Vulnerability of Complementary Cell Suppression to Intruder Attack

Lawrence H. Cox

National Center for Health Statistics, Centers for Disease Control and Prevention
3311 Toledo Road
Hyattsville, MD 20782
lcox@cdc.gov

Introduction

National statistical offices develop an institutional sense of statistical disclosure based on accumulated experience and judgment, from which agency confidentiality policies and procedures are codified. To ensure that disclosure is defined unambiguously and can be limited effectively, disclosure must be expressed quantitatively. For tabular data, this is accomplished by means of a linear functional over the measurement data called a *sensitivity measure*. The sensitivity measure identifies *sensitive* (disclosure) cells to be those cells achieving a positive value for the sensitivity measure. In addition, the sensitivity measure can be used to measure disclosure and to compute a lower bound on how much the value x of a particular sensitive cell would have to be increased (to $x + r$) to reach a hypothetical cell that is nonsensitive. Defining a corresponding lower bound, e.g., equal to $x - r$, defines the cell's *protection interval* (x-r, x+r)—equal to the set of all *unsafe estimates* of x. Statistical disclosure limitation (**SDL**) of tabular data is *complete* if and only if only the tightest (*exact*) interval estimate of each sensitive cell value x computable from the released tabulations is *safe*, viz., strictly contains the protection interval. The theory of sensitivity measures is developed fully in [1].

Complementary cell suppression (**CCS**) is a methodology for statistical disclosure limitation in tabular data. CCS replaces the value of each sensitive cell by a symbol (**D** for "disclosure"). Suppressing only these *primary suppressions* is insufficient to ensure complete SDL, and consequently additional, nonsensitive cells, called *complementary suppressions*, must also have their values replaced by **D**. CCS methodology is focused on assuring good choices for the complementary cells, viz., a collection of cells that assures complete SDL, while suppressing as little useful data as possible. These concepts are fully developed in [2, 3].

In this paper, we examine the data protection capabilities of complementary cell suppression. There are two widely used rules to define disclosure--the p-percent rule and the p/q-ambiguity rule. Treating magnitude data as continuous data, sophisticated users can compute exact interval estimates of suppressed cell values using linear programming. A fundamental question arises: shouldn't the data releaser make the same information available to all users to assist in their understanding, interpretation and analysis of suppressed magnitude data? In the next section we provide SDL preliminaries on sensitivity measures and CCS, and discuss exact intervals. Then we examine CCS mathematically, and demonstrate how a suppression pattern protects sensitive data. Finally, we present vulnerabilities of CCS: first, for CCS performed in a heuristic, as opposed to rigorous, manner; second, under the p-percent rule; third, for exact intervals under the p/q-ambiguity rule; and, finally, for reporting intervals, exact or not, under fairly general assumptions about intruder knowledge.

Statistical Disclosure Limitation Preliminaries

Complementary cell suppression has been applied to major, important data collections of economic and other magnitude data in the US, Canada and the European Union, in some cases for decades, based on software developed at the US Census Bureau, the US National Center for Health Statistics, Statistics Canada, and the EU CASC Project. Suppression has, for the most part, been used exclusively for SDL purposes in such applications. For this reason, we discuss suppression in terms of magnitude data, but remark that our conclusions are equally valid for contingency (count) data.

A simple, widely used disclosure rule for magnitude data is the *p-percent rule* which, in simplified form, states: a tabulation cell **X** is sensitive if, after subtracting the second largest contribution from the cell value, the remainder is within p-percent of the largest contribution. This rule is designed to prevent narrow estimation of any contribution to a cell value by a second contributor or third party. It follows from simple algebra that protecting the largest contributor from the second largest assures that each contributor is protected from each of the others and from an outside third party. Note that magnitude data SDL focuses on contributors as intruders, and that the risk of such *insider disclosure* exceeds that from the outside. p-percent rules have been used in US Economic and Canadian AgricultureCensuses.

X denotes a tabulation cell, and its cell value is x. Order the contributions to x from largest to smallest and denote these contributions x_i, so that $x=\sum_i x_i$; $x_1 \geq x_2 \geq .... x_i \geq ....$ Express p as a decimal; e.g., 20% = 0.20. Disclosure under the p-percent rule is expressed by the sensitivity measure: $S_p(X) = px_1 - \sum_{i \geq 3} x_i > 0$.

