.0278 and 16.9722. This also is important and examined in Section 3.

An argument that can be raised is that infeasible tables do not arise in statistics and therefore are of no practical interest. Clearly, given feasible internal entries in n-dimensions, a complete, feasible n-dimensional table and its marginals can be defined by appropriate addition of the feasible entries. However, complete feasible tables are not always available, for several reasons. Potential marginal totals are not always



Example 1. (a–c) Consistent but infeasible 3-D table; (d–e) Feasible 3-D table.

(d)

			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

(e)

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 1. (*continued*).

fixed, derived 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, the analyst might attempt to fit internal entries to them. By Property 1, this can be done in two-dimensions without violating feasibility. However, owing to Examples 1a–c, this can fail in three or higher dimensions. Furthermore, the failure can go undetected, with the result that an investigator can be analyzing and drawing conclusions from a table that in fact does not exist.

There are other situations where consistent marginal totals are 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 (Cox, 1987), such methods can fail in higher dimensions. Consider Examples 1c–e. If Example 1e demonstrates that the original marginals (of Example 1e) are feasible. Assume that the analyst is given these marginals, and that, e.g., they represent expected values of sample counts. The analyst wants to produce a complete, rounded three-dimensional table to enable actual sampling. Examples 1c and d each represent acceptable roundings of the original marginal totals. The analyst next wants to round internal entries consistent with the rounded marginals. Starting with Example 1d, a feasible rounded table can be produced, and sampling performed, whereas, starting with Example 1c, infeasibility results.

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, the problem might for practical purposes be ignored. However, as demonstrated in Section 3, infeasibility is anything but rare in $n \geqslant 3$ dimensions.

Infeasibility also affects statistical methods: several statistical procedures proven in two dimensions are *insensitive to infeasibility* in $n \geqslant 3$ dimensions, viz., they will produce a final result regardless of whether or not a feasible table exists. Statistical methods that are *insensitive to infeasibility* can have disastrous effects on large-scale statistical data processing operations. Such operations are commonplace at national statistical offices and large survey organizations. Cox (1999,2002) examine several insensitive statistical methods proposed in the statistical disclosure limitation literature.

2.3. Bounds on feasible values for internal entries

The *Fréchet upper bound* for an internal entry of a two-dimensional table is the minimum of its two one-dimensional marginal totals. 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. An upper or lower bound on an internal entry is *exact* if it is achieved in at least one feasible table.

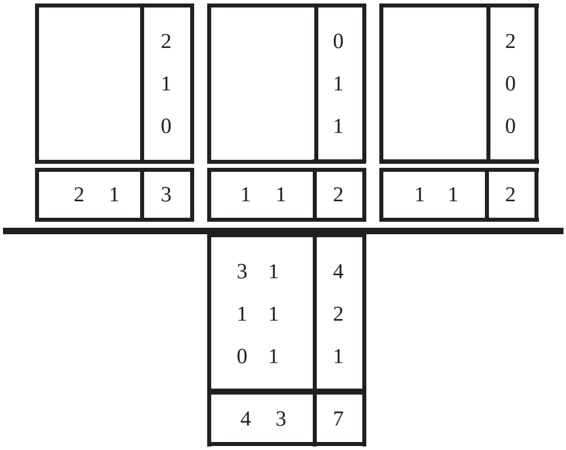
Property 2. Fréchet upper and lower bounds in two-dimensional tables are exact.

The exactness of the Fréchet upper bound is demonstrated by 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 has been reduced to zero. Exactness of the Fréchet lower bound follows by setting all values under the summation to zero.

In n-dimensions, the *n-dimensional Fréchet upper bound* of an internal entry is the minimum of the $(n-1)$-dimensional marginal totals to which the entry contributes. There are n of these $(n-1)$-dimensional marginals. For purposes here, the *n-dimensional Fréchet lower bound* equals the maximum of zero and the maximum of the entry's two-dimensional Fréchet lower bounds. There are $n(n-1)/2$ of the latter.

Example 2. Fréchet upper and lower bounds are not exact in $n \geqslant 3$ dimensions.

