BUREAU OF THE CENSUS
STATISTICAL RESEARCH DIVISION REPORT SERIES
SRD Research Report Number: Census/SRD/RR-90/14

Minimizing λ-*Measures* for Table Additivity in Three Dimensions

by

James T. Fagan
Brian V. Greenberg
Statistical Research Divison
Bureau of the Census
Washington, D.C. 20233

Report completed: October 14, 1990

Report issued: October 14, 1990

Making Tables Additive in Three Dimensions Under λ-Measures

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

Given a contingency table of non-negative numbers in which the internal entries do not sum to the corresponding marginals, there is often a need to adjust internal entries to achieve additivity. Tables that can be so adjusted while maintaining the zero structure are termed feasible. In earlier work, the authors showed how to determine whether a given two-dimensional table is feasible. They also provided algorithms to adjust feasible tables and showed that the algorithms converged to minimize measures of closeness corresponding to raking, maximum likelihood, and minimum Chi-Square. In this paper the authors extend those results to three dimensions and to the one parameter, power-divergence family of goodness-of-fit statistics, referred to as λ-measures, which has been introduced by Cressie and Read.

Key Words and Phrases: contingency tables, power-divergence statistic, goodness-of-fit, 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 revised 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 that will determine whether a given table can be so adjusted, and such adjustable tables were called feasible. In Section 4 of this report we present comparable procedures for three-dimensional tables.

In general, given a feasible table, one seeks a derived table which is close. The notion of "close" is not unique, and for every criterion of closeness a different dervied table may be obtained. Four of the most cited criteria of closeness are: (a) Raking, (b) Maximum Likelihood, (c) Minimum Chi-Square, and (d) Weighted Least Squares. In an earlier paper, Fagan and Greenberg (1988), the authors provide algorithms which, when applied to a feasible table, converge to a revised table optimizing the respective measure of closeness for (a)-(c). Since an optimum revised table for weighted least squares can be solved exactly in closed form, that objective function was not treated in detail in the earlier paper.

In that paper each measure of closeness was couched as a non-linear function to be minimized subject to linear marginal constraints. Starting with the primal (original) objective function we formed the dual which we maximized. Maximizing the dual function is an optimization problem amenable to iterative coordinate descent methods. These techniques yielded iterative algorithms converging to a solution of the dual problems and subsequently to the original.

In this paper we extend findings to encompass the goodness-of-fit measures defined by the power-divergence statistic. This one parameter family of statistics was introduced by Read and Cressie (1988) and for specific values of the parameter, one obtains each of the objective functions (a)-(d) above. We use techniques similar to those employed earlier to derive algorithms which converge to best fit tables for the power-divergence statistics.

In Section 2 we introduce the power-divergence statistic, show how it relates to the earlier goodness-of-fit measures and formalize the objective functions to be minimized. In Section 3 we set up the dual function to be optimized, employ cyclic coordinate descent to derive algorithms, and provide a few examples and summary remarks. In section 4 we define feasibility for three-dimensional tables, provide examples, and derive a test for feasibility.

Tables are adjusted 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); Thompson (1981); Kim, Thompson, Woltman and Vajs (1981) and Woltman, Miskura, Thompson, and Bounpane (1981). Five recent papers relating to table adjustment for estimation and weighting are: Copeland, Peitzmeier, and Hoy (1987); Alexander (1987 and 1990); Lemaitre and Dufour (1987); and Oh and Scheuren (1987). Additional information and bibliography in table adjustment is contained in Fagan and Greenberg (1988).

An abreviated version of this report was presented at the 1990 Annual Meetings of the American Statistical Association and will appear in the proceedings of that meeting (Fagan and Greenberg 1990).

1.2 <u>Feasible Tables</u>. By a <u>table</u> we mean a triple $\mathbf{A} = \{(a_{ij}),r,c\}$ 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 <u>additive</u> if

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

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

That table $\mathbf{A}$ is said to be <u>feasible</u> 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 B is <u>derived</u> from **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:

relating to table adjustment for estimation and weighting are: Copeland, Peitzmeier, and Hoy (1987); Alexander (1987 and 1990); Lemaitre and Dufour (1987); and Oh and Scheuren (1987). Additional information and bibliography in table adjustment is contained in Fagan and Greenberg (1988).

An abreviated version of this report was presented at the 1990 Annual Meetings of the American Statistical Association and will appear in the proceedings of that meeting (Fagan and Greenberg 1990).