An enhancement of the p-percent rule that incorporates prior information available to an intruder is the *p/q-ambiguity rule*: the releaser assumes that an intruder can estimate any contribution to within q-percent, $1 \geq q >> p$. Express q as decimal. Disclosure under the p/q-ambiguity rule is expressed by the sensitivity measure: $S_{p/q}(X) = (p/q)x_1 - \sum_{i \geq 3} x_i > 0$. When $q = 1$, the p-percent and p/q-ambiguity rules are identical; otherwise, it is evident that the p/q-ambiguity rule is *stricter* than the p-percent rule, viz., it identifies as sensitive all p-sensitive cells and possibly more. The p/q-rule has been used in the Canadian Census of Manufactures.

A sensitivity measure is a continuous function. If, as above, it is normalized with final coefficient = -1, then its value r provides an exact lower bound on the distance from a sensitive cell value x to hypothetical larger cell values that are nonsensitive. We refer to r as the *upper protection limit*. Typically, $-r$ is selected as the lower protection limit, and the open interval $(x - r, x + r)$ is the *protection interval*. Upper and lower protection limits here are equal, as is typical in practice, but in general can be unequal. See [1] for complete details.

Complementary cell suppression is a very difficult theoretical and computational problem. CCS can be accomplished using mathematical programming, as follows. Represent the tabular structure as **Ay** = **t.** Entries of **A** are 0 or 1. Original data are **a** = (a_1,…., a_n), so that **Aa** = **t**. Sensitive cell values are denoted $a_{d(i)}$, i = 1, …, s, and their protection limits $r_{d(i)}$, $0 \leq r_{d(i)} \leq a_{d(i)}$, with $r_k = 0$ otherwise. A mathematical programming model for CCS is given by (1), e.g., [4].

$$
\begin{aligned}
&\min \sum_k c_k z_k \\
&i=1,....,s;\ j=1,2;\ k=1,....,n: \\
&\quad Ay_{i,j} = t \\
&\quad l_k \leq y_{i,1,k} \leq a_k - r_k z_k \\
&\quad u_k \geq y_{i,2,k} \geq a_k + r_k z_k \\
&\quad z_j = 0,1; \quad z_{d(i)} = 1
\end{aligned}
\qquad (1)
$$

The first constraint of (1) preserves the tabular structure. The second and third enforce the sensitivity measure. $M \geq 1$ is a suitable constant. Nonnegative lower (l_k) and upper bounds (u_k) on feasible cell values represent intruder prior knowledge, e.g., percentages under the p/q-rule or pseudo-bounds (0 and infinity) if prior knowledge is not assumed. The objective function is selected to preserve or optimize the releaser's notion of data quality: to minimize number of cells suppressed, set $c_k = 1$; to minimize total value suppressed, $c_k = a_k$; and, to minimize Berg entropy (a compromise between number and total value suppressed), $c_k = \log(1 + a_k)$.

Magnitude data are treated as continuous data, and therefore exact interval estimates of suppressed cell values y_k can be obtained via linear programming: min y_k (respectively, max y_k) subject to **Ay** = **t.** The exact interval for a_k is [min y_k, max y_k]. The model constraints of (1) assure that exact intervals contain protection intervals. It has been argued that the releaser should provide exact intervals in lieu of suppressed data under the p-percent rule. Similarly, for the p/q-rule, as the releaser assumes that intruder can estimate data within q-percent, shouldn't the releaser provide these estimates for use by legitimate analysts? The question of releasing intervals has been asked since the 1970s, reemerging recently as *partial cell suppression* [5, 6]. If exact interval estimates of suppressed values are released, the user could impute interval midpoints for suppressed data and analyze the "midpoint tables." Less sophisticated users are likely to do this because it is simple. The resulting tables may fail to be additive, but additivity could be regained through *controlled tabular adjustment* ([7]) or application of iterative proportional fitting (IPF) to impute suppressed values ([8]).

Mathematical Properties of CCS

Complementary cell suppression replaces all sensitive and selected nonsensitive values, which are fixed, by symbols, which can be treated as variables. By definition, a particular CCS solution is complete if all exact intervals for variables corresponding to sensitive values x contain the cell's protection interval $(x - r, x + r)$. Table 1 provides a working example.

