ALGORITHMS FOR MAKING TABLES ADDITIVE: RAKING, MAXIMUM LIKELIHOOD, AND MINIMUM CHI-SQUARE

James T. Fagan and Brian V. Greenberg, Bureau of the Census [1/]
Washington, D.C. 20233

KEY WORDS: contingency tables, iterative proportional fitting, raking, marginal constraints.

1. FEASIBLE CONTINGENCY TABLES

1.1. Introduction. Given a contingency table of non-negative reals in which the internal entries do not sum to the corresponding marginals, there is often the need to adjust internal entries to achieve additivity. In many applications, the objective is to have zero entries in the original table remain zero in the table and positive entries remain positive. Not all two-way contingency tables can be adjusted to achieve additivity subject to these constraints, and in (Fagan and Greenberg, 1987) the authors presented a procedure to determine whether a table can be adjusted, and such adjustable tables were called feasible.

In general, given a feasible table one seeks a derived table that is close. For every criterion of closeness a different objective function must be optimized. Three of the most cited criteria of closeness are: (a) Raking, (b) Maximum Likelihood, and (c) Minimum Chi-Square. In this paper we provide algorithms which converge to a revised table optimizing each objective function. We provide a unified method showing that each of the algorithms converges to a unique table of specified closeness.

Each measure of closeness is framed as a function to be minimized subject to marginal constraints. Starting with the primal (original) objective function we form the dual which we maximize. Maximizing the dual function is an unconstrained optimization problem amenable to iterative coordinate descent methods. These techniques yield iterative algorithms converging to a solution of the dual problems and subsequently to the original.

Table adjustment is used to reconcile tabular data when marginals and internal entries arise from different sources. Internal entries are adjusted when marginals are considered more reliable -- for example, marginals may be derived from 100% census data whereas internal entries may arise from a sample. One application of raking at the Census Bureau is to weight responses to the census long-form which was mailed on a sample basis. Marginals were obtained from the full census count and internal cells are weighted to be comparable to marginal distributions. An excellent discussion of these procedures is contained in a series of four papers: (Fan, Woltman, Miskura, and Thompson, 1981); (Kim, Thompson, Woltman, and Vajs, 1981); (Thompson, 1981); and (Woltman, Miskura, Thompson, and Bounpane, 1981). Four recent papers relating to table adjustment for estimation and weighting are: (Copeland, Peitzmeier, and Hoy, 1987); (Alexander, 1987); (Lemaitre and Dufour, 1987); and (Oh and Scheuren, 1987).

In the next section are definitions, preliminary results, and simplifying reductions. In Section 2 objective functions are obtained, and in Section 3 we introduce duality for non-linear optimization, derive dual functions, and relate their optimal values to the original problems. In Section 4 we introduce cyclic coordinate descent to derive algorithms. In Section 5 we provide an example and concluding remarks. Extensive details and proofs omitted from this paper due to space limitations are contained in the report (Fagan and Greenberg, 1985), from which this paper is an extract. Although this paper is presented in terms of two-way tables, the results extend for higher dimensions.

1.2. Feasible Tables. By a contingency table we mean a triple $\mathbf{A} = \{(a_{ij}), \mathbf{r}, \mathbf{c}\}$ of arrays of non-negative reals where (a_{ij}) is an RxC matrix, $\mathbf{r} = (r_1, \ldots, r_R)$, $\mathbf{c} = (c_1, \ldots, c_C)$, and

$$\sum_{i=1}^{R} r_i = \sum_{j=1}^{C} c_j .$$

We say that $\mathbf{A}$ is additive if

$$\sum_{j=1}^{C} a_{ij} = r_i \qquad i = 1, \ldots, R$$

$$\sum_{i=1}^{R} a_{ij} = c_j \qquad j = 1, \ldots, C .$$

The table $\mathbf{A}$ is said to be feasible if there exists an RxC matrix (b_{ij}) such that $b_{ij} = 0$ if and only if $a_{ij} = 0$ and $\mathbf{B} = \{(b_{ij}), \mathbf{r}, \mathbf{c}\}$ is additive, and we say that $\mathbf{B}$ is derived from $\mathbf{A}$. That is, A is feasible if and only if there exists an RxC matrix (x_{ij}) such that $(b_{ij}) = (x_{ij} a_{ij})$, satisfying:

$$(1) \qquad \sum_{j \in V_R(i)} x_{ij} a_{ij} = r_i \qquad i=1,\ldots,R$$

$$(2) \qquad \sum_{i \in V_C(j)} x_{ij} a_{ij} = c_j \qquad j=1,\ldots,C$$

$$(3) \qquad x_{ij} > 0 \qquad (i,j) \in V,$$

where $V = \{(i,j) \mid (i,j) \in RxC \text{ and } a_{ij} \neq 0\}$,

$V_R(i) = \{j \mid (i,j) \in V\} \quad i=1,\ldots,R$

$V_C(j) = \{i \mid (i,j) \in V\} \quad j=1,\ldots,C$.

1.3. Connected Tables. We say that matrix A is connected if A cannot be written as a direct sum of submatrices. What we are calling a connected matrix has been called inseparable by Savage (1973). We say $\mathbf{A}$ is a connected table if A is a connected matrix. Since every table can be written as the direct sum of connected tables, without loss of generality we can confine our attention to connected tables, and will do so for the remainder of this paper.

2. DERIVED TABLES OPTIMIZING CLOSENESS CRITERIA

2.1. Criteria For Optimal Derived Tables. Given a feasible table **A** one seeks a derived additive table **B** "close" to **A**. Four measures of closeness are:

$$(d_1): \sum_{(i,j)\varepsilon V} b_{ij}\ln(b_{ij}/a_{ij})$$

$$(d_2): \sum_{(i,j)\varepsilon V} -a_{ij}\ln(b_{ij}/a_{ij})$$

$$(d_3): \sum_{(i,j)\varepsilon V} (a_{ij}-b_{ij})^2/b_{ij}$$

$$(d_4): \sum_{(i,j)\varepsilon V} (a_{ij}-b_{ij})^2/a_{ij} .$$

If

$$\sum_{(i,j)\varepsilon V} a_{ij} = \sum_{(i,j)\varepsilon V} b_{ij} = 1,$$

the a_{ij} and b_{ij} could be probabilities, and **A** an observed distribution and **B** a revised distribution conforming to marginal constraints.

Replacing b_{ij} by $a_{ij}x_{ij}$ in (d_1) - (d_4), and making reductions it suffices to minimize the functions of x_{ij} below subject to constraints (1)-(3):

$$(f_1): \sum_{(i,j)\varepsilon V} a_{ij}x_{ij}\ln x_{ij}$$

$$(f_2): \sum_{(i,j)\varepsilon V} -a_{ij}\ln x_{ij}$$

$$(f_3): \sum_{(i,j)\varepsilon V} a_{ij}/x_{ij}$$

$$(f_4): \sum_{(i,j)\varepsilon V} a_{ij}x_{ij}^2 .$$

Criterion (d_4) was introduced by Deming and Stephan (1940) as weighted least squares. The function (f_4) can be optimized by solving a system of linear equations however, the optimal may occur when some x_{ij} are not positive. For this reason (f_4) will not be discussed further in this report. Least squares is in an Appendix to (Fagan and Greenberg, 1985). Criterion (d_3) has been referred to as Minimum Chi-Square because of its resemblence to the Chi-Square statistic. It was introduced in regard to table adjustment by Smith (1947) and is discussed by Causey (1983, 1984).

The objective function in (d_2) and (f_2) is a likelihood function if internal entries of **A** are counts based on a multinomial distribution whose total is N. Under these assumptions, we have the likelihood:

$$L=L(b_{ij}: a_{ij})= \frac{N!}{\prod_{(i,j)\varepsilon V} a_{ij}} \prod_{(i,j)\varepsilon V} (b_{ij})^{a_{ij}} .$$

Note that $\ln(L)$ achieves its maximum at the same point as does:

$$\sum_{(i,j)\varepsilon V} a_{ij}\ln x_{ij}.$$

Criteria (d_1) and (f_1) have an interesting history. A procedure called raking was introduced by Deming and Stephan (1940). One starts with a table and alternately scales rows and columns to achieve additivity, and if a table is feasible the algorithm converges. The authors claimed this procedure would yield a table optimizing (d_4). It was later pointed out by Stephan (1942) that raking did not converge to a table optimizing least squares. Raking did appear to converge for positive tables, however, and workers in the area were quite pleased to use it.