In Example 1d, the $(3,3,1)$ entry (lower right corner of left-most box) has Fréchet upper bound equal to one. However, this entry achieves a unique value of zero,



Example 2. Feasible (unique) 3-D table with inexact Fréchet lower bound.

as there is precisely one feasible table corresponding to these marginal totals. In Example 2, the Fréchet lower bound for the $(1,1,1)$ entry equals one, but this entry has a unique value of two.

Bounds and feasibility are related concepts. Consider Example 1a. Although the $(n-1)$-dimensional marginal totals are consistent, the marginal totals are inconsistent. For entry $(2,1,1)$, the Fréchet upper bound equals 0, while its Fréchet lower bound equals 1, a contradiction: a feasible table cannot exist. The converse is not true: even if, for each internal entry, the Fréchet upper bound is equal to or greater than the Fréchet lower bound, a feasible table may not exist. This is illustrated by Example 1c.

Property 2 is important in statistical disclosure limitation (US 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. Imputation and adjustment of entries also rely on Property 2, 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.

2.4. Data perturbation, swapping and balanced adjustment in tables

Given a two-dimensional table and a selected entry, it is possible to construct one or more *covers* containing the selected entry, i.e., a set of entries such that, for each entry in the set, at least one additional entry from its row and likewise one from its column

0	+	0	−
+	−	0	0
−	0	+	0
0	0	−	+

Fig. 1. An alternating cycle.

are also in the set. Minimal covers form *alternating cycles*, i.e., two entries from each covered row and column are present and each can be assigned a different sign ($\pm$). (Entries assigned the $-$ sign are usually required to be nonzero.) Alternating cycles are necessarily of even length. Alternating cycles are always available in two-dimensions. Alternating cycles underlie methods such as *balanced adjustment* (a.k.a. *controlled perturbation*) of internal entries subject to preserving additivity to marginal totals, for statistical purposes including rounding, data perturbation or data swapping (Cox et al., 1986). The use of alternating cycles is as follows. If all entries along the cycle are positive, then it is possible to move a small positive quantity among these entries in a manner that preserves both nonnegativity of entries and additivity to totals. In particular, if entries are integer, then at least one unit can be moved enabling, e.g., *swapping* of characteristics between sample cases. Doing so repeatedly allows the analyst, e.g., to move from the original table which contains disclosures to an alternative table that does not. 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 $n \geqslant 3$ 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 (Cox and Ernst, 1982). 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 (Cox, 1987), and to methods for imputing/adjusting/perturbing table entries. See Cox et al. (1986) for details.

Feasible zero-restrictions do not pose obstacles in two-dimensional tables. For example, consider the alternating cycle in Fig. 1 and assume that the marginal entries are multiples of some integer rounding base. Then, it is possible to adjust the ($\pm$) entries so that the resulting table is both additive and rounded to the base.

0	*	*
*	**0**	*
*	*	**0**

*	*	**0**
*	**0**	*
0	*	*

*	**0**	*
0	**0**	**0**
*	**0**	*

(a)

+	−	**0**
0	−	+
−	+	−

−	+	**0**
+	**0**	−
0	−	+

0	**0**	**0**
−	+	**0**
+	−	**0**

(b)

Example 3. (a) $3 \times 3 \times 3$ table with a unique cover but no circuit; (b) A unique odd circuit.

Property 3. Covers and alternating cycles exist for nonzero entries in two-dimensional tables.

See Cox (1987) for details on alternating cycles and preserving zero-restrictions in two dimensions.

Example 3. Circuits can fail to exist or be of odd length in $n \geqslant 3$ dimensions.

Example 3a illustrates a $3 \times 3 \times 3$ table with 11 of its 27 entries zero-restricted. The remaining 16 entries, denoted "$*$", are positive. This pattern of * entries is a cover for these 16 values, but this cover is not a circuit because 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: controlled perturbation, such as for disclosure limitation (Cox et al., 1986) is impossible. Similarly for controlled rounding and controlled stratified sampling (Cox and Ernst, 1982).

Example 3b exhibits a $3 \times 3 \times 3$ 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 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.

Examples 3a and b demonstrate that conditions for controlled perturbation and related statistical methods are quite different in higher dimensions from those in two dimensions. In particular, *moves* between contingency tables with fixed marginals can be complex in higher dimensions (Diaconis and Sturmfels, 1998).

2.5. Integrality

Property 4. A consistent pair of integer one-dimensional marginal totals assure the existence of a feasible integer two-dimensional table. And, often many integer feasible solutions exist.

