[illegible] 2001, 20: 1347-1356

Integer Rounding: Finding Bounds on Cell Entries in N-Way Tables When Some Marginal Totals Are Known

Stephen F. Roehrig
Carnegie Mellon University

September 6, 2001

1 Problem Statement

Let $T = [t_{i_1,\ldots,i_n}], i_1 \in 1,\ldots,I_1,\ldots,i_n \in 1,\ldots,I_n$ be an n-way table of unknown nonnegative integers, and suppose some or all of its $(n-1)$-way and lower-dimensional marginal totals are known. Our problem is to find sharp upper and lower *integer* bounds on the entries of T. This problem is of interest to statistical agencies who publish categorical tables that are aggregations of responses to surveys and censuses. From this perspective, the agency wants to ensure that identification of individual respondents, or inference of sensitive attributes, cannot occur. If the bounds on a cell entry in the n-way table are too narrow, given information contained in published lower-dimensional margins, a disclosure of sensitive information may have occurred.

A natural first choice for finding the desired bounds is linear programming [11]. Unfortunately, if at least three of $I_1,\ldots,I_n$ are greater than three, there is no guarantee that the optima found by LP will be integer [7]. Since there are standard examples in the integer programming literature that show that the optimum for the corresponding integer program may be an arbitrary distance from the LP solution, using LP for disclosure auditing of integer tables has been criticized [1]. We show in this paper that, in certain circumstances, the optimal LP bound for any cell in the n-way table may simply be rounded to obtain the optimal integer bound.

2 The Integer Rounding Property

Any collection of marginal tables $\{T_1,\ldots,T_k\}$ of T imply a set of equality constraints on the underlying cell values $t_{i_1,\ldots,i_n}$. In this section we focus on an n-way table where all of the $(n-1)$-way marginals are known. Let $A^\top$ (where $\top$ indicates transpose) be the matrix of coefficients in the equations that define all $(n-1)$-way marginal totals of T, and let c be the vector of the entries of the marginal tables (defined in the same order as the rows of $A^\top$). The reason for denoting this matrix as a transpose will become clear shortly. We define the values

$$\begin{aligned} p &= \sum_{s=1,\ldots n} I_1 \cdots I_{s-1} \cdot I_{s+1} \cdots I_n \qquad (1) \\ q &= I_1 \cdots I_n \end{aligned}$$

$$
\begin{aligned}
r &= (I - 1)\cdots(I_n - 1) \\
s &= q - r
\end{aligned}
$$

and note that matrix $A^\top$ has p rows and q columns.

Example: A $3 \times 2 \times 2$ table T will have the following constraint matrix. There are $p = 3 * 2 + 3 * 2 + 2 * 2$ rows; the first six rows correspond to the 3×2 marginal $T_{i_1 i_2 +}$, the second group of six to the 3×2 marginal $T_{i_1 + i_3}$, and the last four rows to the 2×2 marginal $T_{+ i_2 i_3}$.

$$
A^\top = \begin{bmatrix}
1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\
1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\
1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1
\end{bmatrix}
$$

In general the rows of such a table will not be linearly independent, since the number of degrees of freedom in choosing values $t_{i_1,\ldots,i_n}$ is not necessarily the same as the number of marginal table cells. Below we will give a simple method for directly writing a minimal set of linearly independent constraint equations.

We consider the dual problems

$$
\max\{cx \mid Ax \le b\} \quad \text{and} \quad \min\{yb \mid y \ge 0; yA = c\},
$$

where the minimization problem is the one we are interested in. The vector c is, again, the vector of the released marginal totals, while vector b is chosen as $\pm e_i$ according to the cell in T whose bound is sought (i.e., $+e_i$ for the lower bound of x_i, $-e_i$ for the upper bound). We call this the *bounds problem.*

Schrijver [8] presents a series of definitions and results that we use here.

Definition 1 *A rational system $Ax \le b$ of linear inequalities is called totally dual integral (TDI) if the minimum in the LP-duality equations*

$$
\max\{cx \mid Ax \le b\} = \min\{yb \mid y \ge 0; yA = c\} \tag{2}
$$

has an integral optimal solution y for each integral vector c for which the minimum is finite.