Sinkhorn (1967) showed that raking converges for positive tables. Ireland and Kullback (1968) observed that when raking converged, it converged to a table optimizing (d_1), and their findings was modified by Thompson (1981). Darrach and Ratcliff (1972) independently showed that for feasible tables raking converged to optimize (d_1). Raking has been the object of extensive analysis; see (Oh and Scheuren, 1978) for an extensive bibliography, and (Bishop, Fienberg, Holland, 1975) and (Fienberg, 1970) for further discussion.

Relating raking to information theory, Ireland and Kullback (1968) observe that if **A** is a table of probabilities, the table derived by raking **B** will minimize the discrimination information of **A** with respect to **B**:

$$I((b_{ij}),(a_{ij})) = \sum_{(i,j)\varepsilon V} b_{ij}\ln(b_{ij}/a_{ij}),$$

namely objective function (d_1), see (Kullback, 1959). Note further that the discrimination function of **B** with respect to **A**,

$$I((a_{ij}),(b_{ij})) = \sum_{(i,j)\varepsilon V} a_{ij}\ln(a_{ij}/b_{ij}) ,$$

is objective function (d_2). In the derivations below we will minimize the objective function

$$f_1= \sum_{(i,j)\varepsilon V} a_{ij}x_{ij}(-1+ \ln x_{ij})$$

instead of

$$\sum_{(i,j)\varepsilon V} a_{ij}x_{ij}\ln x_{ij},$$

since both have the same critical values and the former is easier to work with.

3. DERIVING THE DUAL FUNCTIONS

3.1. Introduction. If **A** is feasible then at least one derived table exists, and there exist a unique derived table for which each of the functions f_1,f_2 and f_3 attains its minimum subject to (1) - (3). The point at which these functions attain their minimum subject to only (1) and (2) has $x_{ij} > 0$ for all $(ij)\varepsilon V$. Thus, condition (3) is not required, see (Fagan and Greenberg, 1985). In this section, we provide a unified method to find the optimal table for each of these functions. For $f\varepsilon\{f_1, f_2, f_3\}$, our goal is to solve the following primal problem:

Minimize: $f(\underline{x})$ subject to (1) and (2).

We first form the Lagrangian:

$$L(\underline{x},\underline{\mu},\underline{\lambda}) = f(\underline{x}) \pm \sum_{i=1}^{R} \mu_i\Big(\sum_{j\epsilon V_R(i)} a_{ij}x_{ij} - r_i\Big)$$

$$\pm \sum_{j=1}^{C} \lambda_j\Big(\sum_{i\epsilon V_C(j)} a_{ij}x_{ij} - c_j\Big).$$

Minimize $L(\underline{x},\underline{\mu},\underline{\lambda})$ as a function of $\underline{x}$, $\underline{\mu}$, and $\underline{\lambda}$ and solve for critical $\underline{x}$ values in terms of $\underline{\mu}$ and $\underline{\lambda}$ resulting in the dual function:

$$H(\underline{\mu},\underline{\lambda}) = \underset{\underline{x}>0}{\mathrm{Min}}\ \{L(\underline{x},\underline{\mu},\underline{\lambda})\}\ .$$

Note that $H(\underline{\mu},\underline{\lambda})$ is a function of $\underline{\mu}$ and $\underline{\lambda}$ which we maximize, that is, we solve the dual problem. Under conditions satisified by each of the functions f_1, f_2, and f_3, the maximum of $H(\underline{\mu},\underline{\lambda})$ equals the minimum of the corresponding $f(\underline{x})$ constrained by (1) and (2). The solution for $\underline{x}$ in terms of $\underline{\mu}$ and $\underline{\lambda}$ which maximize $H(\underline{\mu},\underline{\lambda})$ yields the value of $\underline{x}$ that minimizes f subject to (1) and (2). The advantage in going from the primal problem to the dual is that we replace a constrained optimization with a non-constrained problem.