The first part of Property 4 is assured in two dimensions 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. Here, noninteger maximum likelihood estimates for column-row interactions that are additive to column and row totals must be replaced by integer entries or sample sizes that are additive to integer row and column totals. This is accomplished by base 1 controlled rounding of the table of noninteger expected values (Causey et al., 1985). Ernst (1989) showed that in general controlled rounding is not extendible to higher dimensions. However, in those cases where extension is possible (see Section 5), multi-way controlled stratified sampling (a.k.a. *controlled selection*) is possible.

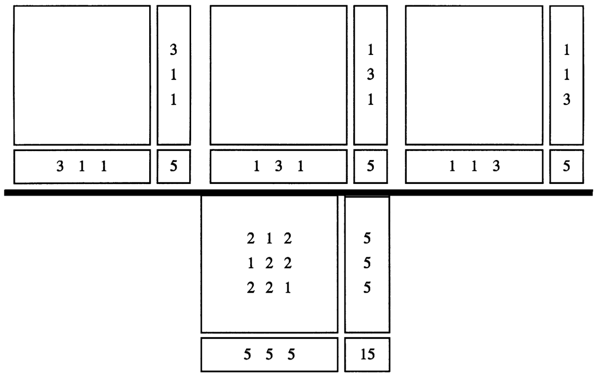
As regards number of solutions, in a two-dimensional table, even if integer one-dimensional marginal totals are small, often there exist many integer feasible tables. For example, a $b \times b$ table with all one-dimensional marginal totals equal to one contains $b!$ feasible integer tables. For disclosure limitation purposes, the existence of relatively few alternative integer feasible solutions could constitute disclosure.

Example 4. The number of integer solutions in higher dimensions can be small, even zero. (Example 4)

Even though the continuous feasible solutions to Example 4 define a constrained subset of three-dimensional space, only four integer solutions exist. For a confidentiality standpoint, a contingency table with so few realizations likely would be considered risky. Vlach (1986) presents an example with integer marginals and non-integer solutions but no integer solutions, viz., *integer infeasibility*.

2.6. Integer bounds

Property 5. In a two-dimensional table, consistent integer one-dimensional (viz., $(n-1)$-dimensional) marginal totals assure that exact lower and upper bounds on internal entries are integer.



Example 4. 3-D table with 3 constrained df but only 4 integer solutions.

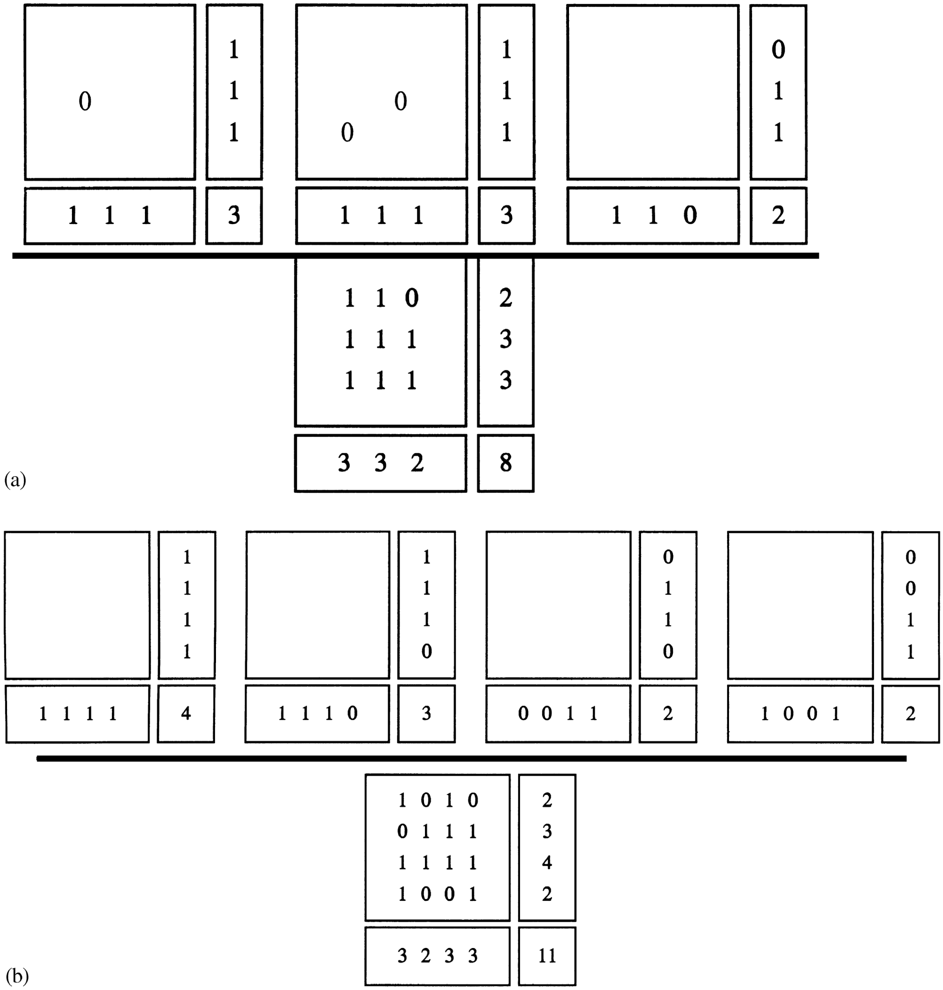
Property 5 is assured by Properties 1 and 2.