Table 1. 4x5 Working Example.

T	O	T	A	L	T
					O
	X(10)		**B(5)**		T
					A
	C(7)		**A(8)**		L

X denotes a sensitive cell, and **A, B, C** denote **X**'s complementary suppressions. For our analysis, we extract the essence of Table 1 and provide hypothetical values for the reduced marginal totals, represented in Table 2.

Table 2. Essentials of Table 1.

17	13	30
x=10	b=5	15
c=7	a=8	15

As it contains only four suppressions, Table 2 is the simplest possible example in two or more dimensions. Nevertheless, Table 2 is a building block, as complete suppression patterns in two dimensions can be decomposed into patterns containing 4 or 6 or 8 … or 2k cells (*even cycles*) and our mathematical analysis applies mutatis mutandis. It is simply easier to see and present in the simplest, 4-element, case. In higher dimensions or linked tables, each two dimensional slice of a complete pattern is a complete two dimensional pattern and consequently amenable to this analysis.

Let $r = 2$. Then **X** is protected if and only if any interval derivable for x contains $(x\text{-}r, x\text{+}r) = (10\text{-}2, 10\text{+}2) = (8, 12)$. This condition holds if **X** is in an *alternating cycle* of suppressed cells and if the cycle permits a *flow* of $r = 2$ units from x = 10 in both **+** and **–** directions. The alternating cycle is given by

17	13	30
X (10)+/-	**B** (5)-/+	15
C (7)-/+	**A** (8)+/-	15

In the + direction, we can move up to 5 units into **X**--more would force $b < 0$.

17	13	30
X (15)	**B** (0)	15
C (2)	**A** (13)	15

In the – direction, we can move up to 8 units out of **X**--more would force $a < 0$.

17	13	30
X (2)	**B** (13)	15
C (15)	**A** (0)	15

Verification that Table 2 protects **X** is demonstrated by exact intervals below. As we can move $r = 2$ units in either direction, **X** is protected.

17	13	30
X [2, 15]	**B** [0, 13]	15
C [2, 15]	**A** [0, 13]	15

Movement of up to 5 (respectively, 8) units through sensitive cell **X** is represented by the following alternating cycle.

17	13	30
x+/-	**b**-/+	15
c-/+	**a**+/-	15

Cells marked with +/- have the *same parity* as *x;* those with -/+ have *opposite parity* to *x*. In general,

- maximum increase to x = minimum value with opposite parity (here, b = 5)
- maximum decrease to x = minimum value with same parity (here, a = 8)
- exact interval for $x = [x\text{-}a,\ x\text{+}b]$ (here, = [2, 15])
- *width* of exact interval = $(b+a)$ (here, = 13)
- *radius* of exact interval = $(b+a)/2$ (here, = 6.5)
- interval *midpoint* = $x + (b\text{-}a)/2$ (here, = 8.5)
- *bias* in midpoint estimate of $x = (b\text{-}a)/2$ (here, = -1.5)

CCS is based on creating cycles that

- contain the sensitive cells **X**
- collectively permit increase/decrease of x to at least $(x\text{-}r(X),\ x\text{+}r(X))$
- minimize information loss measured by linear cost function $\sum_k c_k z_k$

Confidentiality Characteristics of CCS

If complementary cell suppression is performed using a mathematical model that incorporates protection constraints explicitly, such as (1), exact intervals for suppressed sensitive cells must be nonsensitive and disclosure limitation is complete. Model (1) is a mixed integer linear program, which can be difficult or impossible to solve computationally by direct means such as branch and bound, except for small problems. Recent research has focused on solving medium to large CCS problems using branch and cut and specialized techniques [4]. Unfortunately, many organizations continue to solve CCS problems "by hand" or using computer programs based on "by hand" reasoning. These programs are faster than humans, but in the absence of CCS methodology, offer little improvement in terms of protection or data quality. We give two examples based on a typical disclosure rule for counts that defines the unsafe protection interval to be the interval (0, 5) (*5-threshold rule*).