Thus if one can show that $Ax \le b$ is TDI, then optimal solutions to the dual problem (the RHS of (2)) are guaranteed to be integral, provided c is. For the bounding problem, we know that $Ax \le b$ is *not* TDI, since we have examples of non-integral bounds.

Definition 2 *The cone of a set of vectors* $S = \{a_1, \dots a_t\}$ *is the set*

$$cone\{a_1, \dots a_t\} = \{\lambda_1 a_1 + \cdots + \lambda_t a_t \mid \lambda_1, \dots, \lambda_t \ge 0\}.$$

Definition 3 *A finite set of vectors* $a_1, \dots a_t$ *is called a Hilbert basis if each integral vector in* $\text{cone}\{a_1, \dots a_t\}$ *is a nonnegative integral combination of* $a_1, \dots a_t$.

It is well known that every rational polyhedral cone is generated by an integral Hilbert basis [8].

Definition 4 *A rational system* $Ax \le b$ *of linear inequalities has the integer rounding property if*

$$\min\{yb \mid y \ge 0; yA = c; y \text{ integral}\} = \lceil \min\{yb \mid y \ge 0; yA = c\} \rceil,$$

where $\lceil x \rceil$ *is the smallest integer greater than* x.

We wish to show that $Ax \le b$ has the integer rounding property for the bounds problem, thus assuring that any fractional bound returned by an LP can be used to obtain the correct integer bound. The following result is central.

Theorem 1 *(Giles and Orlin [4]) Let* $Ax \le b$ *be a feasible system of rational inequalities, with* b *integral. Then* $Ax \le b$ *has the integer rounding property if and only if the system* $Ax - b\nu \le 0$ *is TDI, or equivalently, if and only if the rows of the matrix*

$$\begin{bmatrix} A & b \\ 0 & 1 \end{bmatrix}$$

form a Hilbert basis.

Example: We illustrate the use of these definitions and theorem with the $3 \times 2 \times 2$ problem presented earlier. The *row-reduced* version of the matrix for the 2-way marginal sums is the following.

$$A^\top = \begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & -1 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & -1 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1
\end{bmatrix}$$

There are $s = 10$ rows, which is the number of degrees of freedom. To show that the bounds problem has the integer rounding property, we must show that the rows of the matrix in Theorem 1 form a Hilbert basis.

Because the vector b is always a unit vector (in the minimization problem), it is obvious that if the rows of A form a Hilbert basis, so will the rows of the augmented matrix in Theorem 1 (although the bases will be for different cones). So the bounds problem becomes one of showing that the rows of

yes, subpolytope of the 3rd-hypersimplex

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & -1 & -1 & 1 & -1 & 1 & 1 & 0 & 0 & 0 \\ 1 & -1 & -1 & 1 & 0 & 0 & 0 & -1 & 1 & 1 \end{bmatrix}$$

form a Hilbert basis (the rows of A have been slightly rearranged to show some structure). In general, matrix A will have the special structure displayed in the example.

Lemma 1 *After row reduction and deletion of rows of zeros, the matrix* $A^\top$ *has the form* $[I\ B]$ *where* I *is the identity matrix of dimension* $s = I_1 \cdots I_n - (I_1 - 1) \cdots (I_n - 1)$ *and* B *is a* $q - r \times (I_1 - 1) \cdots (I_n - 1)$ *dimensional matrix with entries in the set* $\{-1, 0, 1\}$.

Proof: We explicitly give the rows in the row-reduced matrix $A^\top$. First define the following summation limits.

$$\alpha_t = \begin{cases} i_t & \text{if } \alpha_t \neq 1 \\ i_t \neq 1 & \text{if } \alpha_i = 1 \end{cases}$$

For any complete set of indices $i_1, i_2, \ldots, i_n$, define p to be the number of these indices not equal to 1. Then define

$$z_{i_1,\ldots,i_n} = \begin{cases} x_{i_1,\ldots,i_n} & \text{if } p = n \\ x_{i_1,\ldots,i_n} + (-1)^p \sum_{\alpha_1,\ldots,\alpha_n} x_{i_1,\ldots,i_n} & \text{if } p \neq n \end{cases}$$

Then the left-hand sides of the constraint equations $A^\top x$ consist of those $z_{i_1,\ldots,i_n}$ such that at least one index i_t is a one. There are r such rows, each with a leading coefficient of 1, and any row having -1 coefficients consists of this leading 1 followed by terms all having coefficient -1. Any constraint from the original set (i.e. before row-reduction) can be generated from a suitable linear combination of the z. Finally, rearranging the columns of the set of z produces the form $[I\ B]$ as desired. □

Example: For the $3 \times 2 \times 2$ table, the two levels may be arranged with the cells numbered lexicographically, as follows.