1.2 Feasible Tables. By a table we mean a triple $\mathbf{A} = \{(a_{ij}),r,c\}$ 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

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

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

That 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 B is derived from $\mathbf{A}$. That is, $\mathbf{A}$ is feasible if and only if there exists an RxC matrix (x_{ij}) such that $(b_{ij}) = (x_{ij}a_{ij})$, satisfying:

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

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

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

where $\qquad V = \{(i,j)|(i,j)\varepsilon RxC \text{ and } a_{ij}\neq 0\}.$

2. DERIVING TABLES OPTIMIZING THE POWER-DIVERGENCE STATISTICS

2.1 Criteria for Optimal Derived Tables. Given a feasible table **A**, one seeks a derived additive table **B** "close" to **A**. In Fagan and Greenberg (1988) we discussed four measures of closeness:

$$(m_1): \quad \sum_{(i,j)\varepsilon V} b_{ij}\ell n(b_{ij}/a_{ij})$$

$$(m_2): \quad \sum_{(i,j)\varepsilon V} -a_{ij}\ell n(b_{ij}/a_{ij})$$

$$(m_3): \quad \sum_{(i,j)\varepsilon V} (a_{ij} - b_{ij})^2/b_{ij}$$

$$(m_4): \quad \sum_{(i,j)\varepsilon V} (a_{ij} - b_{ij})^2/a_{ij} ,$$

which are the objective functions subject to constraints (1)-(3) for, respectively, raking, maximum likelihood, minimum Chi-Square, and weighted least squares. Background for these particular functions is discussed in Fagan and Greenberg (1988). Each of these functions can be used as a goodness-of-fit statistics to observe how closely an observed distribution resembles an assumed distribution. Our use of these goodness-of-fit measures is somewhat different. Given a non-additive table **A**, find the closest additive table based on each goodness-of-fit measure. In that paper, we presented algorithms which can be used on an arbitrary non-additive table, which may have zero cells, to obtain a derived table for each measure of goodness-of-fit. We replace b_{ij} by $a_{ij}x_{ij}$, and rewrite the expressions above as

$$(g_1):\quad \sum_{(i,j)\varepsilon V} a_{ij}x_{ij}\ln x_{ij}$$

$$(g_2):\quad \sum_{(i,j)\varepsilon V} -a_{ij}\ln x_{ij}$$

$$(g_3):\quad \sum_{(i,j)\varepsilon V} a_{ij}x_{ij}(x_{ij}^{-1}-1)^2$$

$$(g_4):\quad \sum_{(i,j)\varepsilon V} a_{ij}(x_{ij}-1)^2 \ .$$

In Read and Cressie (1984), the authors present a generalized, one-parameter family goodness-of-fit measure -- the power-divergence statistic -- which we write as:

$$d_\alpha(\mathbf{A},\mathbf{B}) = [2/\alpha(\alpha+1)] \sum_{(i,j)\varepsilon V} a_{ij}[(a_{ij}/b_{ij})^\alpha-1]$$

for $\alpha\neq 0,-1$. It is not hard to see that d_1 equals the measure m_3 and d_{-2} equals m_4 (assuming, without loss of generality, that

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

Letting $x_{ij} = b_{ij}/a_{ij}$ we write the power-divergence statistic as:

$$f_\alpha(x) = [2/\alpha(\alpha+1)] \sum_{(i,j)\varepsilon V} a_{ij}(x_{ij}^{-\alpha}-1) \ .$$

We define

$$\begin{aligned} f_0(\underline{x}) &= \lim_{\alpha\to 0} f_\alpha(\underline{x}) \\ &= \lim_{\alpha\to 0} [2/\alpha(\alpha+1)] \sum_{(i,j)\varepsilon V} a_{ij}(x_{ij}^{-\alpha}-1) \\ &= \lim_{\alpha\to 0} \Big[2 \sum_{(i,j)\varepsilon V} a_{ij}(-\ln x_{ij})x_{ij}^{-\alpha}\Big]/(2\alpha+1) \\ &= -2 \sum_{(i,j)\varepsilon V} a_{ij}\ln x_{ij} \end{aligned}$$

using ℓ'Hôpital's Rule at step 3. The last expression on the right is twice the maximum likelihood measure m_2. We also define

$$\begin{aligned} f_{-1}(\underline{x}) &= \lim_{\alpha\to-1} f_\alpha(\underline{x}) \\ &= \lim_{\alpha\to-1} [2/\alpha(\alpha+1)] \sum_{(i,j)\varepsilon V} a_{ij}(x_{ij}^{-\alpha}-1) \\ &= \lim_{\alpha\to-1} [2/\alpha(\alpha+1)] \sum_{(i,j)\varepsilon V} [a_{ij}x_{ij}(x_{ij}^{-(\alpha+1)}-1)+a_{ij}(x_{ij}-1)] \\ &= 2 \sum_{(i,j)\varepsilon V} a_{ij}x_{ij}\ell n x_{ij} \ , \end{aligned}$$

which is twice the raking measure m_1. Note use of the assumption that

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

in the third equality above. Measures f_0 and f_{-1} are treated in Fagan and Greenberg (1988), so we assume $\alpha\neq0,-1$ in this report.

Let S denote the region defined by the constraints (1)-(3). The Hessian of $f_\alpha(\underline{x})$

$$\nabla^2_{\underline{x}} f_\alpha(\underline{x}) = \text{diag } (2a_{ij}x_{ij}^{-(\alpha+2)})$$

is positve definite so f_α is a strictly convex function over S. The set S is a convex set so every local minimum of f_α over S is a global minimum and there is at most one.

