This module is part of the

Memobust Handbook

on Methodology of Modern Business Statistics

26 March 2014

Method: RAS

Contents

General section

1. Summary

The RAS method is a well-known method for data reconciliation. Its aim is to achieve consistency between the entries of some nonnegative matrix and pre-specified row and column totals.

The method was devised in a time when powerful computers were not available. It is very easy to apply and to understand. However, it has a narrow scope of applicability. It can only be applied to nonnegative matrices. Mathematically, the method is an iterative scaling method.

2. General description of the method

Below we give a non-technical description of the RAS method. For a more detailed explanation we refer to Chapter IX of United Nations (1993). The original paper is by Bacharach (1970), but due to its technical character, we do not advise this for readers who are unfamiliar with the RAS method.

RAS is also called matrix raking or matrix scaling in computer science.

Technically, RAS is the same method as Iterative Proportional Fitting (Deming and Stephan, 1940) but both methods were designed for different problems. Iterative Proportional Fitting (IPF) was introduced to estimate contingency tables, when the marginals are fully observed and incomplete information may be available for the cells in the inner part of the table. Its intended use was for estimating census tables. RAS was introduced for updating a given table to new marginal totals, while preserving as much as possible the structure of the initial matrix. It was intended to use for estimating Input Output tables of the National Accounts.

Mathematically, RAS is an iterative scaling method whereby a non-negative matrix is adjusted until its column sums and row sums equal to some pre-specified totals. It multiplies each entry in one row or column by some factor, that is chosen in such a way that the sum of all entries in the row or column becomes equal to its target total. This operation is first applied to all rows of the matrix. As a consequence the matrix becomes consistent with all target row totals. Then, the columns are made consistent with their required totals. As a result consistency is achieved with the column totals, but the constraints on the row totals may be violated again. The rows and columns are adjusted in turn, until the algorithm converges to a matrix that is consistent with all required row and column totals.

The adjustment of the entries of the matrix always happens to be biproportional to the row and column totals, in order to preserve the structure of the matrix as much as possible. This means that all ratios between an entry of the inner part of a matrix and the corresponding row and column totals are kept as close as possible to their initial values.

In the literature several extensions of the RAS method are given, see for instance Lahr and De Mesnard (2004). Amongst others are the following methods:

- GRAS (Generalised RAS) allows for matrices in which some of the elements are predefined, in addition to the row and the column totals;
- Another GRAS method (Generalised RAS) allows for matrices with negative entries, Lenzen et al. (2007);

- TRAS (Three-stage RAS) extends RAS by including constraints on arbitrary subsets of matrix elements, instead of only fixing row and column sums, see Cole (1992) and Gilchrist and St. Louis (1999);
- KRAS, by Lenzen et al. (2009) includes the aforementioned features of GRAS and TRAS and further generalises RAS for the case of conflicting source data. The most simple case is when two data sources prescribe two different values for the same matrix entry. The initial RAS method will start from either of the conflicting values. KRAS however, will use both values and it allows for different reliabilities of the data sources.

  Another case of conflicting data that can be dealt with by KRAS is for discrepancies between marginals and the entries of a matrix. An example of this is a problem presented by Cole (1992)

  $$\begin{pmatrix} a & b \\ c & 1 \end{pmatrix}$$

  where a, b, $c \geq 0$ and with inconsistent row and column totals (1, 3)' and (1, 3). As this problem cannot be solved by using RAS, it is also called a RAS-infeasible problem. KRAS will still find a solution by treating some of the constraints as nonbinding. It searches for a solution with minimum constraint violation.

3. Preparatory phase

4. Examples – not tool specific

4.1 Example: the RAS method

The entries of the matrix in Table 1 below are inconsistent with the associated row and column totals. The RAS method is applied in order to produce a consistent table.

Table 1. Initial matrix

2	4	12
2	4	6
9	9	**18**

The rows are modified first. The sum of the entries in the first row is six, while the row total is twice as large. The entries of the matrix must therefore be multiplied by two. The second row of the matrix entries is consistent with the row total and is therefore left unmodified. The result of the above is Table 2. The matrix entries of this table are consistent with the row totals, but not with the column totals.