3.2. Deriving the Dual Functions.

Raking

Minimize $f_1(\underline{x}) = \sum_{(i,j)\epsilon V} a_{ij}x_{ij}(-1+\ell n\ x_{ij})$

$$L_1(\underline{x},\underline{\mu},\underline{\lambda}) = \sum_{(i,j)\epsilon V} a_{ij}x_{ij}(-1+\ell n x_{ij})$$

$$-\sum_{i=1}^{R} \mu_i\Big(\sum_{j\epsilon V_R(i)} a_{ij}x_{ij}-r_i\Big)$$

$$-\sum_{j=1}^{C} \lambda_j\Big(\sum_{i\epsilon V_C(j)} a_{ij}x_{ij}-c_j\Big).$$

To minimize $L_1(\underline{x},\underline{\mu},\underline{\lambda})$ we take partial derivatives, set them equal to zero and solve for x_{ij} in terms of μ_i and λ_i and obtain:

(4) $$x_{ij} = e^{\mu_i+\lambda_j} \qquad \text{for } (i,j)\epsilon V.$$

Replacing x_{ij} by $e^{\mu_i+\lambda_j}$ in $L_1(\underline{x},\underline{\mu},\underline{\lambda})$ yields:

$$H_1(\underline{\mu},\underline{\lambda}) = \sum_{i=1}^{R} \mu_i r_i + \sum_{j=1}^{C} \lambda_j c_j - \sum_{(i,j)\epsilon V} a_{ij}e^{(\mu_i+\lambda_j)}.$$

Maximum Likelihood

Minimize $f_2(\underline{x}) = \sum_{(i,j)\epsilon V} -a_{ij}\ell n x_{ij}$

$$L_2(\underline{x},\underline{\mu},\underline{\lambda}) = -\sum_{(i,j)\epsilon V} a_{ij}\ell n x_{ij} +$$

$$\sum_{i=1}^{R} \mu_i\Big(\sum_{j\epsilon V_R(i)} a_{ij}x_{ij}-r_i\Big)$$

$$+\sum_{j=1}^{C} \lambda_j\Big(\sum_{i\epsilon V_C(j)} a_{ij}x_{ij}-c_j\Big)$$

(5) $$x_{ij} = 1/(\mu_i+\lambda_j) \qquad \text{for } (i,j)\epsilon V$$

$$H_2(\underline{\mu},\underline{\lambda}) = \sum_{(i,j)\epsilon V} a_{ij}\ell n(\mu_i+\lambda_j) - \sum_{i=1}^{R} \mu_i r_i - \sum_{j=1}^{C} \lambda_j c_j$$

Minimum Chi-Square

Minimize $f_3(\underline{x}) = \sum_{(i,j)\epsilon V} a_{ij}/x_{ij}$

$$L_3(\underline{x},\underline{\mu},\underline{\lambda}) = \sum_{(i,j)\epsilon V} a_{ij}/x_{ij} +$$

$$\sum_{i=1}^{R} \mu_i\Big(\sum_{j\epsilon V_R(i)} a_{ij}x_{ij}-r_i\Big)$$

$$+\sum_{j=1}^{C} \lambda_j\Big(\sum_{i\epsilon V_C(j)} a_{ij}x_{ij}-c_j\Big)$$

(6) $$x_{ij} = 1/(\mu_i+\lambda_j)^{1/2} \qquad \text{for } (i,j)\epsilon V$$

$$H_3(\underline{\mu},\underline{\lambda}) = 2\sum_{(i,j)\epsilon V} a_{ij}(\mu_i+\lambda_j)^{1/2} - \sum_{i=1}^{R} \mu_i r_i - \sum_{j=1}^{C} \lambda_i c_j.$$

The Hessian of each of the functions H_1, H_2, H_3 is negative definite (Luenberger, 1973), so each has extreme points which in turn provide the unique extreme points of f_1, f_2, and f_3 by using (4), (5), and (6).