Example 5. Exact bounds in higher dimensions can be noninteger.

In the absence of Property 5, an exact continuous bound on an integer entry is not necessarily numerically adjacent to the exact integer bound (Cox, 2002), and consequently is of questionable value in the determination of exact integer bounds. Furthermore, unsupported by other methods, continuous bounding methods, e.g., linear programming, can be insensitive to whether or not integer solutions even exist.

Example 5a is an elaboration of Example 3b, 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 3b (viz., excluding the $(3,2,1)$ entry) is $\frac{1}{2}$. Similarly, the minimum value of each of the eight entries marked "−" is $\frac{1}{2}$. 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 $\frac{1}{2}$. 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 $\frac{1}{2}$. 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.

Linear programming reveals that the maximum continuous value of the $(3,3,1)$ entry in Example 5b is $\frac{1}{2}$. This is the only fractional optimum, and is achieved without explicit zero-restrictions. Examples 5a and b illustrate that, in $n \geqslant 3$ 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 (US Department of Commerce, 1994, Recommendation 11). In addition, heuristic bounding methods often are based on iterative improvement of inexact integer bounds. Many



Example 5. (a) Fractional exact bound under zero-restrictions; (b) Fractional exact bound in the absence of zero-restrictions.

of these methods are variants on the *shuttle algorithm* (Buzzigoli and Guisti, 1999), examined critically by Cox (2002). The shuttle is based on refining at each iteration upper (respectively, lower) bounds by exploiting lower (respectively, upper) bounds obtained at the previous iteration. Each inexact iterative bound is necessarily outside the feasible region. When the continuous exact bound is not equal to the integer exact bound, the integer exact bound is necessarily an interior point of the feasible region. Examination of the iterative computations reveals that these methods cannot cross the boundary of the feasible region, and therefore are incapable of reaching the exact integer bound.

3. Closer examination of these problems in higher dimensions

3.1. Counterexamples are abundant

Examples 1 enable creation of consistent but infeasible three-dimensional tables $T(abc)$ of arbitrary size $d_1, d_2, d_3 \geqslant 3$, by combining these examples with blocks of zeroes. For example, to create an infeasible $5 \times 5 \times 5$ table, place Example 1a in the upper-front-left of the $5 \times 5 \times 5$ space, and place Example 1c 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 $n \geqslant 3$ and nearly arbitrary size (see Section 5 for exceptions). Similar constructions are possible for the other examples.