Table 2. First intermediate result

4	8	12
2	4	6
9	9	**18**

The next step of the algorithm modifies the columns. The sum of the first column is six and the column total is nine. The entries in the first column must therefore be increased by a factor of 9/6, or 1.5. The sum of the entries in the second column is 12, while the column total is 9. The two entries of the matrix must therefore be multiplied by 9/12, or 0.75, which produces Table 3. This table is entirely consistent. The algorithm stops. The algorithm would continue with rows if the table were still inconsistent.

Table 3. Final result

6	6	12
3	3	6
9	9	**18**

This example is quite simple. Real-life examples of this method may involve much more iterations. Further, the final matrix may involve broken numbers, even if all initial values are integers. In that case a stopping criterion has to be applied, to stop the algorithm when the discrepancies are sufficiently small.

5. Examples – tool specific

6. Glossary

For definitions of terms used in this module, please refer to the separate “Glossary” provided as part of the handbook.

7. References

Bacharach, M. (1970), *Biproportional matrices & input-output change*. Cambridge University Press, Cambridge.

Cole, S. (1992), A Note on a Lagrangian Derivation of a General Multi-Proportional Scaling Algorithm. *Regional Science and Urban Economics* **22**, 291–297.

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

Gilchrist, D. and St. Louis, L. (1999), Completing input–output tables using partial information with an application to Canadian data. *Economic Systems Research* **11**, 185–193.

Günlük-Şenesen, G. and Bates, J. M. (1988), Some experiments with methods of adjusting unbalanced data matrices. *Journal of the Royal Statistical Society, Series A* **151**, 473–490.

Lahr, M. L. and de Mesnard, L. (2004), Biproportional Techniques in Input-Output Analysis: Table Updating and Structural Analysis. *Economic Systems Research* **16**, 115–134.

Lenzen, M., Wood, R., and Gallego, B. (2007), Some Comments on the GRAS Method. *Economic Systems Research* **19**, 461–465.

Lenzen, M., Gallego, B., and Wood, R. (2009), Matrix balancing under conflicting information. *Economic Systems Research* **21**, 23–44.

United Nations (1993), *Handbook of National Accounting; Handbook of Input-Output Table Compilation and Analysis*, Series F, No. 74. United Nations, New York. http://unstats.un.org/unsd/publication/SeriesF/SeriesF_74E.pdf

Specific section

8. Purpose of the method

The method is used for data reconciliation which is a specific process step used in the context of macro-integration (cf. "Macro-Integration – Main Module").

9. Recommended use of the method

1. This method is only interesting if it is possible to represent the data in one rectangular matrix, in which the entries of the matrix are subject to change, while the row and column totals have to be equal to some pre-specified values.
2. The matrix does not have to be square: the number of rows may differ from the number of columns.
3. There is no way of differentiating between the reliability of different variables in the entries of the matrix. It is therefore impossible to make sure that a given value in the entries of the matrix is to undergo minimal or no change. However, some of the extensions of the RAS method that are described in the literature do allow for differences in reliability.
4. There is likewise no facility to impose constraints other than the row and column totals. It is therefore impossible to prescribe that the sum of two entries in the matrix equals a third entry, even if the condition was met in the initial situation. However, some extended versions of the RAS method in the literature do allow for these constraints.
5. The method is especially useful that the row and column totals are measured with a higher precision than the entries of the matrix.

10. Possible disadvantages of the method

1.

11. Variants of the method

1.

12. Input data

1. Ds-input1 = the entries of a matrix (required);
2. Ds-input2 = the required row and column totals of Ds-input1 (required).

Remark: the input data may involve micro- or macrodata.

13. Logical preconditions

1. Missing values
    1. In Ds-input1 individual missing data values are not allowed.
    2. In Ds-input2 individual missing data values are not allowed.
2. Erroneous values

    1.

3. Other quality related preconditions

    1.

4. Other types of preconditions

    1. The sums of the row and column totals in Ds-input2 are equal.
    2. Rows and columns consisting entirely of zeros do not occur in combination with a nonzero row or column total.
    3. All values in Ds-input1 and Ds-input2 are non-negative.