4. DEVELOPING ITERATIVE PROCEDURES

4.1. Cyclic Coordinate Descent. Given an function $F(\underline{x})$ to optimize, one can sometimes employ an iterative descent procedure. Descent with respect to the coordinate x_i means that one minimizes F as a function of x_i leaving all other coordinates fixed. The cyclic coordinate descent algorithm minimizes F cyclically with respect to each coordinate variable (Luenberger, 1973). We derive iterative procedures based on cyclic coordinate descent for raking, maximum likelihood, and minimum chi-square.

4.2. Raking.

$$H(\underline{\mu},\underline{\lambda}) = \sum_{i=1}^{R} \mu_i r_i + \sum_{j=1}^{C} \lambda_j c_j - \sum_{(i,j)\epsilon V} a_{ij}e^{(\mu_i+\lambda_j)}$$

Taking partial derivatives we get:

$$\frac{\partial H}{\partial \mu_i} = r_i - e^{\mu_i} \sum_{j\varepsilon V_R(i)} a_{ij} e^{\lambda_j}$$

$$\frac{\partial H}{\partial \lambda_j} = c_j - e^{\lambda_j} \sum_{i\varepsilon V_C(j)} a_{ij} e^{\mu_i} \ .$$

Setting each equal to zero and solving respectively for μ_i and λ_j we obtain

$$\mu_i = \ln(r_i / \sum_{j\varepsilon V_R(i)} a_{ij} e^{\lambda_j})$$

$$\lambda_j = \ln(c_j / \sum_{i\varepsilon V_C(j)} a_{ij} e^{\mu_i}) \ .$$

<u>Iterative Algorithm for Raking</u>

1) Initialize $\mu_i^{(0)} = \lambda_j^{(0)} = 0$ and $k=0$.

2) $\mu_i^{(k+1)} = \ln(r_i / \sum_{j\varepsilon V_R(i)} a_{ij} e^{\lambda_j^{(k)}})$.

3) Repeat step 2) for $i=1,\ldots,R$.

4) $\lambda_j^{(k+1)} = \ln(c_j / \sum_{i\varepsilon V_C(i)} a_{ij} \ e^{\mu_i^{(k+1)}})$

5) Repeat step 4) for $j=1,\ldots,C$.

6) Terminate based on some criterion of convergence, or increment k and return to step 2).

This derivation of the raking algorithm was first demonstrated by Bigelow and Shapiro (1977), however, the authors assumed that all a_{ij} were greater than zero.

4.3. <u>Maximum Likelihood.</u>

$$H(\underline{\mu},\underline{\lambda}) = \sum_{(i,j)\varepsilon V} a_{ij} \ln(\mu_i + \lambda_j) - \sum_{i=1}^{R} \mu_i r_i - \sum_{j=1}^{C} \lambda_j c_j$$

Taking partial derivatives we get

$$\frac{\partial H}{\partial \mu_i} = \sum_{j\varepsilon V_R(i)} (a_{ij}/(\mu_i + \lambda_j)) - r_i$$

$$\frac{\partial H}{\partial \lambda_j} = \sum_{i\varepsilon V_C(j)} (a_{ij}/(\mu_i + \lambda_j)) - c_j \ .$$

Setting each equal to zero, the objective is to find the unique μ_i and λ_j that are zeros of the respective functions:

$$F(\mu_i) = \sum_{j\varepsilon V_R(i)} (a_{ij}/(\mu_i + \lambda_j)) - r_i$$

$$G(\lambda_j) = \sum_{i\varepsilon V_C(j)} (a_{ij}/(\mu_i + \lambda_j)) - c_j \ .$$

In contrast to the situation with raking, zeros of $F(\mu_i)$ and $G(\lambda_j)$ cannot be found in closed form. Let us assume we can find the zeros for these functions, to yield:

$\mu_i^{(k+1)}$ is the zero of $F(\mu_i) =$

$$\sum_{\varepsilon V_R(i)} a_{ij}/(\mu_i + \lambda_j^{(k)}) - r_i$$

$\lambda_j^{(k+1)}$ is the zero of $G(\lambda_j) =$

$$\sum_{i\varepsilon V_C(j)} a_{ij}/(\mu_i^{(k+1)} + \lambda_j) - c_j ,$$

and hence an algorithm for maximum likelihood similar to the one for raking. To find the unique zeros of $F(\mu_i)$ and $G(\lambda_j)$, we use Newton's method. We employ a single iterative step of Newton's method within iterative of cyclic coordinate descent, and the composite algorithm is below. Due to limitations of space, we will not present the details of deriving that procedure here, however, they are explicitly worked out in (Fagan and Greenberg, 1985).

<u>Iterative Algorithm for Maximum Likelihood</u>

1) Initialize $\mu_i^{(0)} = \lambda_j^{(0)} = .5$ and $k=0$.

2) $\mu_i^{(k+1)} = \mu_i^{(k)} + (\sum_{j\varepsilon V_R(i)} (a_{ij}/(\mu_i^{(k)} + \lambda_j^{(k)})) - r_i) /$

$$\sum_{j\varepsilon V_R(i)} a_{ij}/(\mu_i^{(k)} + \lambda_j^{(k)})^2 \ .$$

2') Let $b = \max_{j\varepsilon V_R(i)} \{-\lambda_j^{(k)}\}$. If $\mu_i^{(k+1)} - b \leq 0$

set $\mu_i^{(k)} = (\mu_i^{(k)} + b)/2$ and go to 2).

3) Repeat steps 2) and 2') for $i=1,\ldots,R$.

4) $\lambda_j^{(k+1)} = \lambda_j^{(k)} + (\sum_{i\varepsilon V_C(j)} (a_{ij}/(\lambda_i^{(k+1)} + \lambda_j^{(k)})) - c_j) /$

$$\sum_{i\varepsilon V_C(j)} a_{ij}/(\mu_i^{(k+1)} + \lambda_j^{(k)})^2 .$$

4') Let $a = \max_{i\varepsilon V_C(j)} \{-\mu_i^{(k+1)}\}$. If $\lambda_j^{(k+1)} - a \leq 0$

set $\lambda_j^{(k)} = (\lambda_j^{(k)} + a)/2$ and go to 4).

5) Repeat steps 4) and 4') for $j=1,\ldots,C$.

6) Terminate based on some criterion, or increment k and return to step 2)

4.4. <u>Minimum Chi-Square.</u>

$$H(\underline{\mu},\underline{\lambda}) = 2 \sum_{i,j)\varepsilon V} a_{ij} (\mu_i + \lambda_j)^{1/2} - \sum_{i=1}^{R} \mu_i r_i - \sum_{j=1}^{C} \lambda_j c_j$$

We proceed exactly as in maximum likelihood, and all the earlier remarks pertain; in particular, we embed Newton's method within cyclic coordinate descent. Full details are worked out in (Fagan and Greenberg, 1985).

Iterative Algorithm for Minimum Chi-Square

1) Initialize $\mu_i^{(0)} = \lambda_j^{(0)} = .5$ and $k=0$.

2) $$\mu_i^{(k+1)} = \mu_i^{(k)} + 2\Big(\sum_{j\varepsilon V_R(i)} a_{ij}\,(\mu_i^{(k)}+\lambda_j^{(k)})^{-1/2} - r_i\Big) \Big/ \sum_{j\varepsilon V_R(i)} a_{ij}(\mu_i^{(k)}+\lambda_j^{(k)})^{-3/2}$$

2') Let $b = \max_{j\varepsilon V_R(i)} \{-\lambda_j^{(k)}\}$. If $\mu_i^{(k+1)} - b \leq 0$ set $\mu_i^{(k)} = (\mu_i^{(k)}+b)/2$ and go to 2).

3) Repeat steps 2) and 2') for $i=1,\ldots,R$.