$T(d_1, \ldots, d_n; A)$ denotes an n-dimensional statistical table of *size* $(d_1, \ldots, d_n)$. T comprises $\prod_{j=1}^{n} d_j$ nonnegative internal entries, $d_j \geqslant 2$, 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 $T(2^n)$ or $T(bc)$ is used when the meaning is clear, and $mT(d_1, \ldots, d_n; A) = T(d_1, \ldots, d_n\colon mA)$. 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 $M = (M_B, M_N)$ 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 $T(d_1, \ldots, d_n; A)$ exist in dimension $n \geqslant 3$. The first theorem demonstrates that 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 for $n \geqslant 3$. Let $T_\mathrm{f} = T(d_1, \ldots, d_n; A_\mathrm{f})$ denote a feasible table and $T_\mathrm{i} = T(d_1, \ldots, d_n; A_\mathrm{i})$ denote an infeasible table. Then there exists an integer p such that $T_\mathrm{f} + mT_\mathrm{i}$ is infeasible for all $m \geqslant p$.*

Proof. That T_i is infeasible is equivalent to saying that every basic solution of $MX = A_\mathrm{i}$ contains at least one negative entry. As the set of basic solutions of $MX = A_\mathrm{f}$ is bounded, then there exists an integer p such that, whenever $m \geqslant p$, every basic solution of $MX = A_\mathrm{f} + mA_\mathrm{i}$ contains at least one negative entry. □

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.

Infeasibility in statistical environments such as at national statistical offices can, if not detected, bring a large data processing activity to its knees. Feasibility can be tested using linear programming, but without insight into what conditions on the $(n-1)$-dimensional marginals ensure feasibility. 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.

3.2. A test for feasibility

Despite the fact that Property 1 fails in $n \geqslant 3$ 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 $n \geqslant 3$ dimensions, as follows (for complete details, see: Bishop et al., 1975, Sections 3.5.1-2).

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 an unsaturated log-linear model, all sufficient statistics can be derived from the $(n-1)$-dimensional marginal totals and the iteration can begin with *starting values* containing the value 1 in each internal cell. The log-linear model corresponding to the $(n-1)$-dimensional marginal totals is the *no n-factor effect* model, for which the usual n sets of $(n-1)$-dimensional marginal totals $A_q = \{a_{i_1,..+,..i_n}\}$ are sufficient statistics (summation over the qth index only). Starting values are $a^{(0)}_{i_1,\ldots,i_q,\ldots,i_n} = 1$. Here, the superscript denotes the iteration number, and the iterations are organized in n cycles corresponding to fitting to each of the n sets of $(n-1)$-dimensional marginals in turn. At the 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 $q \pmod n = 1,\ldots,n-1,\ 0$ (middle loop), and each cycle $m = 1,\ldots,$ (outer loop), as follows:

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

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 equal or are derivable from the $(n-1)$-dimensional marginals.

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:

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

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 therefore holds in the limit $m \to \infty$. Consequently, if the iterative proportional fitting algorithm converges for all internal entries, then there must exist M^* such that given $\varepsilon > 0$, whenever $M \geqslant M^*$, we have: $|\sum_{i_q} a^{((M-1)n+q)}_{i_1,\ldots,i_q,\ldots,i_n} - a_{i_1,\ldots,+,\ldots,i_n}| < \varepsilon$ for all $q = 1, \ldots, n$. But this is precisely the statement that the convergence limits define a feasible table. □

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, when divergence manifests itself as n subsequences converging to two or more distinct limits (e.g., Example 1b), infeasibility can be detected with relative ease. Theorem 3.2 holds for all minimal sufficient statistics.

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 $a^{(0)}$ 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 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. All entries of a totally unimodular matrix must equal $-1, 0$, or $+1$, as does the coefficient matrix of the system of *aggregation equations* defined by the $(n-1)$-dimensional marginals of an n-dimensional table. Conditions under which this coefficient matrix is totally unimodular are explored in the next section.

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 Fig. 2 (for $b=c=3$). Note that arc flows are designated by variables x_{ij} (these appear as x_{ij} in Fig. 2), and that $x_{ij} = a_{ij}$ is one, but in general not the only, *feasible solution* to the network.

5. Integer optima in *n*-dimensions

Properties 4 and 5 depend on the total unimodularity of two-dimensional tables. The efficiency of associated computations, viz., for computing exact bounds (Gusfield, 1988), owes to the network structure (Kennington and Helgason, 1980). The counterexamples presented in Section 2 demonstrate that these properties can be lost in $n \geqslant 3$ dimensions. In this section, we investigate circumstances under which they will be preserved.