Let T be the set of vectors satisfying (1), (2) and

$$x_{ij}\geq 0 \qquad (i,j)\varepsilon V$$

and let L be the boundry points of T, that is, L consists of vectors satisfying (1), (2) and

$$x_{ij} = 0 \qquad \text{for some} \qquad (i,j)\varepsilon V.$$

Every point of L is a limit point for S and f_α is continuous over S, so for $\underline{z}\varepsilon L$, we can define

$$f_\alpha(\underline{z}) = \lim_{\underline{x}_k \to \underline{z}} f_\alpha(\underline{x}_k)$$

where $\{\underline{x}_k\}_{k=1}^{\infty}$ is a sequence in S converging to $\underline{z}$. Hence, f_α is defined and continuous over all of T if we define

$$f_\alpha(\underline{x}) : T \to R \cup \{\infty\} .$$

Note that

$$f_\alpha(\underline{z}) = \begin{cases} \infty & \text{if} \quad \alpha > 0 \\ [1/\alpha(\alpha+1)] \sum_{(i,j)\varepsilon V} a_{ij}(z_{ij}^{-\alpha} - 1) & \text{if} \quad \alpha < 0. \end{cases}$$

The set T is closed and bounded and f_α is continuous, so f_α has a minimum over T.

Note that for all $(i,j)\varepsilon V$

$$[\nabla_{\underline{x}} f_\alpha(\underline{x})]_{(i,j)} = [(-2)/(\alpha+1)] a_{ij} x_{ij}^{-(\alpha+1)} .$$

As $x_{ij} \to 0$,

$$[\nabla_{\underline{x}} f_\alpha(\underline{x})]_{(i,j)} \to -\infty$$

for $\alpha > -1$. If $\underline{z}\varepsilon L$ and $\{\underline{x}_k\}_{k=1}^{\infty} \to \underline{z}$ for $\underline{x}_k \varepsilon S$, then for some $(i,j)\varepsilon V$,

$$\lim_{\underline{x}_k \to \underline{z}} [\nabla_{\underline{x}} f_\alpha(\underline{x}_k)]_{(i,j)} = -\infty .$$

For each $\underline{z}\varepsilon L$, there exists a sequence $\{x_k\}$ in S converging to $\underline{z}$. Since $\underline{z}\varepsilon L$, there exists an $(i_0,j_0)\varepsilon V$ such that $\underline{z}_{(i_0,j_0)} = 0$. Thus in each neighborhood of $\underline{z}$, f_α is decreasing in at least one direction; hence f_α cannot achieve a minimum over T at $\underline{z}\varepsilon L$. Thus, for $\alpha > -1$, f_α has its minimum over T in S. Since S is an open convex set and f_α is a convex function, there is a unique global minimum for f_α over S.

To find the global minimum of f_α over S, it suffices to use standard optimization techniques for a convex function with linear constraints. In the next section we form the Lagrangian, set up the dual function which we proceed to maximize, and finally interpret the results in the primal problem.

2.2 Forming the Dual Function

To solve the primal problem

$$(P_\alpha): \qquad \text{Minimize} \quad f_\alpha(\underline{x}) \quad \text{over } S,$$

we form the Lagrangian by incorporating conditions (1) and (2) into the primal to obtain

$$L_\alpha(\underline{x},\underline{\mu},\underline{\lambda}) = f_\alpha(\underline{x}) + \sum_{i=1}^{R} \mu_i \Big(\sum_{(i,j)\varepsilon V} a_{ij}x_{ij} - r_i \Big)$$

$$+ \sum_{j=1}^{C} \lambda_j \Big(\sum_{(i,j)\varepsilon V} a_{ij}x_{ij} - c_j \Big) \,.$$

We minimize $L_\alpha(\underline{x},\underline{\mu},\underline{\lambda})$ as a function of $\underline{x}$, $\underline{\mu}$, and $\underline{\lambda}$ and solve for critical $\underline{x}$ values in terms of $\underline{\mu}$ and $\underline{\lambda}$ which we replace in $L_\alpha(\underline{x},\underline{\mu},\underline{\lambda})$ resulting in the dual function:

$$H_\alpha(\underline{\mu},\underline{\lambda}) = \min_{\underline{x}>0} \{L_\alpha(\underline{x},\underline{\mu},\underline{\lambda})\} \,.$$

Note that $H_\alpha(\underline{\mu},\underline{\lambda})$ is a function of $\underline{\mu}$ and $\underline{\lambda}$ which we maximize, thus solving the dual problem. The maximum of $H_\alpha(\underline{\mu},\underline{\lambda})$ equals the minimum of the corresponding $f_\alpha(\underline{x})$ constrained by (1) and (2). Adding the condition that $\underline{x}>0$ in terms of $\underline{\mu}$ and $\underline{\lambda}$ when maximizing $H_\alpha(\underline{\mu},\underline{\lambda})$ yields the value of $\underline{x}$ that minimizes f_α over S.