    Remark: A more complete description of the preconditions is given in Bacharach (1970). In order to avoid technical details we do not state the preconditions mentioned in this article.

14. Tuning parameters

1. Threshold values (optional), which specifies the maximum tolerated violation between the sum of the entries in one row or column of a matrix and the required row or column total.

15. Recommended use of the individual variants of the method

1.

16. Output data

1. Ds-output1 = the entries of a matrix

17. Properties of the output data

1. The output matrix (Ds-output1) is consistent with the pre-specified row and column sums (Ds-input2).
2. The amount of adjustments always happens to be biproportional to the row and column totals (Ds-input2).

18. Unit of input data suitable for the method

Processing full data sets

19. User interaction - not tool specific

1. Before execution of the method, the tuning parameters and input datasets must be specified.
2. During operation no user interaction is needed, but inspection of quality indicators and subsequent adjustment of tuning parameters and recurrent use is optional.
3. After use of the method the quality indicators and logging should be inspected.

20. Logging indicators

1. The running time of the application.
2. The number of iterations.

3. Characteristics of the input data, for instance problem size, and the largest discrepancies between the row and column totals of the initial matrix (Ds-input1) and the desired totals (Ds-input2).

21. Quality indicators of the output data

1. The most important quality indicator is *how* much the figures are adjusted. Relative or absolute differences may be explored. Because of the relationships between the various entries of the input matrix, the differences must be examined in their mutual context.

Remark 1: It would be possible, for example, to explore how the ratios between matrix cells and the row and column totals change in the reconciliation process. Remind that the RAS method attempts to preserve these ratios as much as possible. If a ratio has to change in the reconciliation process nonetheless, it is advisable to review the suitability of RAS.

Remark 2: Special attention is needed for zeros, both before and after reconciliation. The RAS method may create zeros if a row or column total is zero, but cannot adjust existing zero entries of the matrix. In both cases it must be verified that the data set has the correct structure.

22. Actual use of the method

1. At Statistics Netherlands the RAS method is used for updating the Input and Output Tables of National Accounts. At the current time, t, the row and column totals are fixed, since they have to be consistent with another statistic: the supply and use tables. Figures about the entries of the matrix are available only up to and including $t - 1$. Updating involves modifying the entries of the matrix at $t - 1$ in such a way as to make them consistent with the row and column totals at time t, also preserving the structure at $t - 1$ as much as possible.

Interconnections with other modules

23. Themes that refer explicitly to this module

1. Macro-Integration – Main Module

24. Related methods described in other modules

1. Macro-Integration – Stone’s Method
2. Macro-Integration – Denton’s Method
3. Macro-Integration – Chow-Lin Method for Temporal Disaggregation

Remark 1. The RAS method is the easiest method to apply, small problems can even be solved without a computer. However its field of application is more narrow than for the other methods. It is restricted to updating a matrix in such a way that it conforms to predefined row and column totals. The Stone method is more general: it does not require input data that can be represented in one matrix and it allows for more general type of constraints than the alignment to pre-specified row and column totals.

Remark 2. For the specific problem of updating a matrix to known row and column totals a lot of alternative methods are described in the literature as well, for an overview see for instance Lahr and De Mesnard (2004) and Lenzen et al. (2009). We choose to describe the RAS method, since this method is the most well-known and most rudimentary.

25. Mathematical techniques used by the method described in this module

1. Iterative Scaling (also called raking)

26. GSBPM phases where the method described in this module is used

1. GSBPM phase 6.2 “Validate Outputs”

27. Tools that implement the method described in this module

1.

28. Process step performed by the method

Data reconciliation

Administrative section

29. Module code

Macro-Integration-M-RAS

30. Version history

Version	Date	Description of changes	Author	Institute
0.1	31-03-2011	first version	Jacco Daalmans	CBS
0.2	05-04-2012	second version	Jacco Daalmans	CBS
0.3	21-06-2013	third version	Jacco Daalmans	CBS
0.3.1	06-09-2013	preliminary release		
0.3.2	09-09-2013	page numbering adjusted		
1.0	26-03-2014	final version within the Memobust project		

31. Template version and print date

Template version used	1.0 p 4 d.d. 22-11-2012
Print date	26-3-2014 13:25