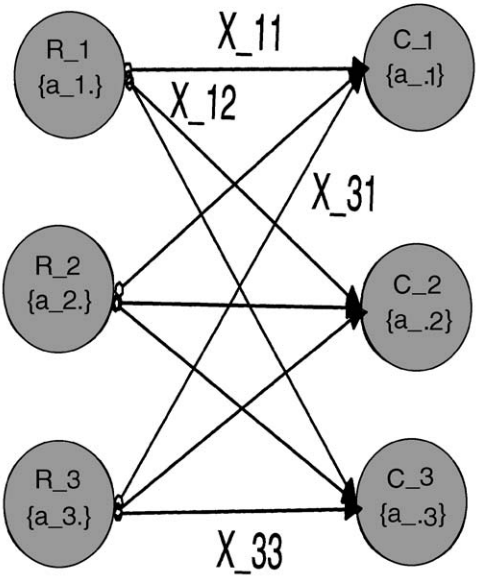


Fig. 2. Network for a 3×3 two-dimensional table.

$T \times T$ denotes a generic $(n+1)$-dimensional statistical table of the form $T \times T(d_1, \ldots, d_n, 2)$, viz., comprising two copies of a generic n-dimensional table. Feasible $(n+1)$-dimensional tables are nonnegative solutions of

$$\begin{pmatrix} M_1 & 0 \\ 0 & M_2 \\ I & I \end{pmatrix} (X_1, X_2) = \begin{pmatrix} A_1 \\ A_2 \\ A_3 \end{pmatrix}, \quad \text{for } \{0,1\}\text{-matrices } M_1,\ M_2$$

and integer matrices $A_1,\ A_2,\ A_3$.

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 vectors A of appropriate size, the solutions of $MX \leqslant A$ are integer (Nemhauser and Wolsey, 1988, Chapter III.1). Totally integer is a very desirable, but restricted, property.

Example 6. Let M denote the coefficient matrix corresponding to the table in Example 5a. M is obtained by deleting all columns corresponding to entries marked "0" from the coefficient matrix M_{333} of a full $3 \times 3 \times 3$ table. A feasible solution x^* for the table is to set the $(3,2,1)$ entry to 1 and all other nonzero entries to $\frac{1}{2}$. This is a noninteger solution. 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 $\frac{1}{2}$, and thus M is not totally unimodular. As x^* maximizes each + entry, it is an extreme point of $MX = A$. The corresponding table T (Example 5a) is not totally integer.

Theorem 5.1. *T is totally integer if and only if $T \times T$ is totally integer.*

Proof. (If) Trivial.

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

$$\begin{pmatrix} M_1 & 0 \\ 0 & M_2 \\ I & I \end{pmatrix} (X_1, X_2) = \begin{pmatrix} A_1 \\ A_2 \\ A_3 \end{pmatrix}, \quad \text{for } M_1,\ M_2 \in \{0,1\},\ \text{integer } A_1, A_2, A_3 \geqslant 0,$$

$$\text{and } X_1, X_2 \geqslant 0. \tag{4}$$

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

$$c(x) = \sum_{x^*_{1j} \neq 0} x_{1j} + \sum_{x^*_{2j} \neq 0} x_{2j} = \sum_{x^*_{1j} \neq 0} x_{1j} + \sum_{x^*_{1j} \neq a_{3j}} (a_{3j} - x_{1j}). \tag{5}$$

The latter optimization is over only the X_1-constraints corresponding to one copy T_1 of T (viz., over $M_1 X_1 = A_1$) subject to additional integer capacity constraints given by: $X_1 \leqslant A_3$. As capacity constraints do not affect the total unimodularity of M_1, then x^*_1 is an integer point of the polyhedron defined by: $M_1 X_1 = A_1$, $X_1 \leqslant A_3$. As A_3 is integer, then $x^*_2 = A_3 - x^*_1$ is integer, and hence x^* is integer. Therefore, $T \times T$ is totally integer. □

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

Theorem 5.2. *T is network if and only if $T \times T$ is network.*

Proof. (If) Trivial.