To find the minimum of $L_\alpha(\underline{x},\underline{\mu},\underline{\lambda})$ subject to $\underline{x}>0$, for each $(i,j)\varepsilon V$ we form

$$\frac{\partial L_\alpha}{\partial x_{ij}} = [-2/(\alpha+1)]a_{ij}x_{ij}^{-(\alpha+1)}+a_{ij}(\mu_i+\lambda_j) \ .$$

Setting this expression to zero yields

$$x_{ij}^{-(\alpha+1)} = [(\alpha+1)/2](\mu_i+\lambda_j) \ .$$

Since $x_{ij}>0$ we have

$$[(\alpha+1)/2](\mu_i+\lambda_j)>0 \ ,$$

and

$$x_{ij} = [[(\alpha+1)/2](\mu_i+\lambda_j)]^{-1/(\alpha+1)} \ .$$

Replacing these values in $L_\alpha(\underline{x},\underline{\mu},\underline{\lambda})$ for x_{ij} and simplifying yields:

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

Our objective is to solve the Dual Problem

(D_α): Maximize $H_\alpha(\underline{\mu},\underline{\lambda})$ subject to $[(\alpha+1)/2](\mu_i+\lambda_j)>0$.

Note that the function $H_\alpha(\underline{\mu},\underline{\lambda})$ is concave since P_α is a convex problem and the set

$$W = \{(\underline{\mu},\underline{\lambda}): \ [(\alpha+1)/2](\mu_i+\lambda_j)>0 \quad \text{for all} \quad (i,j)\varepsilon V\}$$

is a convex set. Thus, any local maximum of H_α is a global maximum and a local maximum of H_α does exist whenever f_α has a minimum. In fact, if $\underline{x}^*$ is the minimum of f_α over S, then there exist $(\underline{\mu}^*,\underline{\lambda}^*)$ in W such that $(\underline{\mu}^*,\underline{\lambda}^*)$ maximizes $H_\alpha(\underline{\mu},\underline{\lambda})$, where for all $(i,j)\varepsilon V$

$$x^*_{ij} = [[(\alpha+1)/2](\mu^*_i+\lambda^*_j)]^{-1/(\alpha+1)}>0 .$$

That is,

$$(\underline{\mu}^*,\underline{\lambda}^*) \text{ solves } D_\alpha \text{ if and only if } \underline{x}^* \text{ solves } P_\alpha .$$

Our objective in the next section is to find points $(\underline{\mu}^*,\underline{\lambda}^*)$ to solve D_α .

3. DEVELOPING ITERATIVE PROCEDURES

3.1 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 (1984). The function F is minimized with respect to x_1 first and then with respect to x_2 and so forth through x_n. We derive an iterative procedure based on cyclic coordinate descent to maximize $H_\alpha(\underline{\mu},\underline{\lambda})$ over W .

We begin by taking partial derivitives:

$$\frac{\partial H\alpha}{\partial \mu_i} = \sum_{(i,j)\varepsilon V} a_{ij}[((\alpha+1)/2)(\mu_i+\lambda_j)]^{-1/(\alpha+1)}-r_i \qquad \text{for } i=1,\ldots,R$$

$$\frac{\partial H\alpha}{\partial \lambda_j} = \sum_{(i,j)\varepsilon V} a_{ij}[((\alpha+1)/2)(\mu_i+\lambda_j)]^{-1/(\alpha+1)}-c_j \qquad \text{for } j=1,\ldots,C.$$

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

$$\frac{\partial H\alpha}{\partial \mu_i}(\mu_i) \quad \text{and} \quad \frac{\partial H\alpha}{\partial \lambda_j}(\lambda_j) \;.$$

Our iterative procedure to find $(\underline{\mu}^*,\underline{\lambda}^*)$ to maximize $H_\alpha(\underline{\mu},\underline{\lambda})$ over W is (in principle) as follows. Initialize $\mu_i^{(0)}$ and $\lambda_j^{(0)}$, find $\mu_i^{(k+1)}$ as a function of $\lambda_j^{(k)}$, and find $\lambda_j^{(k+1)}$ as a function of $\mu_i^{(k+1)}$. In particular, we let $\mu_i^{(k+1)}$ be the unique zero of

$$\frac{\partial H\alpha}{\partial \mu_i}(\mu_i) = \sum_{(i,j)\varepsilon V} a_{ij}[[(\alpha+1)/2)](\mu_i+\lambda_j^{(k)})]^{-1/(\alpha+1)} - r_i$$

such that $[(\alpha+1)/2][\mu_i^{(k+1)}+\lambda_j^{(k)}]>0$ and let $\lambda_k^{(k+1)}$ be the unique zero of

$$\frac{\partial H\alpha}{\partial \lambda_j}(\lambda_j) = \sum_{(i,j)\varepsilon V} a_{ij}[((\alpha+1)/2)(\mu_i^{(k+1)}+\lambda_j)]^{-1/(\alpha+1)} - c_j \;,$$

