CONVERGENCE PROPERTIES OF THE ITERATIVE 1980 CENSUS ESTIMATOR

John H. Thompson, Bureau of the Census

BACKGROUND

This paper presents a discussion of some of the properties of the 1980 census estimation and weighting procedures. As has been thoroughly discussed (1), the 1980 census data was collected on a 100-percent basis for certain basic characteristics, and on a sample basis for a wide collection of demographic data items. The census estimation procedure is performed on a state by state basis to assign weights to the sample person and housing unit records. The first step in the estimation procedure is to divide or partition a state into mutually exclusive portions of geography referred to as weighting areas. Weighting areas are required to have a minimum sample count of 400 persons, and are required to respect certain municipal and census geographic boundaries within the 400 sample person constraint. Furthermore, weighting areas are never permitted to cross county boundaries. Within each weighting area, the estimation and weighting operations are performed separately for the sample of persons and the sample of housing units. Both procedures, however, are very similar. Counts are obtained from the sample for the interior cells of a multidimensional weighting array defined by demographic characteristics that were collected on a 100-percent basis during the 1980 census. The counts from the 100-percent census are also available for each marginal category of the weighting array. The sample and complete counts for each marginal category are next tested against a set of criteria. Rows and columns of the array which correspond to the marginal categories which fail this test are collapsed or combined together until an array is obtained with a set of marginal counts which satisfy the collapsing criteria. The sample interior cell counts are then scaled and rescaled via iterative proportional fitting, or raking, so that the sum in each row or column consecutively agrees with the 100-percent census row or column count. This procedure is iterated a number of times, and the adjusted sample counts are then allocated to the weighting area sample person or housing unit records as weights.

For a two dimensional weighting matrix; with R rows and C columns (after collapsing), and for a given weighting area, the procedure may be described as follows:

Let;

- N denote the total census 100-percent count for the weighting area
- n denote the total census sample count for the weighting area
- N_{ij}, n_{ij} denote the 100-percent and sample counts for the cell in the i^{th} row and j^{th} column of the weighting matrix.

The iterative adjustment is then performed by computing values of $\hat{N}_{ij}^{(t)}$ for $t = 0,1,2,\ldots$ as

$$\hat{N}_{ij}^{(0)} = \frac{N}{n} n_{ij}$$

$$\hat{N}_{ij}^{(1)} = \hat{N}_{ij}^{(0)} \frac{N_{i.}}{\hat{N}_{i.}^{(0)}}$$

$$\hat{N}_{ij}^{(2)} = \hat{N}_{ij}^{(1)} \frac{N_{.j}}{\hat{N}_{.j}^{(1)}}$$

$$\vdots$$

$$\hat{N}_{ij}^{(2t-1)} = \hat{N}_{ij}^{(2t-2)} \frac{N_{i.}}{\hat{N}_{i.}^{(2t-2)}}$$

$$\hat{N}_{ij}^{(2t)} = \hat{N}^{(2t-1)} \frac{N_{.j}}{\hat{N}_{.j}^{(2t-1)}}$$

Letting $\pi_{ij} = \frac{n_{ij}}{n}$ 1/

$$P_{i.} = \frac{N_{i.}}{N} \quad i = 1,2,\ldots,R$$

$$P_{.j} = \frac{N_{.j}}{N} \quad j = 1,2,j\ldots,C$$

the above procedure may be seen to be equivalent to computing

$$N \hat{p}_{ij}^{(0)} = N \pi_{ij}$$

$$N \hat{p}_{ij}^{(1)} = N \hat{p}_{ij}^{(0)} \frac{P_{i.}}{\hat{p}_{i.}^{(0)}}$$

$$N \hat{p}_{ij}^{(2)} = N \hat{p}_{ij}^{(1)} \frac{P_{.j}}{\hat{p}_{.j}^{(1)}}$$

$$N \hat{p}_{ij}^{(2t-1)} = N \hat{p}_{ij}^{(2t-2)} \frac{P_{i.}}{\hat{p}_{i.}^{(2t-2)}}$$

$$N \hat{p}_{ij}^{(2t)} = N \hat{p}_{ij}^{(2t-1)} \frac{P_{.j}}{\hat{p}_{.j}^{(2t-1)}}$$

