



Max-Planck-Institut für demografische Forschung
Max Planck Institute for Demographic Research
Konrad-Zuse-Strasse 1 · D-18057 Rostock · GERMANY
Tel +49 (0) 3 81 20 81 - 0; Fax +49 (0) 3 81 20 81 - 202;
<http://www.demogr.mpg.de>

MPIDR TECHNICAL REPORT 2014-003
DECEMBER 2014

**An *R*-script for the assessment of the
cross-sectional and the longitudinal
components of a difference between two
values of an aggregate demographic
measure by contour replacement**

Dmitri A. Jdanov (jdjanov@demogr.mpg.de)
Vladimir M. Shkolnikov (shkolnikov@demogr.mpg.de)

For additional material see www.demogr.mpg.de/tr/

© Copyright is held by the authors.

Technical reports of the Max Planck Institute for Demographic Research receive only limited review. Views or opinions expressed in technical reports are attributable to the authors and do not necessarily reflect those of the Institute.

An *R*-script for the assessment of the cross-sectional and the longitudinal components of a difference between two values of an aggregate demographic measure by contour replacement

by Dmitri A. Jdanov and Vladimir M. Shkolnikov

Abstract

An algorithm for the contour decomposition of differences between two values of an aggregate demographic measure is realized in *R*. The contour replacement algorithm is an extension of the stepwise replacement algorithm. It allows us to split the cross-sectional difference at a later time point by age, and to further split it through contributions of the initial conditions and of the trend. The input arguments are matrices of event rates for two populations under comparison. Each of the matrices contains two vectors of event rates for the initial and the final time points. The aggregate measure is computed by an external *R* function.

Background

The proposed *R*-script realizes the algorithm of the contour replacement proposed by Jdanov and Shkolnikov (2014). In the script, we keep the same notation and the same steps of calculation as those given in the paper. We tried to avoid the use of programming tricks which may shorten the code, but complicate its understanding. Below we summarize the main ideas of the method. But we recommend that readers who wish to gain a better understanding of what the program does and how best to interpret its results read the paper first.

The general decomposition problem is to estimate the additive contributions of the inter-population differences between the elementary event rates according to age and/or other demographic dimensions—such as cause of death, birth order, and sub-population—to the overall inter-population difference in the aggregate demographic index. The best known discrete decomposition techniques permit us to split the overall inter-population difference in life expectancy by age (Andreev, 1982; Arriaga, 1984; Pressat, 1985).

Andreev, Shkolnikov, and Begun (2002) proposed the stepwise replacement algorithm for solving the general decomposition problem. For some aggregated measures (life expectancy and healthy life expectancy) this method results in analytical solutions, but for many other measures there are numerical solutions, because the analytical expressions either cannot be deduced, or they result in overly complicated formulae.

The contour replacement algorithm is a natural extension of the stepwise replacement algorithm, and provides a solution for a more complex decomposition task.

In the following, we describe software for solving the core problem of one-dimensional age decomposition by contour replacement.

Imagine that an aggregate index (say, life expectancy) is measured in two populations at two time points t and T . In order to understand the reasons for the inter-population difference at time T , we might want to assess the contributions of different ages to the total between-population difference at time point T . This implies a conventional decomposition that can be

accomplished with any of the above-mentioned classic techniques of the 1980s. Yet it is clear that every age component depends on the initial differences between the respective event rates at time t (initial conditions), and on the temporal changes in these rates in the two populations (trends) from moment t to moment T (Figure 1). In other words, the final difference Δ_{AB} is to be decomposed by age and (within each age) by the initial conditions (the differences between event rates m_a and m_b) and the trend (temporal changes from m_a to m_A and from m_b to m_B).

The contour decomposition allows us to split the final inter-population difference at time T into age-specific contributions due to the initial inter-population differences in age-specific mortality at time t , and age-specific contributions due to (within-country) age-specific mortality trends over time from t to T .

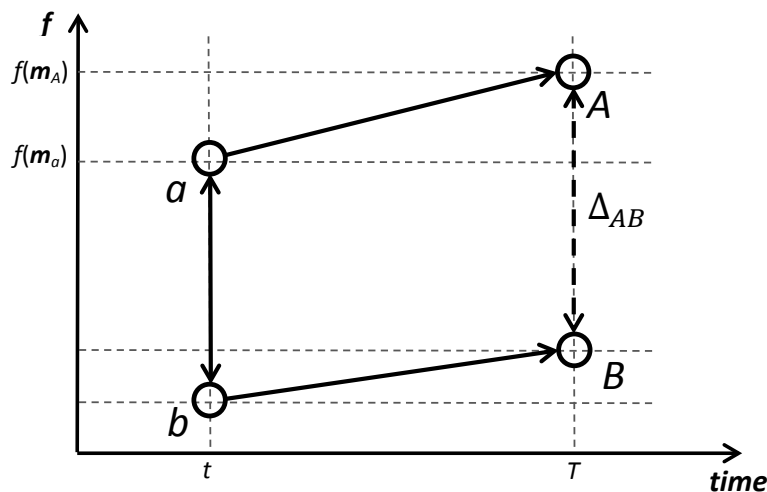


Figure 1. Cross-sectional differences and longitudinal changes in an aggregate demographic measure $E = f(m)$.