4) $$\lambda_j^{(k+1)} = \lambda_j^{(k)} + 2\Big(\sum_{j\varepsilon V_C(j)} a_{ij}(\mu_i^{(k+1)}+\lambda_j^{(k)})^{1/2} - c_j\Big) \Big/ \sum_{i\varepsilon V_C(j)} a_{ij}(\mu_i^{(k+1)}+\lambda_j^{(k)})^{-3/2}$$

4') Let $a = \max_{i\varepsilon V_R(j)} \{-\mu_i^{(k+1)}\}$. If $\lambda_j^{(k+1)} - a \leq 0$ set $\lambda_j^{(k)} = (\lambda_j^{(k)}+a)/2$ and go to 4).

5) Repeat steps 4) and 4') for $j=1,\ldots,C$.

6) Terminate based on some criterion, or increment k and return to step 2).

5. CONCLUDING REMARKS

Let $\mathbf{A} = \{(a_{ij}),\mathbf{r},\mathbf{c}\}$ be the the following table:

0	1	2	3	4	4
1	4	5	6	7	5
0	0	0	1	2	2
3	6	7	8	9	5
4	7	8	9	10	5
3	4	4	5	5	21.

The maximum likelihood adjusted table is

0	1	1	1	1	4
1	1	1	1	1	5
0	0	0	1	1	2
1	1	1	1	1	5
1	1	1	1	1	5
3	4	4	5	5	21,

the table obtained through raking is

.0	.624	.949	1.208	1.219	4
.594	1.168	1.110	1.130	.998	5
.0	.0	.0	.796	1.204	2
1.131	1.112	.987	.956	.814	5
1.275	1.097	.953	.910	.765	5
3	4	4	5	5	21,

and the table using minimum chi-square is

.0	1.360	1.007	.758	.875	4
1.426	.758	.894	.915	1.007	5
.0	.0	.0	1.183	.817	2
.806	.934	1.048	1.066	1.146	5
.768	.948	1.051	1.078	1.155	5
3	4	4	5	5	21.

As can be seen, adjusted tables are quite different, and when the need to adjust a table to additivity arises, care must be exercised in selecting the most appropriate procedure.

When deciding upon methods, one must ask why the table is not additive in the first place. For example, if table entries were obtained through a survey, some factors can be:

a) Sampling considerations,
b) Coverage problems,
c) Non-response and classification errors,
d) Errors induced by earlier processing.

Sources of non-additivity should be investigated, and to the extent possible, corrected for before using any of the procedures discussed. After survey specific adjustments are made based on causes of non-additivity, one uses general adjustment procedures as described here. We are not suggesting any one of these procedures is superior to the others, and most likely, the best procedure for a given application will depend on the application itself and model assumptions. By having algorithms to adjust tables to the maximum likelihood and chi-square criteria, one has the option of using the procedure of choice and studies can be conducted comparing all three. One should attempt to select an adjustment strategy focusing on subject-based information, uses to which the adjusted tables will be put, and analytic needs.

Code has been developed to implement the algorithms discussed here in two and three dimensions. Raking and maximum likelihood seem to take about the same running time and minimum chi-square takes more time than the others.

ACKNOWLEDGMENT

The authors thank Chip Alexander and Ed Robinson for their careful reading of an earlier version of this paper and their thoughtful comments. They also thank Bev Causey and John Thompson for helpful discussions when developing the basic ideas.

REFERENCES

Alexander, C.H. (1987). A Class of Methods for Using Person Controls in Household Weighting. Survey Methodology. **13**, 183-198.

Bigelow, J. and Shapiro, N. (1977). A Scaling Theorem and Algorithms. SIAM Journal of Applied Mathematics. **33**, 348-352.

Bishop, Y., Fienberg, S., and Holland, P. (1975). Discrete Multivariate Analysis: Theory and Practice. MIT Press, Cambridge, Mass.

Causey, B. (1983). Estimation of Proportions for Multinomial Contingency Tables Subject to Marginal Constraints. Communications in Statistics (Ser. A). **12**, 2581-2587.

Causey, B. (1984). Estimation Under Generalized Sampling of Cell Proportions for Contingency Tables Subject to Marginal Constraints. Communications in Statistics (Ser. A). **13**, 2487-2494.