Table 3. 3x3 Table With Internal Entries Suppressed.

$$\begin{pmatrix} * & * & * \\ * & * & * \\ * & * & * \end{pmatrix}\begin{pmatrix} 11 \\ 5 \\ 5 \end{pmatrix} \quad \begin{pmatrix} * & * & * \\ * & * & * \\ * & * & * \end{pmatrix}\begin{pmatrix} 5 \\ 11 \\ 5 \end{pmatrix} \quad \begin{pmatrix} * & * & * \\ * & * & * \\ * & * & * \end{pmatrix}\begin{pmatrix} 5 \\ 5 \\ 11 \end{pmatrix}$$

$$\begin{pmatrix} 11 & 5 & 5 \end{pmatrix}(21) \quad \begin{pmatrix} 5 & 11 & 5 \end{pmatrix}(21) \quad \begin{pmatrix} 5 & 5 & 11 \end{pmatrix}(21)$$

$$\begin{pmatrix} 1 & 10 & 10 \\ 10 & 1 & 10 \\ 10 & 10 & 1 \end{pmatrix}$$

Table 3 is a 3x3x3 contingency table with all internal entries suppressed. Release of Table 3 is equivalent to release of all 2-dimensional table cells ("line" marginal totals) in lieu of the 3-dimensional cells (internal table entries). Table 3 is not a realistic confidentiality example because it contains published marginal totals with value = 1, failing the 5-threshold rule. We ignore this issue momentarily, returning to it in the next paragraph. Meantime, compute 2-dimensional Frechet lower bounds for cells (1,1,1), (2,2,2) and (3,3,3) within planes k = 1, 2, 3, respectively. Each of these lower bounds = 1. Each of these three cells is constrained by a marginal total (vertical) = 1, and consequently these cells cannot achieve value > 1. Hence, each has value =1, which has been revealed and is sensitive--disclosure limitation in this hypothetical example would have been entirely unsuccessful.

We return to the issue of realism. Replace Table 3 by a table comprising 5 copies of Table 3 stacked vertically, viz., a 3x3x15 table with two sets of planar marginals unchanged and the third (vertical) set with values five times those of Table 1 (table not drawn here). This is a realistic example (no marginals < 5) for which 15 cells are revealed to have value = 1, so that disclosure limitation fails completely.

A second example illustrating the failure of non-mathematical CCS methods to protect data is given by Table 4, a two-dimensional table with suppressions.

Table 4. 4x5 Table with Suppressions.

18	**21**	**18**	**23**	**80**
D_{11}	D_{12}	D_{13}	9	**20**
6	D_{22}	D_{23}	6	**20**
D_{31}	5	5	D_{34}	**15**
D_{41}	5	6	D_{44}	**25**

CCS in Table 4 may appear successful, as each suppressed cell is contained in a row and a column containing one or two additional suppressions and as corresponding sums are ≥ 5. But, in fact, $D_{11} = 1$ can be deduced: add the first two rows: $D_{11} + D_{12} + D_{13} + D_{23} + D_{33} = 19$; add the second and third columns: $D_{12} + D_{13} + D_{23} + D_{33} = 18$; subtract the latter from the former, to obtain $D_{11} = 1$. Disclosure limitation has failed.

Both examples illustrate that CCS should be done based on a verifiable mathematical model and NOT "by hand" or by software based, in essence, on "by hand" reasoning. For continuous data, these flaws can be detected by linear programming. For contingency data, efficient methods for computing exact intervals are available only for specialized classes of tables ([10, 11]). A heuristic iterative min-max method--a *shuttle algorithm* ([12])—to refine inexact intervals and on iterative midpoint refinement to construct consistent tables, is presented in ([13]).

The data releaser may opt to release exact intervals [l, u] for the suppressed cells. Even if the releaser does not do so, the sophisticated user can compute these intervals independently, at least for continuous data. So, it suffices to assume that exact intervals are available. We return to our working example. Again, we remind the reader that all situations are not as simple as this 4-element two-dimensional cycle; but, that all situations do comprise two-dimensional cycles that the intruder can analyze in precisely the same manner as we now proceed to do.

17	13	30
x+/-	**b**-/+	15
c-/+	**a**+/-	15

Assume for concreteness that $b \leq c$ and $a \leq x$. By virtue of the polyhedral geometry of linear constraint systems, the intruder can determine the following.