such that $[[(\alpha+1)/2][\mu_i^{(k+1)}+\lambda_j^{(k+1)}]>0$. The sequence of vector pairs $(\underline{\mu}^{(k)},\underline{\lambda}^{(k)})$ will converge to a vector pair $(\underline{\mu}^*,\underline{\lambda}^*)$ such that $H_\alpha(\underline{\mu}^*,\underline{\lambda}^*)$ is maximum (subject to $[(\alpha+1)/2](\mu_i^*+\lambda_j^*)>0$) and hence such that if

$$x_{ij}^* = [[(\alpha+1)/2](\mu_i^* + \lambda_j^*)]^{-1/(\alpha+1)} \;,$$

then $\underline{x}^*$ minimizes $f_\alpha(\underline{x})$ over S. That is, the solution of the dual problem, D_α, is used to obtain the solution of the primal problem, P_α.

Details of cyclic coordinate descent are discussed in Luenberger (1984, p. 228) and as applied to table adjustment problems in Fagan and Greenberg (1985). To find the unique zeros of

$$\frac{\partial H\alpha}{\partial \mu_i}(\mu_i) \quad \text{and} \quad \frac{\partial H\alpha}{\partial \lambda_j}(\lambda_j)$$

we use Newton's method within each iteration of cyclic coordinate descent and the composite algorithm is below. We will not present the details of the derivation here, but they follow closely along the lines presented in Fagan and Greenberg (1985).

3.2 Iterative Procedure to Maximize $H_\alpha(\mu,\lambda)$ for $\alpha\neq 0,-1$

1) Initialize $\mu_i^{(0)} = \lambda_j^{(0)} = 1/(\alpha+1)$ and set $k=0$

2) $$\mu_i^{(k+1)} = \mu_i^{(k)} + \frac{2\left(\sum_{(i,j)\varepsilon V} a_{ij}[[(\alpha+1)/2](\mu_i^{(k)}+\lambda_j^{(k)})]^{-1/(\alpha+1)} - r_i\right)}{\sum_{(i,j)\varepsilon V} a_{ij}[[(\alpha+1)/2]\mu_i^{(k)}+\lambda_j^{(k)}]^{-(\alpha+2)/(\alpha+1)}}$$

2') Let $\lambda = \underset{(i,j)\varepsilon V}{\mathrm{Max}} \{-[(\alpha+1)/2]\lambda_j^{(k)}\}$.

If $[(\alpha+1)/2]\mu_i^{(k+1)}-\lambda\leq 0$, set $\mu_i^{(k)} = [\mu_i^{(k)}+ 2\lambda/(\alpha+1)]/2$ and go to 2).

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

4) $$\lambda_j^{(k+1)} = \lambda_j^{(k)} + \frac{2\left(\sum_{(i,j)\varepsilon V} a_{ij}[[(\alpha+1)/2](\mu_i^{(k+1)}+\lambda_j^{(k)})]^{-1/(\alpha+1)} - c_j\right)}{\sum_{(i,j)\varepsilon V} a_{ij}[[(\alpha+1)/2](\mu_i^{(k+1)}-\lambda_j^{(k)})]^{-(\alpha+2)/(\alpha+1)}}$$

4') Let $\mu = \underset{(i,j)\varepsilon V}{\mathrm{Max}} \{-[(\alpha+1)/2]\mu_i^{(k+1)}\}$.

If $[(\alpha+1)/2]\lambda_j^{(k+1)}-\mu\leq 0$ set $\lambda_j^{(k)} = [\lambda_j^{(k)}+2\mu/(\alpha+1)]/2$ and go to 4).

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

6) Increment k and return to step 2) else terminate if:

(a.) the sequence of values $\mu_i(k)$ and $\lambda_j(k)$ converges for all i and j

(b.) the sequence of values $\mu_i(k)$ or $\lambda_j(k)$ gets too large or too close to zero (c.) the program begins to oscilate between steps 2) and 2') or 4) and 4')

(d.) the number of iterations becomes excessively large.

When terminating for criterion (a) above, the values $\mu_i(k)$ and $\lambda_j(k)$ will converge to μ_i^* and λ_j^*, and

$$x_{ij}^* = [[(\alpha+1)/2)(\mu_i^* + \lambda_j^*)]^{-1/(\alpha+1)}$$

for $(i,j)\varepsilon V$ will minimize f_α over S. There will not be an optimal over S if one must terminate for conditions (b), (c) or (d). Under these conditions one typically has an optimal on the bound y, L, and this does not tell us very much. The algorithm will converge for all $\alpha>-1$.

3.3 Examples

In Fagan and Greenberg (1988) the authors introduced Table 1 (below) and found the adjusted tables under raking, maximum likelihood, and minimal Chi-Square, (corresponding to f_α for $\alpha = -1,0,1$, respectively). We now discuss the adjusted tables based on Table 1 for various other α .

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

Table 1

(a) For $\alpha = -4$ the solution appears to be on the boundary of S and we cannot find it using the algorithm above. We terminate the algorithm for this example when $\alpha = -4$ for reason (c) above. The algorithm oscilated between 4) and 4').