111	121
211	221
311	321

112	122
212	222
312	322

The lemma says that the reduced constraint equation set is the following.

$$
\begin{aligned}
z_{111} &= x_{111} + (-1)^0 \sum_{i_1 \neq 1, i_2 \neq 1, i_3 \neq 1} x_{i_1 i_2 i_3} = x_{111} + (x_{222} + x_{322}) \\
z_{112} &= x_{112} + (-1)^1 \sum_{i_1 \neq 1, i_2 \neq 1, i_3 = 2} x_{i_1 i_2 i_3} = x_{112} - (x_{222} + x_{322}) \\
\ldots &= \ldots \\
z_{321} &= x_{321} + (-1)^2 \sum_{i_1 = 3, i_2 = 2, i_3 \neq 1} x_{i_1 i_2 i_3} = x_{321} - x_{322}
\end{aligned}
$$

which is identical to the matrix A shown above.

To prove the integer rounding property for the bounds problem, it remains to show that the rows of matrix A form a Hilbert basis for the cone generated by the rows. One way to do this is computationally, that is, actually compute the Hilbert basis for cone(A) and show that it is identical to A. A number of algorithms exist for computing Hilbert bases, but all of them are derived in one way or another from the Buchberger algorithm for computing the Gröbner basis for a polynomial ideal [5, 9, 3]. Each algorithm is designed for a different representation of the cone; we chose the version of Henk and Weismentel, which assumes the cone is described, not in terms of its generating vectors (as in Definition 2), but rather by a set of homogeneous inequalities. In particular, their algorithm requires that the cone C be described as

$$C = \{x \mid Mx \leq 0\}.$$

Each inequality is called a *support* of the cone.

In order to use the algorithm of Henk and Weismentel, it is thus necessary to find the matrix M corresponding to the cone of the row vectors $a_i, i = 1, \ldots, q$. A procedure for this is given in Weyl [10], and is the following.

1. For each set of $s-1$ linearly independent vectors $\{a_{i_1}, \ldots, a_{i_{s-1}}\}$, find λ_i such that the hyperplane

$$\lambda_1 a_1 + \ldots + \lambda_{s-1} a_{s-1} = 0$$

   contains them.

2. Test whether one or the other of the halfspaces

$$\pm(\lambda_1 a_1 + \ldots + \lambda_{s-1} a_{s-1}) \leq 0$$

   contains (supports) all the a_i. If so, add the support to the matrix M.

Example: The Weyl procedure was coded and run for our $3 \times 2 \times 2$ example; M is a 28×10 matrix of -1's and 0's. The algorithm of Henk and Weismentel was then run for this matrix, and produced a Hilbert basis precisely equal to A.

The example problem is not realistic, for two reasons. First, it is known [2] that three-dimensional problems where at least one $I_i \leq 3$ are actually network problems, and so are guaranteed to have integer bounds, provided the data (the marginal values used as input) are integral. So such problems vacuously have the integer rounding property. Second, the execution times for Buchberger-like algorithms grow exceedingly

Σ

quickly with increases in the I_i. For example, one fast implementation of Buchberger's algorithm for finding Gröbner bases [6] takes milliseconds to find the basis for a $3 \times 3 \times 3$ problem, 20 minutes for the $4 \times 3 \times 3$ problem, and 3 months for the $5 \times 3 \times 3$ problem.

The smallest bounds problem that is known to have non-integer linear programming bounds is the $3\times3\times3$ problem. Evidence that the step from $3 \times 3 \times 2$ to $3 \times 3 \times 3$ is a major one can also be seen in the size of the support sets for the two. The matrix M for the $3 \times 3 \times 2$ problem has 47 integral rows (up from the $3 \times 2 \times 2$, having 24 rows), while that for the $3 \times 3 \times 3$ has 919 rows, many of them with fractional entries. Computing the Hilbert basis for the latter seems out of the question with current algorithms and computing capapbilities.