- $l(x) = x - a$: a of same parity as x, and $l(a) = 0$
- $u(x) = x + b$: b of opposite parity to x, and $l(b) = 0$
- intruder knows the width of the exact interval = $a + b$
- if intruder can determine a or b or b-a or b/a, then x is revealed

Consequently, protection on a cycle hinges on the intruder's ability to determine a single quantity. If $(b-a)/(2x)$ is small, then the midpoint estimate is precise. Similarly, if **A**, **B** are not historically sensitive, then the intruder can examine historical data to estimate a or b or $b - a$ or b/a, directly, or via regression, and consequently estimate x.

Often contributor counts are released, so the intruder knows precisely the one contributor cells. If **X** involves two contributors, then by subtracting its own value, the second contributor can obtain even sharper bounds than those below.

If **X** involves only one contributor, then the intruder can deduce

$$l(x) \leq (1-p)x$$

$$(1+p)x \leq u(x)$$

Consequently,

$$l(x)/(1-p) \leq x \leq u(x)/(1+p) \qquad (2)$$

which is sharper than $l(x) \leq x \leq u(x)$. In addition, $u(x)/l(x) \geq (1+p)/(1-p)$, so that $(u(x)-l(x)/(u(x)+l(x)) \geq p$.

By virtue of (2), exact intervals can be "shrunk" if p is known. It is often discussed as to whether the releaser should make the value of p public to enhance analyzability of the data. It would appear that the answer to that question is a resounding "NO". The next question is whether the releaser should release the minimal safe interval (x – r(X), x + r(X)**).** Again the answer is "NO" because in so doing, x = midpoint of (x-r(X), x+r(X)) is divulged, as are $r = r(X) = (x + r(X))$

– *(x)* and $p = r/x$. Under *sliding protection* ([14]), which requires only that the width of the protection interval be at least $2r(X)$, the second question is moot.

If releasing *p* erodes protection, how well protected is the value of this parameter? For each one or two contributor cell **X**

$$\begin{aligned} p \le p(X) &= (u(x)-l(x))/(u(x)+l(x)) \\ &= ((u(x)-l(x))/2)/((u(x)+l(x))/2) \qquad (3) \\ &= \text{(radius/midpoint) of the protection interval for } \mathbf{X} \end{aligned}$$

These inequalities provide (many) upper bounds for *p*. In the context of a national census, or a set of different censuses or censuses conducted over multiple years, and a great many one contributor cells, many upper bounds (3) for *p* become available. The smallest, $p' = p(X')$, could be very precise.

A lower bound p'' on *p* can be substituted into (2) to sharpen estimation of sensitive single contributor values *x*. The intruder thus obtains a tighter interval than the exact interval $l \le x \le u$:

$$l \le l/(1-p'') \le l/(1-p) \le x \le u/(1+p) \le u/(1+p'') \le u \qquad (4)$$

A lower bound can obtained via trial and error as follows.

- begin with any solution (e.g,, adjusted midpoint or IPF)
- choose p'' and protect **X** to within p"-percent
- if the current cycle is not selected, then $p > p''$
- do this for each one contributor cell **X**
- the largest p'' is a lower bound for *p*

Enhancements to the p/q–ambiguity related to weighting and imputation are examined in ([15]). In this section, we discover fundamental weaknesses related to release of exact intervals under a p/q-rule. **X** denotes a sensitive cell under a p/q-rule, and is suppressed with complementary suppressions **A**, **B**, **C**. Assume that exact intervals are released in place of suppressions.

Table 5. Alternating Cycle for Cell Suppression

X $[l_X, u_X]$……+/-	**B** $[l_B, u_B]$ -/+
C $[l_C, u_C]$ -/+	**A** $[l_A, u_A]$ +/-

Assume $l_A, l_C, l_X \ge l_B$ (other cases analogous). Thus, $a, c, x \ge b$. From the polyhedral geometry of linear constraints, the intruder can deduce

$$u_X - l_X = u_B - l_B = u_A - l_A = u_C - l_C = 2q \min\{a, b, c, x\} = 2qb$$

$$\begin{aligned} l_B &= (1-q)b & l_C &= c - qb \\ u_B &= (1+q)b & u_C &= c + qb \\ l_A &= a - qb & l_X &= x - qb \\ u_A &= a + qb & u_X &= x + qb \end{aligned}$$

Should the releaser release the value of *q*? The answer is definitely "NO" because these equations would reveal *a, b, c* and *x*. Indeed, it makes no difference whether or not the releaser reveals *q*, as *q* is in fact knowable. For $q < 1$,

$$u_B / l_B = (1+q)/(1-q)$$

$$q = (u_B - l_B)/(u_B + l_B)$$

resulting in

$$\begin{aligned} b &= l_B/(1-q) \\ a &= l_A + qb \\ c &= l_C + qb \\ x &= l_X + qb \end{aligned}$$

Consequently, release of exact intervals for a p/q-rule results in completely failed disclosure limitation when the suppression pattern corresponds to an alternating cycle. The general case is examined in the next section.