(b) For $\alpha=-3$, the adjusted table is in Table 2:

0	.431	.817	1.201	1.551	4
.408	1.034	1.097	1.221	1.241	5
0	0	0	.672	1.328	2
1.122	1.209	1.036	.985	.649	5
1.471	1.327	1.550	.922	.231	5
3	4	4	5	5	21

Table 2

(c) For α= 2/3, the adjusted table is below

0	1.275	.998	.822	.906	4
1.318	.816	.924	.936	1.006	5
0	0	0	1.136	.864	2
.857	.949	1.037	1.048	1.108	5
.824	.960	1.041	1.559	1.116	5
3	4	4	5	5	21

Table 3

3.4 Discussion

(a.) This algorithm will converge to a solution of D_α and hence of P_α for arbitrary α and arbitrary table **A** if the function $f_\alpha(\underline{x})$ has a minimum at a positive $\underline{x}^*$.

(b.) Algorithm steps 2') and 4') ensure that the solution remains positive, that is,

$$[(\alpha+1)/2][\underline{\mu}^*+\lambda^*]>0.$$

(c.) For $\alpha>-1$, for every table **A**, the function $f_\alpha(\underline{x})$ has a minimum at a positive $\underline{x}^*$, so this algorithm will find it.

(d.) For an arbitrary α and arbitrary table **A**, if $f_\alpha(\underline{x})$ has a positive minimum at $\underline{x}^*$, then $f_\sigma(\underline{x})$ will have a positive minimum at some $\underline{y}^*$ for all σ in a neigborhood of α . In fact, $\underline{y}^*$ will be a continuously differentiable function of α .

(e.) Read and Cressie (1988) remark that they favor α= 2/3 as a desirable measure of goodness-of-fit. Note that for α= 2/3, all feasible tables have a solution.

4. THREE-DIMENSIONAL TABLES

The preceeding sections of this report were couched in terms of two-dimensional tables as were our earlier reports on this topic Fagan and Greenberg (1984,1985 and 1988). Virtually all procedures and algorithms that can be applied to two-dimensional tables also can be applied to tables of higher dimension after minor modifications. In particular, the problem set-up and algorithms in Sections 2 and 3 have virtual identical counterparts in three-dimensions for feasible tables.

The definition for table feasibility also goes over to three-dimensions (and higher) and procedures to determine if a three-dimensional table is feasible are similar to those for two-dimensions; see Fagan and Greenberg (1985). The only exception to this rule is that in the earlier work one sets up a linear programming problem which has the structure of a transportation problem, see Luenberger (1984). In three-dimensions, one does not have the corresponding transportation problem, so one must stick with the more general linear programming problem throughout. With that understanding, if the linear programming problem has a solution, Lemmas 1, 2 and 3 and Theorem 1 in Fagan and Greenberg (1987) hold completely in three-dimensional tables. Accordingly, one can apply the corresponding iterative procedures to determine whether an arbitrary three-dimensional table is feasible.

To be a little more specific, we define a three-dimensional contingency table as a four-tuple: $\mathbf{A} = \{(a_{ijk}),\underline{r},\underline{c},\underline{\ell}\}$ of arrays of non-negative reals where:

(a) (a_{ijk}) is an MxNxP matrix

(b) $\underline{r}$ is an MxN matrix

(c) $\underline{c}$ is an NxP matrix

(d) $\underline{\ell}$ is an MxP matrix, and

$$\sum_{j=1}^{N} r_{ij} = \sum_{k=1}^{P} \ell_{ik} \qquad (i=1,\ldots,M),$$

$$\sum_{i=1}^{M} r_{ij} = \sum_{k=1}^{P} c_{jk} \qquad (j=1,\ldots,N),$$

$$\sum_{i=1}^{N} \ell_{ik} = \sum_{j=1}^{N} c_{jk} \qquad (k=1,\ldots,P) .$$

We say **A** is additive if

$$\sum_{k=1}^{P} a_{ijk} = r_{ij} \qquad \begin{array}{l} i=1,\ldots,M \\ j=1,\ldots,N \end{array}$$

$$\sum_{i=1}^{M} a_{ijk} = c_{jk} \qquad \begin{array}{l} j=1,\ldots,N \\ k=1,\ldots,P \end{array}$$

$$\sum_{j=1}^{N} a_{ijk} = \ell_{ik} \qquad \begin{array}{l} i=1,\ldots,M \\ k=1,\ldots,P \end{array} .$$

Table **A** is feasible if there exists an additive table **B** such that

$$b_{ijk} = a_{ijk}x_{ijk}$$

for $x_{ijk} \neq 0$ whenever $a_{ijk} \neq 0$, for all $(i=1,\ldots,M;\ j=1,\ldots,N;\ k=1,\ldots,P)$.