Nonetheless, it is possible to exploit the structure of the matrix A to extend the range of bounds problems that provably have the integer rounding property. We note that the real problem is not to actually compute the Hilbert basis from scratch, but rather only to verify that a given set of vectors (the rows of A) actually are the basis. This can make the difference between success and failure. We give next an analysis if the $3 \times 3 \times 3$ case as an illustration, and then extend this to larger three-dimensional problems.

3 The $3 \times 3 \times 3$ Problem

The matrix A for the $3\times3\times3$ bounds problem has size 27×19, with the usual structure of a 19×19 indentity matrix on top of an 8×19 array of rows containing 1's, 0's and -1's. The last 8 rows are the following, blocked off to show the structure.

1	-1	0	-1	1	0	0	0	0	-1	1	0	1	0	0	0	0	0	0
1	0	-1	-1	0	1	0	0	0	-1	0	1	1	0	0	0	0	0	0
1	-1	0	0	0	0	-1	1	0	-1	1	0	0	1	0	0	0	0	0
1	0	-1	0	0	0	-1	0	1	-1	0	1	0	1	0	0	0	0	0
1	-1	0	-1	1	0	0	0	0	0	0	0	0	0	-1	1	0	1	0
1	0	-1	-1	0	1	0	0	0	0	0	0	0	0	-1	0	1	1	0
1	-1	0	0	0	0	-1	1	0	0	0	0	0	0	-1	1	0	0	1
1	0	-1	0	0	0	-1	0	1	0	0	0	0	0	-1	0	1	0	1

Thus the overall structure looks like the following.

	I	
D	E	0
D	0	E

The structure of A suggests that we examine the following submatrices, where the identity matrix I is chosen appropriately.

$A'=$

I	
D	E

$A''=$

I
E

We identify the D and E submatrices in A' with the submatrices in the second horizontal block in A, and E in A'' with the E in the third horizontal block of A.

Lemma 2 *If the rows of A' and A'' are both Hilbert bases (for their respective cones), so are the rows of A.*

Proof: Let $I_S = \{1, \ldots, s\}$ and suppose that for some $\lambda_i \geq 0, h \equiv \sum_{I_S} \lambda_i a_i$ is integral. Let $I_{A'} \subset I_S$ denote the set of indices corresponding to A' in A, and similarly for A''. Then $\sum_{I_{A''}} \lambda_i a_i''$ is an integer vector, and because A'' is a Hilbert basis by assumption, there are integer $\alpha_i, i \in I_{A''}$ such that $\sum_{I_{A''}} \alpha_i a_i'' = \sum_{I_{A''}} \lambda_i a_i''$. Let $h' = h - \sum_{I_{A''}} \alpha_i a_i$, so that h' is integer and a non-negative linear combination of the rows of A, but with its trailing columns (those corresponding to A'') equal to zero. Since A' is a Hilbert basis, there are integers $\beta_i \geq 0, i \in I_{A'}$ such that $h' = \sum_{I_{A'}} \beta_i a_i'$. Then the coefficients $\alpha_i, i \in I_{A''}, \beta_i, i \in I_{A'}$ are integers such that $h = \sum_{I_{A''}} \alpha_i a_i + \sum_{I_{A'}} \beta_i a_i$, showing that A is a Hilbert basis. □

Both matrices A' and A'' can be checked to be Hilbert bases in a reasonable time. Computing the supports takes only seconds, and verifying that these matrices are Hilbert bases is similarly fast. Thus we have the following.

Theorem 2 *The $3 \times 3 \times 3$ bounds problem has the integer rounding property.*

Applying the procedure of Lemma 1, it is immediate that any $k \times 3 \times 3$ problem will also have the integer rounding property. Indeed, moving to the $4 \times 3 \times 3$ problem adds $(3-1)*(3-1) = 4$ rows and $3*3-4 = 5$ columns to M. An examination of the entries in these new rows and columns shows that the pattern shown in the $3 \times 3 \times 3$ case is simply continued, with a repitition of the horizontal block of four rows, along with an additional identity matrix, as follows.