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

A network representation $N(T \times T)$ of $T \times T$ is obtained by applying the method illustrated in Fig. 3 successively for each n-dimensional index j, as follows. Prior to introducing the $(n+1)$th set of $(n-1)$-dimensional totals $X_1 + X_2 = A_3$, two networks $N(T_1)$ and $N(T_2)$ identical in node-arc structure exist, viz., one for each copy of T. Portions of these networks containing the arc corresponding the variable x_j is illustrated in Fig. 3a. The two networks differ only in the individual node requirements, which are determined, respectively, by the $(n-1)$-dimensional totals A_1 and A_2. Fig. 3b illustrates the state of network $N(T \times T)$ at arc j just after introducing the $(n+1)$th set of $(n-1)$-dimensional totals. Note that there are no longer j-arcs to nodes O_{2j}

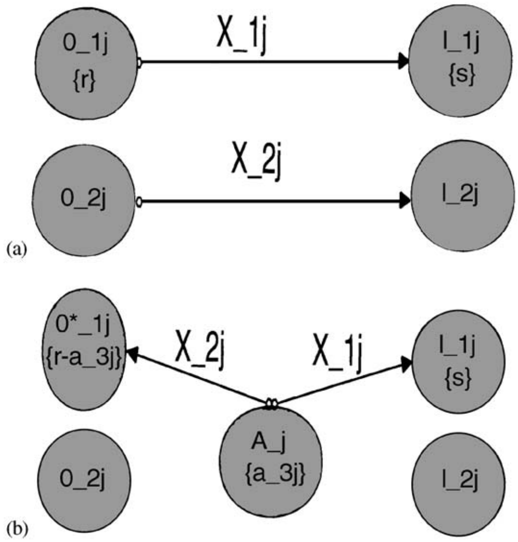


Fig. 3. (a) Original networks and (b) Creating $N(T \times T)$.

and I_{2j}: once the alterations of Fig. 3b are made for all j-indexes, these nodes can be removed from the network because they represent linearly dependent equations. □

Corollary 5.3. *$T(2^n)$, $T(2^n b)$ and $T(2^n bc), b, c \geqslant 3, n \geqslant 0$, are network.*

Theorem 5.3. *T is totally integer, and network, if and only if $T = T(2^n)$, $T(2^n b)$ or $T(2^n bc); b, c \geqslant 3, n \geqslant 0$.*

Proof. (If) Corollary 5.3 provides the proof.

(Only if) Example 5a provides the proof: T of any other size must contain $T(3^3)$, which by Example 5a fails to be totally integer, and therefore T must fail as well. □

The class of higher-dimensional statistical tables covered by the preceding theorems is an important one. It is not uncommon for categorical variables to be dichotomous, or dichotimized for analysis.

6. Discussion

We have shown in Section 2 that important mathematical properties that hold for two-dimensional tables do not hold in all cases in higher dimensions, and consequently that statistical methods and algorithms for higher-dimensional problems based on assumptions that these methods are extendible to all higher-dimensional cases are faulty and can 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.

Three questions arose: (1) in general, how pervasive is the failure of certain statistical methods to extend from two to higher dimensional tables, (2) is it possible to detect failure in advance, and, (3) is it possible to identify one or more general classes of higher-dimensional tables for which suitable extension of these methods is guaranteed? The first two questions were answered in Section 3. We showed that infeasibility and failure of integrality, not present in two dimensions, are ubiquitous and numerous in higher dimensions and therefore not mathematical anomalies that can be ignored. Because data are often subjected to statistical adjustment, imputation, rounding, etc., independently and due to the need to 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 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. The third question was answered in Section 5. We characterized completely those higher-dimensional tables for which integrality is assured for prescribed $(n-1)$-dimensional marginals.

Extension of properties and algorithms for two-dimensional tables to more complex structures evidently needs to be attempted with caution. 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 advanced mathematically and extremely demanding computationally, viz., even for $3 \times 3 \times 3$ tables.

This paper is concerned with mathematical properties of higher-dimensional statistical tables and their effects on extending important statistical methods from two-dimensional to higher-dimensional tables. Strong interest is emerging in the combinatorial mathematics community in more general properties of n-dimensional tables with $(n-1)$-dimensional marginals fixed, e.g., enumeration and integer programming for contingency tables (Sturmfels, 1996, Chapter 5), n-dimensional magic squares and hypercubes (Beck et al., 2002), volumes of polytopes defined by doubly stochastic matrices (Beck and Pixton, 2002), and computational complexity (de Loera and Onn, 2002). Combined, these results prompt a general question: why are two-dimensional tables fundamentally different from higher-dimensional tables? In each instance, this question is complex mathematically. However, in broad generality some observations can be made. Moving beyond two dimensions increases both the number variables and the number of marginal constraints and their interactions. Geometrically, this results in a mathematically more diverse set of polytopes that, on one extreme, can limit the size of the feasible region or eliminate it all together (viz., infeasibility) and at the other extreme can result in a richer set of feasible and extremal solutions including fractional extremal points (viz., failure of integrality). The latter can also be attributed algebraically to the failure of total unimodularity: the existence of one or more integer determinants of absolute value greater than one which, as the denominator in Cramer's rule, can cause fractional extremal solutions.