The convergence properties of the values of $\hat{N}_{ij}^{(t)}$ or equivalently of $\hat{p}_{ij}^{(t)}$ as t becomes large have been widely discussed (2), particularly when the values of π_{ij} are non-zero for each i,j. Of interest, Ireland and Kullback (3) have shown that in the above instance, $\hat{p}_{ij}^{(t)}$ converges to a set of values p_{ij}^{*} such that the

p_{ij}^{*}, minimize $I(p;\pi)$, where

$$I(p;\pi) = \sum_{i=1}^{R} \sum_{j=1}^{C} p_{ij} \ln (p_{ij}/\pi_{ij}),$$

subject to the restrictions that $p_{ij} > 0$ and that

$$\sum_{i} p_{ij} = P_{i.} \quad i = 1,2,\ldots, R, \text{ and}$$

$$\sum_{j} p_{ij} = P_{.j} \quad j = 1,2,\ldots, C.$$

Furthermore, Ireland and Kullback showed that the resulting estimates of the "true" interior weighting matrix cell probabilities (or totals) possess BAN properties and that the convergence is geometric. Thus, any estimates of sample characteristics highly correlated wit' the weighting matrix interior cell estimates would also have highly desirable properties.

Darroch and Ratcliff (4), give a more general discussion of the convergence of the $p_{ij}^{(t)}$, when some of the values of π_{ij} are allowed to be zero.

Based on Taylor Series expansions, it is possible to derive expressions which approximate the variance of estimates of sample characteristics which result from this procedure. Arora and Brackstone (5), and Rao (6) give an excellent documentation of these findings, and also give estimators of this variance.

In applying these iterative procedures to obtain sample estimates at the Census Bureau, the observed sample interior weighting matrix cell counts are often seen to include zero's which occur in a variety of structural patterns. Some of the properties of this iterative procedure will be discussed in four structural settings which hopefully will include all possible realizations of the interior cell structure of the observed sample weighting matrix counts. The discussions will center on an arbitrary two dimensional weighting matrix in any given weighting area. From the discussions below, it should be clear that the arguments given for converence are easily extrapolated to situations involving weighting arrays with a higher number of dimensions.

The four structural settings in which convergence of the 1980 census estimation procedure will be discussed are defined in terms of the separability (7) of the observed weighting matrix. Simply put, a weighting matrix will be said to be separable if the observed non-zero sample counts can be arranged into block diagonal form via elementary row and column operations. The following example involving a 5 by 4, weighting matrix will demonstrate this concept.

Example 1 - Separable Weighting Matrix

Row Number \ Column Number	1	2	3	4
1	X	0	0	X
2	0	X	X	0
3	0	0	0	X
4	0	X	X	0
5	X	0	0	X

If X represents an observed non-zero cell then this weighting matrix is separable, since by interchanging the rows and columns, the matrix can be put in the form shown below.

Example 1--Separable Weighting Matrix Cont'd

Row Number from above \ Column Number from above	1	4	3	2
1	X	X	0	0
3	0	X	0	0
5	X	X	0	0
2	0	0	X	X
4	0	0	X	X

An inseparable weighting matrix is simply a matrix which cannot be put into this form. A much more detailed and rigorous discussion of the concept of separability and inseparability may be found in Bishop, Fienberg and Holland (7).

Clearly, the operation of changing the order of the rows and columns of a weighting matrix via elementary row and column permutations will not effect the mathematics of the iterative census estimation procedure. It is also apparent that any adjustments made to the interior cells of the inseparable submatrices will only depend on the marginal controls of the submatrix, and not on the adjustments made to cells in any other inseparable submatrix. This motivates the four distinct structural settings on which the convergence properties of the iterative procedure will be discussed. These are as follows:

Structural Setting (1)

The observed sample weighting matrix interior cell counts are all non-zero. That is,

$$\pi_{ij} \neq 0 \text{ for } i = 1,2,\ldots, R,\ j = 1,2,\ldots,C.$$

Structural Setting (2)