<table>
<tr><td colspan="4">I</td></tr>
<tr><td>D</td><td>E</td><td>0</td><td>0</td></tr>
<tr><td>D</td><td>0</td><td>E</td><td>0</td></tr>
<tr><td>D</td><td>0</td><td>0</td><td>E</td></tr>
</table>

This extends in the obvious way to the $k \times 3 \times 3$ case. Empirical results suggest the following extensions.

Conjecture 1 *The $I \times J \times K$ bounding problem, when all $(I-1) \times (J-1) \times (K-1)$-way margins are known, has the integer rounding property.*

Conjecture 2 *The bounding problem for an n-way table, when arbitrary lower-dimensional margins are given, has the integer rounding property.*

4 Finding an Optimal Integer Solution

It may be useful to exhibit an instance of a table that achieves the optimal integer solution for a cell in T (for instance, in Senate testimony!). Experience suggests that non-integral optima are rare, so the method given here will only occasionally be required. We use a Gomory cut [8] derived from the objective function row of the simplex tableau. The idea is to use some facts evident from the optimal (continuous) simplex tableau to generate new constraints that eliminate non-integer optima.

Let's start with the example of a $2 \times 2 \times 2 \times 2$ table for which all six two-way margins are known. This is the most elementary example that has no decomposition, so should illustrate some of the complications that can arise. The dimensions are labeled W, X, Y and Z.

(z_1, y_1)	x_1	x_2
w_1	5	4
w_2	5	12

(z_1, y_2)	x_1	x_2
w_1	18	10
w_2	17	13

(z_2, y_1)	x_1	x_2
w_1	16	10
w_2	16	6

(z_2, y_2)	x_1	x_2
w_1	15	11
w_2	9	9

With variables numbered lexicographically, the constraint equations look like this:

	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	x11	x12	x13	x14	x15	x16	b
y1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	54
y2	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	35
y3	0	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0	47
y4	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	40
y5	1	1	0	0	0	0	0	0	1	1	0	0	0	0	0	0	42
y6	0	0	1	1	0	0	0	0	0	0	1	1	0	0	0	0	59
y7	0	0	0	0	1	1	0	0	0	0	0	0	1	1	0	0	32
y8	0	0	0	0	0	0	1	1	0	0	0	0	0	0	1	1	43
y9	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	26
y10	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	48
y11	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	58
y12	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	44
y13	1	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0	35
y14	0	0	1	1	0	0	1	1	0	0	0	0	0	0	0	0	54
y15	0	0	0	0	0	0	0	0	1	1	0	0	1	1	0	0	39
y16	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	1	48
y17	1	0	1	0	0	0	0	0	1	0	1	0	0	0	0	0	45
y18	0	1	0	1	0	0	0	0	0	1	0	1	0	0	0	0	56
y19	0	0	0	0	1	0	1	0	0	0	0	0	1	0	1	0	39
y20	0	0	0	0	0	1	0	1	0	0	0	0	0	1	0	1	36
y21	1	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0	37
y22	0	1	0	1	0	1	0	1	0	0	0	0	0	0	0	0	52
y23	0	0	0	0	0	0	0	0	1	0	1	0	1	0	1	0	47
y24	0	0	0	0	0	0	0	0	0	1	0	1	0	1	0	1	40

After row reduction to find an independent set, we have

	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	x11	x12	x13	x14	x15	x16	b
y1	1	0	0	0	0	0	0	1	0	0	0	1	0	1	1	3	71
y2	0	1	0	0	0	0	0	-1	0	0	0	-1	0	-1	0	-2	-28
y3	0	0	1	0	0	0	0	-1	0	0	0	-1	0	0	-1	-2	-33
y4	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	44
y5	0	0	0	0	1	0	0	-1	0	0	0	0	0	-1	-1	-2	-44

y6	0	0	0	0	0	1	0	1	0	0	0	0	0	1	0	1	36
y7	0	0	0	0	0	0	1	1	0	0	0	0	0	0	1	1	43
y8	0	0	0	0	0	0	0	0	1	0	0	-1	0	-1	-1	-2	-41
y9	0	0	0	0	0	0	0	0	0	1	0	1	0	1	0	1	40
y10	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	1	48
y11	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	40