The p/q-ambiguity rule assumes that intruder knowledge is uniform across all contributions and cells, in that the intruder can obtain lower and upper bounds for any value with q-percent accuracy, for fixed $q < 1$ (expressed as a decimal). A natural generalization is to permit *symmetric intruder knowledge* which may vary from cell to cell, viz., for cells A, B, C, X, the releaser assumes that the intruder can obtain lower or upper bounds for cell values or its contributions accurate to within fractions $0 \le q_A, q_B\ q_C\ q_X < 1$, respectively. In this scenario, the releaser incorporates the corresponding constraints into the mathematical model (1) for cell suppression, solves the model for a final suppression pattern, and reports an exact interval for each suppressed cell value. Unfortunately, this scheme is also vulnerable, because along a suppression pattern based on an alternating cycle such as Table 5, we have:

$$\begin{aligned} a &= u_A - \frac{u_B - l_B}{2} \\ b &= u_B - \frac{u_B - l_B}{2} \\ c &= u_C - \frac{u_B - l_B}{2} \\ x &= u_X - \frac{u_B - l_B}{2} \end{aligned} \tag{5}$$

The reason why this scheme fails is given by the following proposition.

Proposition 1 Exact intervals $[l_X, u_X]$ for a suppression pattern based on an alternating cycle that are computed assuming symmetric intruder knowledge are symmetric around true values x, viz., $x = \frac{u_X + l_X}{2}$, and thus true values are revealed.

It is straightforward to generalize Proposition 1 to suppression patterns in two-way tables and further to tables of network type ([10]). Indeed, the same result holds in complete generality.

Let **A'y = t'** denote the *system of suppression equations.*, derived from the original system **Ay = t** by replacing unsuppressed entries by their true values. Let $l_k \le y_k \le u_k$ denote constraints corresponding to an assumption of symmetric intruder knowledge and let **l**, **u** denote the corresponding vectors. S is the number of suppressed cells. Define the linear program

$$L: \quad \begin{aligned} &\max(0) \\ &A' \begin{pmatrix} y_1 \\ \vdots \\ y_S \end{pmatrix} = t' \\ &l_k \le y_k \le u_k \qquad k = 1, \ldots, S \end{aligned} \tag{6}$$

Theorem 1 Exact intervals for a suppression pattern computed assuming symmetric intruder knowledge $l_k \le y_k \le u_k$ are symmetric around true values, regardless of the underlying tabular structure.

Thus, if the releaser wishes to provide bounds for suppressed entries, these bounds must be nonsymmetric. If the original disclosure rule gives rise to nonsymmetric protection intervals, then $x_k^{(0)} - l_k, u_k - x_k^{(0)}$ are not necessarily equal, and interval midpoints are not necessarily equal to true values. If, in addition, no functional relationship exists between l_k, u_k, then the releaser might be able to provide exact intervals for true values. Otherwise, the releaser might adopt the following procedure or an adaptation for the case $x_k^{(0)} - l_k = u_k - x_k^{(0)}$.

Procedure: Randomly select values $l_k^{'}$, $l \le l_k^{'} \le a_k - r_k$, $k = 1, \ldots, S$ from S independent uniform distributions. If $l_k^{'} - l_k \ge a_k - r_k - l_k^{'}$, define $n_\kappa = a_k + r_k + (a_k - r_k - l_k^{'})$ and $\sigma_\kappa = \frac{1}{4}(a_k - r_k - l_k^{'})$; otherwise $n_\kappa = u_k - (l_k^{'} - l_k)$ and $\sigma_\kappa = \frac{1}{4}(l_k^{'} - l_k)$. Randomly select values $u_k^{'}$, $u_k \ge u_k^{'} \ge a_k + r_k$, $k = 1, \ldots, S$ from S independent truncated normal distributions $N(n_k, \sigma_k^2)$, $k = 1, \ldots, S$, truncated to intervals $[n_\kappa - 4\sigma_\kappa, n_\kappa + 4\sigma_\kappa)]$. In (1), replace symmetric bounds l_k, u_k by general bounds $l_k^{'}, u_k^{'}$.