For simplicity we represent a 2x2x2 three-dimensional table by

$$\begin{array}{cc|c} a_{111} & a_{121} & a_{101} \\ a_{211} & a_{221} & a_{201} \\ \hline a_{011} & a_{021} & a_{001} \end{array} \qquad \begin{array}{cc|c} a_{112} & a_{122} & a_{102} \\ a_{212} & a_{222} & a_{202} \\ \hline a_{012} & a_{022} & a_{002} \end{array} \qquad \begin{array}{cc|c} a_{110} & a_{120} & a_{100} \\ a_{210} & a_{220} & a_{200} \\ \hline a_{010} & a_{020} & a_{000} \end{array}$$

Level 1 Level 2 Level 0

where Level 1 plus Level 2 add to the total level, Level 0.

Below, we see that Table 1 is feasible, with an additive counterpart in Table 1':

$$\begin{array}{cc|c} 1 & 1 & 2 \\ 1 & 1 & 1 \\ \hline 2 & 1 & 3 \end{array} \qquad \begin{array}{cc|c} 1 & 1 & 1 \\ 1 & 1 & 2 \\ \hline 1 & 2 & 3 \end{array} \qquad \begin{array}{cc|c} 2 & 1 & 3 \\ 1 & 2 & 3 \\ \hline 3 & 3 & 6 \end{array}$$

Level 1 Level 2 Level 0

Table 1

$$\begin{array}{cc|c} 3/2 & 1/2 & 2 \\ 1/2 & 1/2 & 1 \\ \hline 2 & 1 & 3 \end{array} \qquad \begin{array}{cc|c} 1/2 & 1/2 & 1 \\ 1/2 & 3/2 & 2 \\ \hline 1 & 2 & 3 \end{array} \qquad \begin{array}{cc|c} 2 & 1 & 3 \\ 1 & 2 & 3 \\ \hline 3 & 3 & 6 \end{array}$$

Level 1 Level 2 Level 0

Table 1'

Below we display Table 2 which is not feasible:

$$\begin{array}{cc|c} 1 & 1 & 2 \\ 1 & 1 & 1 \\ \hline 2 & 1 & 3 \end{array} \qquad \begin{array}{cc|c} 1 & 1 & 1 \\ 1 & 1 & 2 \\ \hline 1 & 2 & 3 \end{array} \qquad \begin{array}{cc|c} 1 & 2 & 3 \\ 2 & 1 & 3 \\ \hline 3 & 3 & 6 \end{array}$$

Level 1 Level 2 Level 0

Table 2.

For if Table 2' were additive then $b_{122}<1$ (being a summand of $b_{102} = 1$) so $b_{121}>1$ (because $b_{122} + b_{121} = b_{120} = 2$), but b_{121} is a summand of $b_{021} = 1$: a contradiction.

$$
\begin{array}{cc|c} b_{111} & b_{121} & 2 \\ b_{211} & b_{221} & 1 \\ \hline 2 & 1 & 3 \end{array}
\qquad
\begin{array}{cc|c} b_{122} & b_{122} & 1 \\ b_{212} & b_{222} & 2 \\ \hline 1 & 2 & 3 \end{array}
\qquad
\begin{array}{cc|c} 1 & 2 & 3 \\ 2 & 1 & 3 \\ \hline 3 & 3 & 6 \end{array}
$$

Level 1 Level 2 Level 0

Table 2'

This example exhibits a sharp distinction between two and three dimensions. In two dimensions, every table having all positive entries is feasible; whereas Table 2 is a non-feasible table in three dimensions with all entries positive. It is also interesting to observe that there is no non-negative additive table with marginals as shown in Table 3. This is in contrast to the fact that in two dimensions every table with positive marginals has at least one non-negative solution.

$$
\begin{array}{cc|c} x & x & 3 \\ x & x & 1 \\ \hline 1 & 3 & 4 \end{array}
\qquad
\begin{array}{cc|c} x & x & 1 \\ x & x & 3 \\ \hline 1 & 3 & 4 \end{array}
\qquad
\begin{array}{cc|c} 1 & 3 & 4 \\ 3 & 1 & 4 \\ \hline 4 & 4 & 8 \end{array}
$$

Table 3

For the sake of completeness, we provide the algorithm to determine whether an arbitrary three-dimensional table **A** is feasible. The definitions, Lemmas, Theorem, and algorithm are taken from Fagan and Greenberg (1987) and proofs for all assertions can be found there.

From table **A** we form a new table **M** = $\{(m_{ijk}),\underline{r},\underline{c},\underline{\ell}\}$ where

$$
m_{ijk} = \begin{cases} 0 & \text{if } a_{ijk} = 0 \\ 1 & \text{if } a_{ijk} \neq 0 . \end{cases}
$$

A is feasible if and only if **M** is feasible.