Giving this to a simplex algorithm and maximizing x_1, we have this tableau:

		x 5	x16	x 2	x 3	x 9
x 1	22.33	-0.33	-0.33	-0.33	-0.33	-0.33
x 8	7.67	0.33	-0.67	0.33	0.33	-0.67
x10	19.67	0.33	0.33	-0.67	0.33	-0.67
x 4	31.67	0.33	0.33	-0.67	-0.67	0.33
x13	3.67	-0.67	0.33	0.33	0.33	-0.67
x 6	12.67	-0.67	0.33	-0.67	0.33	0.33
x 7	14.67	-0.67	0.33	0.33	-0.67	0.33
x11	22.67	0.33	0.33	0.33	-0.67	-0.67
x14	15.67	0.33	-0.67	0.33	-0.67	0.33
x12	4.67	-0.67	-0.67	0.33	0.33	0.33
x15	20.67	0.33	-0.67	-0.67	0.33	0.33
z	22.33	-0.33	-0.33	-0.33	-0.33	-0.33

The maximum for x_1 is 22.33, not an integer. To find the maximum *integer* solution of x_1, we can construct a Gomory cut (an additional contraint, see [8]) which reduces the feasible region while not eliminating integer lattice points. The final ($z = x_1$) line of the tableau says that, given the values of the basic variables,

$$x_1 = 22.33 - \frac{1}{3}x_5 - \frac{1}{3}x_{16} - \frac{1}{3}x_2 - \frac{1}{3}x_3 - \frac{1}{3}x_9.$$

Since we are seeking an integer solution, it's necessary that one or more of the non-basic variables x_5, x_{16}, x_2, x_3 become positive, in order to reduce the continuous maximum of 22.33 to at least 22.00. Thus we must have

$$\frac{1}{3}x_5 + \frac{1}{3}x_{16} + \frac{1}{3}x_2 + \frac{1}{3}x_3 + \frac{1}{3}x_9 \geq \frac{1}{3}$$

or

$$x_5 + x_{16} + x_2 + x_3 + x_9 \geq 1.$$

This inequality can be converted into an equality by the addition of the non-negative slack variable x_{17}:

$$x_5 + x_{16} + x_2 + x_3 + x_9 - x_{17} = 1$$

When this new constraint is added to the original LP formulation, and x_1 is maximized, we get the following tableau:

		x 9	x16	x 2	x 3	x17
x 1	22.00	0.00	0.00	0.00	0.00	-0.33
x 6	12.00	1.00	1.00	0.00	1.00	-0.67
x14	16.00	0.00	-1.00	0.00	-1.00	0.33

x 5	1.00	-1.00	-1.00	-1.00	-1.00	1.00
x 7	14.00	1.00	1.00	1.00	0.00	-0.67
x15	21.00	0.00	-1.00	-1.00	0.00	0.33
x 4	32.00	0.00	0.00	-1.00	-1.00	0.33
x 8	8.00	-1.00	-1.00	0.00	0.00	0.33
x12	4.00	1.00	0.00	1.00	1.00	-0.67
x10	20.00	-1.00	0.00	-1.00	0.00	0.33
x11	23.00	-1.00	0.00	0.00	-1.00	0.33
x13	3.00	0.00	1.00	1.00	1.00	-0.67
z	22.00	-0.00	-0.00	-0.00	-0.00	-0.33

Thus we have an integer solution, with definite proof of its optimality. Here is another example. Maximizing x_3 we get

		x11	x 7	x14	x 4	x 1
x 5	1.33	0.33	-0.67	0.33	0.33	-0.67
x 8	18.33	0.33	-0.67	0.33	-0.67	0.33
x10	14.33	0.33	-0.67	-0.67	0.33	0.33
x 3	35.67	-0.33	-0.33	-0.33	-0.33	-0.33
x13	15.33	0.33	0.33	-0.67	-0.67	0.33
x 2	18.33	0.33	0.33	0.33	-0.67	-0.67
x15	22.33	-0.67	-0.67	0.33	0.33	0.33
x 9	9.33	-0.67	0.33	0.33	0.33	-0.67
x 6	15.33	-0.67	0.33	-0.67	0.33	0.33
x12	23.33	-0.67	0.33	0.33	-0.67	0.33
x16	2.33	0.33	0.33	-0.67	0.33	-0.67
z	35.67	-0.33	-0.33	-0.33	-0.33	-0.33