In the next section, we describe how the script can be used, and provide two practical examples. These examples are identical to those presented by Jdanov and Shkolnikov (2014). Thus, this technical report provides the data and the calculations which led to the results shown in the paper.

From this point onward, we assume that the reader has basic knowledge of R . If the reader is not familiar with R , we recommend reading a brief introduction to R before continuing on to the next section.

How to use it

To test all of the examples, all three R -files should be sourced. Nevertheless, the method itself is realized in one R -function script `decomp.contour()` (file `decomp_cotour.R`). In addition to

scripts, we provide the data used in the examples. The data for the two examples are extracts from the Human Mortality Database (www.mortality.org) and the Human Fertility Database (www.humanfertility.org).

Here we provide only a description of the function for contour replacement decomposition, its arguments, and the structure of the input and the output data. Examples of the use of the program for the example calculations are given in the following section.

Usage: `decomp.contour(c1, c2, ages = c(0, 15, 40, 65, 80, 111), FUN = ex.per, ...)`

Input arguments:

- **c1, c2** are two numeric matrices or data frames of event rates *c1* and *c2*. The first column of the data frame is used for the age scale. The values in this column are the lower bounds of the age intervals. The length of every interval is the difference between two consecutive lower bounds. The last row of the first column presents an open-ended age interval. The second and the third columns contain event rates at the initial and the final time points, respectively.

For example:

```
> head(c1)
  Age mx1970 mx2009
1   0 0.01530 0.00263
2   1 0.00186 0.00041
3   2 0.00117 0.00020
4   3 0.00093 0.00015
5   4 0.00086 0.00014
6   5 0.00075 0.00012
> tail(c1)
  Age mx1 mx2
106 105 0.68809 0.55888
107 106 0.71054 0.58868
108 107 0.73201 0.61785
109 108 0.75244 0.64619
110 109 0.77179 0.67353
111 110 0.79006 0.69975
```

- **ages** is a vector which defines the age groups for the vectors of the output effects. For example, `ages = c(0, 15, 40, 65, 80, 111)` define the presentation of the results as contributions of aggregations of the age groups 0-14, 15-39, 40-64, 80-110.
- **FUN** is the name of an (external) function for the calculation of the aggregate measure to be decomposed.
- **...** any other arguments to be passed to FUN

Output: The function returns a list with the following elements:

- **\$decomp** is the data frame with the results of the decomposition. The data frame has three columns (the initial conditions effect, the trend component, and the results of the conventional decomposition).
- **\$values** is the data frame with the calculated statistics of interest for both populations for the initial and the end time points.

Examples:

```
> decomp.trend(country1, country2, FUN = ex.per)
```

```
> decomp.trend(country1, country2, ages = 0:111, FUN = ex.per)
```

The detailed examples are given below.

Examples of decompositions of life expectancy and of variability in age at first birth

Two examples presented in this section realize the calculations for the empirical examples from Jdanov and Shkolnikov (2014). The research questions and the interpretation of the results are provided in the paper. In both examples, we assume that the data files from the archive [data.zip](#) are placed in the working directory, and that all of the respective R scripts from the [scripts.zip](#) are sourced.

Example 1. Life expectancy gap between Japan and the USA in 2009, with a reference to 1970.

Using the Human Mortality Database (HMD, 2014) data, we calculate the contributions to the total difference in life expectancy at birth between Japan and the USA in 2009 of age groups, with a further split into the dynamic and the initial (of the year 1970) conditions. We take the data for the mortality rates from the Japanese and the US life tables which we downloaded from the HMD and saved in the [data.zip](#) archive under the names USAfltper_1x1.txt (females, USA), USAmltper_1x1.txt (males, USA), JPNfltper_1x1.txt (females, JPN), and JPNmltper_1x1.txt (males, JPN). The decomposition effects for males are calculated as follows:

```
> cdecomp.ex(' JPNml tper_1x1. txt', ' USAml tper_1x1. txt', c(1970, 2009), sex = ' m' )
```

```
$decomp
  initial      trend conventional
0  0.5642216 -0.14914097  -0.4150806
15 0.7883467  0.04087993  -0.8292266
40 1.9228056 -0.56036691  -1.3624387
65 -0.1709430  0.88146118  -0.7105181
80 -0.5627068  0.66591935  -0.1032125
```

```
$values
```

```
                1970    2009
JPNml tper_1x1. txt 69.31787 79.55168
USAml tper_1x1. txt 67.02132 76.13120
```

Here the first element of the list (`$decomp`) contains the initial conditions (column initial) and the trend (column trend) components split by age groups, as well as the age components of the conventional decomposition of the life expectancy difference in 2009 (third column). The sum of the initial condition and the trend components is equal to the conventional contribution. Each row represents components for the specific age group. The name of each row corresponds to the lower limit of the age group. In the given output, the age-specific components represent the contribution of the age groups 0-14, 15-39, 40-64, and 80+.