Given the table **M**, consider the following sequence of linear programming problems indexed by positive integers, q: minimize

$$(4.)\qquad C^q = \sum_{i=1}^{M} \sum_{j=1}^{N} \sum_{k=1}^{P} c^q_{ijk} x_{ijk}$$

subject to

$$(5.)\qquad \sum_{k=1}^{P} x_{ijk} = r_{ij} \qquad \begin{matrix} i=1,\ldots,M \\ j=1,\ldots,N \end{matrix}$$

$$(6.)\qquad \sum_{i=1}^{M} x_{ijk} = c_{jk} \qquad \begin{matrix} j=1,\ldots,N \\ k=1,\ldots,P \end{matrix}$$

$$(7.)\qquad \sum_{j=1}^{N} x_{ijk} = \ell_{ik} \qquad \begin{matrix} i=1,\ldots,M \\ k=1,\ldots,P \end{matrix}$$

$$(8.)\qquad x_{ijk} \geq 0 \qquad (i=1,\ldots,M; \quad j=1,\ldots,N; \quad k=1,\ldots,P),$$

where

$$c^1_{ijk} = \begin{cases} T & \text{if } m_{ijk} = 0 \\ 0 & \text{otherwise}, \end{cases} \qquad \text{and}$$

$$T = \sum_{i=1}^{M} \sum_{j=1}^{N} r_{ij} = \sum_{j=1}^{N} \sum_{k=1}^{P} c_{jk} = \sum_{i=1}^{M} \sum_{k=1}^{P} \ell_{ik}$$

and for q>1,

$$c^q_{ijk} = \begin{cases} 1 & \text{if } c^{q-1}_{ijk} = 1 \text{ or } x^{q-1}_{ijk} \neq 0 \text{ and } m_{ijk} \neq 0 \\ T & \text{if } m_{ijk} = 0 \\ 0 & \text{otherwise}, \end{cases}$$

where (x^q_{ijk}) minimizes (4) subject to (5)-(8). Let us denote by U the set of X_{ijk} such that conditions (5)-(8) are satisfied. If U is empty, then M is not feasible.

Lemma 1: If $U \neq \emptyset$, there exists a positive integer k such that $C^k \geq T$. Let

$$N = \min \{k \varepsilon Z^+ | \; C^k \geq T\} .$$

Lemma 2: If $U \neq \emptyset$ and if $C^1 \neq 0$, then $C^1 \geq T$ and **M** is not feasible.

Lemma 3: If $U \neq \emptyset$ and if $C^1 = 0$, then $C^N = T$ and C^k is a non-decreasing function of k for k=1,...,N.

Theorem 1: If $U \neq \emptyset$, suppose $C^1 = 0$ and N is as above. Then **M** is feasible if and only if $c^{N-1}_{ijk} > 0$ for all (i=1,...,M; j=1,...,N; k=1,...,P).

Iterative Procedure to Determine Feasibility. Given a contingency table **A**, to determine whether or not **A** is feasible proceed as follows. Form **M** as in the paragraphs above. If $U = \emptyset$, then A is not feasible. Otherwise, obtain C^1: If $C^1 \neq 0$, then **A** is not feasible by Lemma 2. If $C^1 = 0$, form C^2, C^3, etc., until $C^N = T$, and examine the cost matrix (c^{N+1}_{ijk}). If $c^{N+1}_{stv} = 0$ for any (s,t,v) then **A** is not feasible, otherwise **A** is feasible by Theorem 1. Given the marginal values in Table 3, the contraints (5)-(8) cannot be satisfied, and the set U is empty.

V. SUMMARY REMARKS

In this report we extend earlier work and show how to adjust arbitrary non-additive feasible tables into additive tables minimizing the power-divergence statistic introduced by Creesie and Read (1984). We provide examples and theoretical background for the procedures introduced. These methods can be easily extended to tables of dimension greater than two. In additions, we present procedures for determining when three-dimensional tables are feasible. The algorithms presented for this purpose extend directly to tables of dimension greater than three. Background issues for table adjustment and a bibliography are presented in the authors' earlier papers, see Fagan and Greenberg (1985, 1987, and 1988).

REFERENCES

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

Alexander, C.H. (1990) "Incorporating Persons Estimates into Household Weighting Using Verious Measures of Convergence," Proceedings of the Fifth Annual Research Conference, Bureau of the Census, Washington, D.C., 446-462.

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.

Cressie, N. and Read, T. (1984) "Multinomial Goodness-of-Fit Tests", Journal of the Royal Statistical Society Series B, **46**, 440-464.

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 Association, 195-200, Washington, D.C.

Fagan, J. and Greenberg, B. (1988) "Algorithms for Making Tables Additive; Raking; Maximum Likelihood, and Minimum Chi-Square," Proceedings of the Section on Survey Research Methods, American Statistical Association, 467-472, Washington, D.C.. (This is a condensed version of the 1985 Report with the same title.)

Fagan, J. and Greenberg, B. (1990) "Making Tables Additive in Three Dimension Under λ-Measures ," Proceedings of the Section on Survey Research Methods, American Statistical Association, Alexandria, Va. (to appear).

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.

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.

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

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

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.

Read, T. and Cressie, N. (1988). Goodness-of-Fit Statistics to Discrete Multivariate Data, Springs-Verlag, New York.

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.