The observed sample weighting matrix counts may be zero for some cells, but the observed sample weighting matrix is inseparable. Furthermore, the set of probabilities T_{IS} is non empty, where T_{IS} is a set of probabilities such that

$$T_{IS} = \left\{ p_{ij} \;\middle|\; \begin{array}{l} p_{ij} \neq 0 \text{ when } \pi_{ij} \neq 0 \\ p_{ij} = 0 \text{ when } \pi_{ij} = 0, \text{ and} \\ p_{i.} = P_{i.}\ \ i = 1,2,\ldots, R \\ p_{.j} = P_{.j},\ i = 1,2,\ldots, C \end{array} \right\}.$$

Structural Setting (3)

The observed sample weighting matrix is separable with M inseparable submatrices. Furthermore, for each inseparable submatrix, the set of probabilities $T_{S\ell}$, $\ell = 1,2,\ldots, M$ must be non empty, where for each value of ℓ,

$$T_{S\ell} = \left\{ \begin{array}{l} p_{ij} \mid p_{ij} \neq 0 \text{ when } \pi_{ij} \neq 0, \\ p_{ij} = 0 \text{ when } \pi_{ij} = 0, \\ p_{i.} = P_{i.}\ \ i = 1,2,\ldots, R_\ell \\ p_{.j} = P_{.j}\left[\dfrac{\sum_{i=1}^{R_\ell} P_{i.}}{\sum_{j=1}^{C_\ell} P_{.j}}\right] \end{array} \right\}$$

(where R_ℓ and C_ℓ are the row and column sizes for the ℓ^{th} inseparable submatrix).

Structural Setting (4)

The observed sample weighting matrix has at least one inseparable submatrix containing at least one zero for which the set of probabilities discussed above is empty. For an example of this situation, consider the following 2 by 2 matrix

X	X	$P_{1.} = 0.1$
X	0	$P_{2.} = 0.9$
$P_{.1}=0.2$	$P_{.2}=0.8$	1

(where as before X denotes a non-zero sample count).

1. Convergence for Structural Settings (1) and (1)

For structural setting (1) the problems associated with convergence have been solved, as discussed. For structural setting (2)--an inseparable weighting matrix containing zero values of π_{ij} — it may be shown that the same properties hold. The following lemma states this explicitly:

Lemma 1

Consider a weighting matrix with R rows and C columns, and with the structure given by structural setting (2). The iterative procedure converges to a set of counts or, equivalently, of probabilities p^*_{ij} such that:

a) $p^*_{ij} \neq 0$ when $\pi_{ij} \neq 0$

$p^*_{ij} = 0$ when $\pi_{ij} = 0$;

b) $p^*_{i.} = P_{i.}$ $i = 1,2,\ldots, R$

$p^*_{.j} = P_{.j}$ $j = 1,2,\ldots, C$ and

c) $I(P^*;\pi) = \sum_{i,j\,\varepsilon S} p^*_{ij} \ln \frac{p^*_{ij}}{\pi_{ij}}$

is a minimum among all probabilities with properties a) and b), and where S denotes the set of values of i, j for which $\pi_{ij} \neq 0$.

The additional properties related to estimates which minimize the discrimination information function immediately follow.

Proof of Lemma 1:

The proof is omitted due to space limitations. A similar proof is given by Darroch and Ratcliff (4).

2. Structural Setting (3) Convergence

For structural setting (3), the observed sample weighting matrix consists of a number of inseparable submatrices. It is only necessary to consider the convergence properties of the procedure for any one of the inseparable submatrices. This follows since the iterative procedure for each submatrix is equivalent to applying the procedure to an inseparable matrix, with R rows and C columns, but for which

(2.1.1) $$\sum_{i=1}^{R} P_{i.} \neq \sum_{j=1}^{C} P_{.j}$$

The following lemma summarizes the properties of the procedure in this situation.

Lemma 2

Let M_1 and M_2 denote two weighting matrices both with R rows and C columns, and with the same observed sample counts π_{ij}. The marginal counts for M_1 are $P_{1.}, P_{2.},\ldots, P_{R.}, P_{.1}, P_{.2},\ldots, P_{.C}$, the marginal counts for M_2 are $P'_{1.}, P'_{2.},\ldots, P'_{R.}, P_{.1},\ldots, P_{.C}$, and

1) $P'_{i.} = F\,P_{i.}$, where $F = \frac{\sum_{j=1}^{C} P_{.j}}{\sum_{i=1}^{R} P_{i.}}$

2) Let $f^{(n)}_{ij}$ denote the n^{th} application of the iterative procedure to M_1, and $\hat{p}^{(n)}_{ij}$ the n^{th} application to M_2.