The second element of the list (`$values`) represents the value of life expectancy for Japan and the USA in 1970 and 2009, calculated using mortality rates from the respective data files.

The functions in file [ex.R](#) used in example 1 are as follows:

`cdecomp.ex (file_name1, file_name2, years = c(1970,2009), sex = 'm')` performs the contour decomposition of the difference in life expectancy at birth. `file_name1` and `file_name2` are the data files with the HMD life tables for Japan 1970 and 2009 and the USA in 1970 and 2009.

`ex.per (mx, age, sex = 'm')` calculates life expectancy at birth using mortality rates from `mx` and the age groups from the vector `age`.

`mx.select (file_name, years)` creates input matrix of mortality rates for the `decomp.contour` using data from the HMD life table file.

Example 2. Difference in the age at first birth variability between Czech and Russian female cohorts of 1970, with a reference to 1955.

The decomposition of the 1970 cohort's difference in the standard deviation of first births into the initial conditions (in the 1955 mothers' cohort) and the trend components was implemented using the HFD data. The input data files `CZEasfrVHbo.txt` and `RUSasfrVHbo.txt` contain age-specific fertility rates for the Czech Republic and Russia, classified by birth cohort, age, and birth order. For the first birth order, the age components of the decomposition are as follows:

```
> cdecomp.sdmab('CZEasfrVHbo.txt', 'RUSasfrVHbo.txt', bo = 1, years = c(1955, 1970), agerange = c(12, 39))
```

```
$decomp
  initial trend conventional
12      0     0            0
20      0     0            0
25      0     0            0
30      0     0            0
35      0     0            0
```

`$values`

```

              1955      1970
CZEasfrVHbo.txt 8.077747 8.077747
RUSasfrVHbo.txt 8.077747 8.077747
```

As in the first example, the first element of the list (`$decomp`) contains the calculated initial conditions and the trend components, and the total components of the final Czech-Russian difference by the age groups 12-19, 20-24, 25-29, 30-34, and 35+. The values of the standard deviation of the age at first birth for the birth cohorts of 1955 and 1970 are given in `$values`.

The functions in file `sdmab.R` used in this example are as follows:

`cdecomp.sdmab (fname1, fname2, bo = 1, years = c(1955, 1970), agerange = c(12,39))` performs the contour decomposition of the difference in the standard deviation of the age at birth. `file_name1` and `file_name2` are the HFD files with the age-specific fertility rates, and `bo` defines the birth order¹. The standard deviation is calculated over the ages defined in `agerange`.

¹ The functions can aggregate data over birth order. In such cases the argument should be defined as a vector; e.g., `bo=1:3`.

`sdmab(fx, age)` calculates the standard deviation of the age at birth using the age-specific fertility rates `fx`.

`ASFR.select (fname, years, bo, agerange)` prepares the input data for the contour decomposition (`decomp.contour`).

`readHFDfx(fname)` reads the data from the standard HFD file with the age-specific fertility rates.

Acknowledgements

This work was completed as part of the project on Divergence and Causes of Death (MODICOD) funded by the AXA Foundation, and of the project on Disparities in Mortality Trends to Future Health Challenges (DIMOCHA) funded by the Agence nationale de la recherche (ANR) and the Deutsche Forschungsgemeinschaft (DFG).

References

Andreev, E.M. (1982) Metod komponent v analize prodoljitelnosti zizni. [The method of components in the analysis of length of life]. *Vestnik Statistiki*, 9: 42-47.

Andreev E.M., Shkolnikov V.M., Begun A.Z. (2002) Algorithm for decomposition of differences between aggregate demographic measures and its application to life expectancies, healthy life expectancies, parity-progression ratios and total fertility rates. *Demographic Research*, 7: 499-522. Available at www.demographic-research.org

Arriaga E. 1984. Measuring and explaining the change in life expectancies. *Demography* 21(1): 83-96. Jdanov D.A and Shkolnikov V.M. (2014) Assessment of cross-sectional and longitudinal components of a difference with an algorithm of contour replacement. *MPIDR Working Paper* WP-2014-010 at <http://www.demogr.mpg.de/papers/working/wp-2014-010.pdf>

Pressat, R. (1985). Contribution des écarts de mortalité par âge à la différence des vies moyennes. *Population*: 4-5, 766-770.