A triangular or other symmetric distribution can be substituted for the truncated normal. The procedure resolves the symmetry problem (deterministic), and the choice of distribution (stochastic) assures that the expected value of each interval midpoint equals the true value. This property is useful, e.g, for data analysis at higher levels of aggregation. Conversely, if this property is unnecessary, a different choice of distributions can be made.

This procedure resolves the problem raised by Theorem 1. But, there is a caveat. To be effective in moving interval midpoints off true values, any procedure must in some cases narrow the original interval, perhaps considerably. Doing so opens the releaser to the vulnerability discussed previously.

Author's Statement This work solely represents the findings and opinions of the author and should not be interpreted as representing the policies or practices of the Centers for Disease Control and Prevention or any other organization or group. Partial results of this research were presented in [16].

References

1. Cox LH. Linear sensitivity measures in statistical disclosure control. *Journal of Statistical Planning and Inference* **5** (1981) 153-164.
2. Cox LH. Suppression methodology and statistical disclosure control. *Journal of the American Statistical Association* **75** (1980) 377-385.
3. Cox LH. Network models for complementary cell suppression. *Journal of the American Statistical Association* **90** (1995) 1453-1462.
4. Fischetti, M and JJ Salazar. Solving the cell suppression problem on tabular data with linear constraints. *Management Science* **47/7** (2001) 1008-1026.
5. Fischetti M and JJ Salazar. Partial cell suppression: a new methodology for statistical disclosure control. *Statistics and Computing* **13** (2003) 13–21.
6. Salazar JJ. A unified mathematical programming framework for different statistical disclosure limitation methods. *Operations Research* **53** (2005) 819–829.
7. Cox LH, JP Kelly and R Patil. Balancing quality and confidentiality for multivariate tabular data. in: **Privacy in Statistical Databases 2004, Lecture Notes in Computer Science 3050** (J Domingo-Ferrer and V Torra, eds.), Berlin: Springer-Verlag (2004) 87-98.
8. Cox, LH, JG Orelien and BV Shah. A method for preserving statistical distributions subject to controlled tabular adjustment. in: **Privacy in Statistical Data Bases 2006, Lecture Notes in Computer Science 4302** (J Domingo-Ferrer and L Franconi, eds.), Heidelberg: Springer (2006) 1-11.
9. Cox, LH. On properties of multi-dimensional statistical tables. *Journal of Statistical Planning and Inference* **117** (2003) 251-273.
10. Cox, LH. Contingency tables of network type: models, Markov basis and applications. *Statistica Sinica* **17** (2007) 1371-1393.
11. Dobra, A and SE Fienberg. Bounds for cell entries in contingency tables given marginal totals and decomposable graphs. *Proceedings of the National Academy of Sciences* **97** (2000) 11885-11892.
12. Buzzigoli, L and A Giusti. 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, Lisbon, 25 to 27 March 1998, Luxembourg: European Communities (1999) 131-147.
13. Isserman AM and J Westervelt. 1.5 million missing numbers: overcoming employment suppression in *County Business Patterns* data. *International Regional Science Review* **29** (2008) 311-335.
14. Kelly, JP, BL Golden and AA Assad. Cell suppression: disclosure protection for sensitive tabular data. *Networks* **22** (1992) 397-417.
15. O'Malley M and LR Ernst. Practical considerations in applying the pq-rule for primary disclosure suppression. 2007 Proceedings of the American Statistical Association, Survey Research Methods Section, [CD-ROM}, Alexandria, VA: American Statistical Association (2007).
16. Cox, LH. A data quality and data confidentiality assessment of complementary cell suppression. in: **Privacy in Statistical Data Bases 2008, Lecture Notes in Computer Science 5262** (J Domingo-Ferrer and Y Saygin, eds.), Heidelberg: Springer (2008) 13-23.