Then if the set T_{IS} for M_2 is non-empty,

$$\lim_{n\to\infty} \hat{f}^{(2n)}_{ij} = \lim_{n\to\infty} \hat{p}^{(n)}_{ij} \text{ for } i = 1,2,\ldots, R;\ j = 1,2,\ldots, C.$$

and

$$\lim_{n\to\infty} \hat{f}^{(2n-1)}_{ij} = F \lim_{n\to\infty} \hat{p}^{(n)}_{ij} \text{ for } i = 1,2,\ldots, R;\ j = 1,2,\ldots, C.$$

Proof: From Lemma 1, $\lim_{n} \hat{p}^{(n)}_{ij}$ exists.

As an illustration of the method of proof, consider:

(2.1.2) $$p^{(4)}_{ij} = \frac{P_{.j}}{\hat{p}^{(3)}_{.j}} \frac{P_{i.}}{\hat{p}^{(2)}_{i.}} \frac{P_{.j}}{\hat{p}^{(1)}_{.j}} \frac{P_{i.}}{\pi_{i.}} \pi_{ij}$$

Note; 1) $$\hat{p}^{(1)}_{.j} = \sum_{i=1}^{R} \hat{p}^{(1)}_{ij} = \sum_{i=1}^{R} \pi_{ij} \frac{P'_{i.}}{\pi_{i.}} = F \sum_{i=1}^{R} \pi_{ij} \frac{P_{i.}}{\pi_{i.}} = F\,\hat{f}^{(1)}_{.j}$$

2) $$\hat{p}^{(2)}_{ij} = \pi_{ij} \frac{P_{.j}}{\hat{p}^{(1)}_{.j}} \frac{P'_{i.}}{\pi_{i.}} = \pi_{ij} \frac{P_{.j}}{F\,\hat{f}^{(1)}_{1j}}\ F \frac{P_{i.}}{\pi_{i.}} = \hat{f}^{(2)}_{ij}$$

so that $\hat{p}^{(2)}_{i.} = \hat{f}^{(2)}_{i.}$

3) $$\hat{p}^{(3)}_{.j} = \sum_{i=1}^{R} p^{(3)}_{ij} = \sum_{i=1}^{R} \hat{p}^{(2)}_{ij} \frac{P'_{i.}}{\hat{p}^{(2)}_{i.}} = F\,\hat{f}^{(3)}_{.j}$$

Substituting 1), 2), and 3) into 2.1.2 gives

(2.1.3) $$\hat{p}^{(4)}_{ij} = \frac{P_{.j}}{F\,\hat{f}^{(3)}_{.j}} \frac{F\,P_{i.}}{F\,\hat{f}^{(1)}_{i.}} \frac{P_{.j}}{F\,\hat{f}^{(1)}_{i.}} \pi_{ij} = \hat{f}^{(4)}_{ij}$$

It should now be clear, that a straightforward induction proof will establish that

$$\hat{p}^{(2n)}_{ij} = \hat{f}^{(2n)}_{ij} \text{ for } n = 1,2,\ldots$$

$$\hat{p}^{(2n-1)}_{ij} = F\,\hat{f}^{(2n-1)}_{ij} \text{ for } n = 1,2,\ldots$$

Finally, since the $\lim_{n} \hat{p}^{(n)}$ exists, as was shown in Lemma 1, the result stated in Lemma 2 follows.