This leads, in the same way, to a Gomory cut. In this case, one or more of the nonbasics x_{11}, x_7, x_{14}, x_4 or x_1 must become positive, sufficient to reduce x_3 by at least 2/3. Thus we require

$$x_{11} + x_7 + x_{14} + x_4 + x_1 \geq 2$$

Adding the slack variable x_{18} and the constraint

$$x_{11} + x_7 + x_{14} + x_4 + x_1 - x_{18} = 2.$$

we get

		x11	x 7	x19	x 4	x14
x 5	0.00	1.00	0.00	-0.67	1.00	1.00
x 6	16.00	-1.00	0.00	0.33	0.00	-1.00
x 3	35.00	0.00	0.00	-0.33	0.00	0.00
x17	60.00	3.00	3.00	-3.00	3.00	3.00
x12	24.00	-1.00	0.00	0.33	-1.00	0.00
x15	23.00	-1.00	-1.00	0.33	0.00	0.00
x16	1.00	1.00	1.00	-0.67	1.00	0.00
x 8	19.00	0.00	-1.00	0.33	-1.00	0.00

x 2	17.00	1.00	1.00	-0.67	0.00	1.00
x10	15.00	0.00	-1.00	0.33	0.00	-1.00
x 1	2.00	-1.00	-1.00	1.00	-1.00	-1.00
x13	16.00	0.00	0.00	0.33	-1.00	-1.00
x 9	8.00	0.00	1.00	-0.67	1.00	1.00
x18	1.00	0.00	0.00	1.00	0.00	0.00
z	35.00	-0.00	-0.00	-0.33	-0.00	-0.00

The idea of Gomory cuts is general—it's guaranteed to eventually give an integer solution [8], since bit-by-bit the feasible continuous polytope is whittled down to the convex hull of the feasible integer lattice points. For general IPs, however, this process may be slow—lots of required cuts. However, for the bounds problem, the polytope seems to be quite nicely shaped, so few cuts are required.

From a computational perspective, the cuts would ordinarily be added as new constraints in a dual simplex procedure. In this procedure, new constraints can be easily added without having to start the optimization problem all over again. The CPLEX software has a good implementation of the dual simplex, and has functions in its API for adding new constraints.

Conjecture 3 *The result of adding a single Gomory cut, and re-solving the LP, gives an optimal integral solution to the bounds problem.*

References

[1] Cox, L. "Some Remarks on Research Directions in Statistical Data Protection," in *Proceedings of Statistical Data Protection '98*, Eurostat, Luxembourg, 1999.

[2] Cox. L., "On Properties of Multi-Dimensional Statistical Tables," manuscript, U.S. Environmental Protection Agency, 2000.

[3] Cox, D. J. Little and D. O'Shea, *Ideals, Varieties, and Algorithms*, Springer-Verlag, New York, 1996.

[4] Giles, R. and J. Orlin, "Verifying Total Dual Integrality," manuscript, 1981.

[5] Henk, M. and R. Weismantel, "On Hilbert Bases of Polyhedral Cones," Preprint SC 96-12, Konrad-Zuse-Zentrum für Informationstechnik, Berlin, 1996. } Raymond (?)

[6] Roehrig, S., "An Efficient Algorithm for Computing Gröbner Bases for Confidentiality Problems," Technical Report, The Heinz School, Carnegie Mellon University, Pittsburgh, 2001.

[7] Roehrig, S., "Auditing Disclosure in Multi-Way Tables With Cell Suppression: Simplex and Shuttle Solutions," Technical Report, The Heinz School, Carnegie Mellon University, Pittsburgh, 1998.

[8] Schrijver, A., *Theory of Linear and Integer Programming*, Wiley & Sons, New York, 1986.

[9] Sturmfels, B., *Algorithms in Invariant Theory*, Springer-Verlag, Vienna, 1993.

[10] Weyl, H., "The Elementary Theory of Convex Polyhedra," in *Annals of Mathematics Study No. 24*, 195?.

[11] Zayatz, L. "Using Linear Programming Methodology for Disclosure Avoidance Purposes," *Proceedings of the International Seminar on Statistical Confidentiality*, Luxembourg, 1993.