Copeland, K.R., Peitzmeier, F.K., Hoy, C.E. (1987). An Alternative Method for Controlling Current Population Survey Estimates to Population Counts. Survey Methodology. **13**, 173-181.

Darroch, J.M. and Ratcliff, D. (1972). Generalized Iterative Scaling for Log-Linear Models. The Annals of Mathematical Statistics. **43**, 1470-1480.

Deming W.E. and Stephan F.F. (1940). On a Least Squares Adjustment of a Sampled Frequency Table When the Expected Marginal Totals are Known. The Annals of Mathematical Statistics. **11**, 427-444.

Fagan, J. and Greenberg, B. (1985). Algorithms for Making Tables Additive; Raking, Maximum Likelihood, and Minimum Chi-Square. Statistical Research Division Report Series, Census/SRD/RR-85/12, Bureau of the Census, Statistical Research Division, Washington, D.C.

Fagan, J. and Greenberg, B. (1987). Making Tables Additive in the Presence of Zeros. American Journal of Mathematical and Management Sciences. **7**, 359-383. An earlier version of this paper appeared in: (1984). Proceedings of the Section on Survey Research Methods, American Statistical Society, Washington, D.C., 195-200.

Fienberg, S.E. (1970). An Iterative Procedure for Estimation in Contingency Tables. Annals of Mathematical Statistics. **41**, 907-917.

Fan, M.C., Woltman, H.F., Miskura, S.M., and Thompson, J.H. (1981). 1980 Census variance estimation procedure. Proceedings of the Section on Survey Research Methods, American Statistical Association, Washington, D.C., 176-181.

Ireland, C.K. and Kullback, S. (1968). Contingency Tables With Given Marginals. Biometrika. **55**, 179-188.

Kim, J., Thompson, J.H., Woltman, H.F., and Vajs, S.M. (1981). Empirical results from the 1980 Census sample estimation study. Proceedings of the Section on Survey Research Methods, American Statistical Association, Washington, D.C., 170-175.

Kullback, S. (1959). Information Theory and Statistics. Wiley, New York.

Lemaitre, G. and Dufour, J. (1987). An Integrated Method for Weighting Persons and Families. Survey Methodology. **13**, 199-207.

Luenberger, D. (1973). Introduction to Linear and Nonlinear Programming (2nd Ed., 1984). Addison-Wesley, Reading, MA.

Oh, H.L. and Scheuren, F.J. (1978). Multivariate Raking Ratio Estimation in the 1973 Exact Match Study. Proceedings of the Section on Survey Research Methods, American Statistical Association, Washington, D.C., 716-722.

Oh, H.L. and Scheuren, F.J. (1987). Modified Raking Ratio Estimation. Survey Methodology. **13**, 209-219.

Savage, I.R. (1973). Incomplete Contingency Tables: Conditions for the Existence of Unique MLE. In Mathematics and Statistics, Essays in Honor of Harold Bergstrom, edited by P. Jagers and L. Rade. Chalmers Institute of Technology, Goteborg, Sweden, 87-99.

Sinkhorn, R. (1967). Diagonal Equivalence to Matrices with Prescribed Row and Column Sums. American Mathematical Monthly. **74**, 402-405.

Smith, J.H. (1947). Estimation of Linear Functions of Cell Proportions. Annals of Mathematical Statistics. **18**, 231-254.

Stephan, F.F. (1942). An Iterative Method of Adjusting Sample Frequency Tables When Expected Marginal Totals are Known. The Annals of Mathematical Statistics. **13**, 166-178.

Thompson, J. (1981). Convergence Properties of the Iterative 1980 Census Estimator. Proceedings of the Section on Survey Research Methods, American Statistical Association, Washington, D.C., 182-185.

Woltman, H.F., Miskura, S.M., Thompson, J.H., and Bounpane, P.A. (1981). 1980 Census Weighting and Variance Studies: Design and Methodology. Proceedings of the Section on Survey Research Methods, American Statistical Association, Washington, D.C., 164-169.

1/ This paper reports the general results of research undertaken by Census Bureau staff. The views expressed are attributable to the authors and do not reflect those of the Census Bureau.