For structural setting (3), the iterative procedure converges, in a sense, for each inseparable submatrix of the weighting matrix. However, the statistical properties of the resulting estimates became somewhat clouded. Since for the 1980 census, the iterative procedure is stopped after an even number of applications the census estimates may be seen to converge to a set of estimates which minimize.

$$I(P';\pi) = \sum_{i=1}^{R_m} \sum_{j=1}^{C_m} P'_{ij} \ln \frac{P'_{ij}}{\pi_{ij}}$$ for each submatrix, where the P'_{ij}'s have the marginal struc-

ture associated with matrix M_2 above. Letting F_m represent the ratio of the row to the column marginal sums for each submatrix, it seems logical to assume that F_m could be used as a measure of the bias in estimates of characteristics highly correlated with the row categories of the weighting matrix. Unfortunately, more research must be conducted before this statement may be affirmed.

3. Structural Setting (4)

For structural setting (4), there is no firm result known to this author. However, the following example sets a basis for an "educated guess" as to the convergence of the procedure. Consider the following 2 by 2 matrix of interior observations and marginal counts.

.2	.5	.1
.3	0	.9
.2	.8	

Repeated application of the procedure will result in a matrix that tends to ossicilate between the following two matrices:

0	.8	.1
.2	0	.9
.2	.8	

After even Applications

0	.1	.1
.9	0	.9
.2	.8	

After odd Applications

The claim is thus made (without verification), that the iterative procedure will tend to converge to a matrix with structural setting (3), i.e., some of $\hat{p}_{ij}^{(n)}$ will converge to zero.

Ideally, it would be desirable to identify weighting matrices identified with structural settings (3) and (4) prior to conducting the weighting operations. This would permit further collapsing to eliminate this problem. The example given above would only be observed as the result of some unusual nonsampling error, and would be easily recognized, during the census processing procedures. However, it is unlikely that the following situation would be noticed during the processing:

.3	.2	.4
.5	0	.6
.5	.5	

The result would, unfortunately, be the same.

This was, in part, the motivation for only performing the 1980 census ratio estimation procedures twice. It was felt that the high order of the rate of convergence of the procedure would provide adequate results for weighting areas with matrices that corresponded to structural settings (1) and (2). It was also believed that for structural settings (3) and (4) the resulting estimates would have the benefits of the ratio estimation controls, without demonstrating a significant tendency to the oscillating situation. It would be highly desirable to develop a efficient procedure to screen weighting areas to determine those with matrices of structural setting (3) or (4). This would permit the Census Bureau to further collapse these weighting matrices until a structural type was obtained that would permit convergence.

FOOTNOTE

1/ This notation is very similar to that used by Ireland and Kullback (3) and by Darroch and Ratcliff (4).

REFERENCES

(1) Thompson, J., Miskura, S., Woltman., and Bounpane, P., "1980 Census Weighting and Variance Estimation Studies, Design and Methodology", Contributed paper presented at the 1981 ASA meetings.

(2) In addition to the other references given some of the many sources which discuss these estimators are:

Deming, W.E., and Stephen, F.F., "On a Least Squares Adjustment of a Sampling Frequency Table When the Expected Marginal Totals are Known". The Annals of Mathematical Statistics, Vol. XI, 1940.

Fienberg, S.E., "An Iterative Procedure for Estimation in Contingency Tables", The Annals of Mathematical Statistics, 1970.

Ireland, C.T., and Scheuren, F., "The Rakes Progress", Contributed paper presented at the meetings of the ASA, 1974.

(3) Ireland, C.T., and Kullback, S., "Contingency Tables with Given Marginals", Biometrika, 55, 1968.

(4) Darroch, J.M., and Ratcliff, D., "Generalized Iterative Scaling for Log-Linear Models", The Annals of Mathematical Statistics, Vol. 43, No. 5, 1972.

(5) Arora, H.R., and Brackstone, G.J., "An Investigation of the Properties of Raking Ratio Estimators with Simple Random Sampling", Survey Methodology, Vol. 3, No. 1, 1977.

(6) Rao, J.N.K., "Raking Ratio Estimators" Unpublished Document, January 1974.
Rao, J.N.K., "Raking Ratio Estimators II" Unpublished Document, March 1974.
Rao, J.N.K., "Raking Ratio Estimators III" Unpublished Document, October 1974.

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

(8) Kullback, S., Information Theory and Statistics, John Wiley and Sons Inc., New York, 1959.

ACKNOWLEDGEMENTS

The author wishes to acknowledge Nash J. Monsour and Robert Fay for their very helpful comments given in the review of this paper.