This paper is intended to contribute to improved understanding of the structure of multi-dimensional statistical tables, towards improved understanding of how statistical methods valuable for two-dimensional tables might or might not be extended to higher dimensions. Theorem 5.3, which distinguishes between higher-dimensional structures that behave like two-dimensional tables and those that do not, addresses this issue and provides useful guidance. We have focused on the natural extension of two-dimensional tables with prescribed column and row totals, namely, n-dimensional tables with prescribed $(n-1)$-dimensional marginals. Other situations are likely to produce differing results and are worth examination.

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 extension of these inquiries are linked two- and higher-dimensional tables, and more general structures. This will be challenging, viz., this 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 opinions expressed herein are solely those of the author and should not be interpreted as representing the policies or practices of the National Center for Health Statistics, Centers for Disease Control and Prevention, or any other organization. Comments from Shelley Zacks and Matthias Beck are gratefully acknowledged.

References

Beck, M., Pixton, D., 2002. The Ehrhart polynomial of the Birkoff polytope. Manuscript dated February 25, 2002, 14pp. Available: http://front.math.ucdavis.edu/math.CO/0202267.

Beck, M., Cohen, M., Cuomo, J., Gribelyuk, P., 2002. The number of "magic" squares and hypercubes. Manuscript dated May 27, 2002. 11pp. Available: http://front.math.ucdavis.edu/math.CO/0201013.

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

Buzzigoli, L., Guisti, A., 1999. An algorithm to calculate the lower and upper bounds of elements of an array given its marginals. Statistical Data Protection: Proceedings of the Conference, EUROSTAT, Luxembourg, pp. 131–147.

Causey, B.D., Cox, L.H., Ernst, L.R., 1985. Applications of transportation theory to statistical problems. J. Amer. Statist. Assoc. 80, 903–909.

Cox, L.H., 1987. A constructive procedure for unbiased controlled rounding. J. Amer. Statist. Assoc. 82, 520–524.

Cox, L.H., 1999. Some remarks on research directions in statistical data protection. Statistical Data Protection: Proceedings of the Conference, EUROSTAT, Luxembourg, pp. 163–176.

Cox, L.H., 2002. Bounds on entries in 3-dimensional contingency tables. Inference Control in Statistical Databases-From Theory to Practice, Lecture Notes in Computer Science, Vol. 2316. Springer, New York, pp. 21–33.

Cox, L.H., Ernst, L.R., 1982. Controlled rounding. INFOR 20, 423–432.

Cox, L.H., Fagan, J., Greenberg, B., Hemmig, R., 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, pp. 388–393.

de Loera, J., Onn, S., 2002. On the complexity of three-way statistical tables. Manuscript dated, July 30, 2002. Available: http://front.math.ucdavis.edu/math.CO/0207200.

Deming, W.E., Stephan, F.F., 1940. On a least squares adjustment of a sampled frequency table with the expected marginals are known. Ann. Math. Statist. 11, 427–444.

Diaconis, P., Sturmfels, B., 1998. Algebraic algorithms for sampling from conditional distributions. Ann. Statist. 26, 363–397.

Ernst, L.R., 1989. Further applications of linear programming to sampling problems. Technical Report-Census/SRD/RR-89-05. Statistical Research Division, Washington, DC, US Census Bureau, 33pp. Available: http://www.census.gov/srd/papers/pdf/rr89-05.pdf.

Gusfield, D., 1988. A graph theoretic approach to statistical data security. SIAM J. Computing 17, 552–571.

Kennington, J., Helgason, R., 1980. Algorithms for Network Programming. Wiley, New York.

Nemhauser, G., Wolsey, L., 1988. Integer and Combinatorial Optimization. Wiley, New York.

Sturmfels, B., 1996. Gröbner Bases and Convex Polytopes. American Mathematical Society, Providence, RI.

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