Ratios of exponential functions, interpolation

Josef Bukac
Bulharska 298, Jaromer-Josefov 55102
Czech Republic

Abstract: We describe models of proportions depending on some independent quantitative variables. An explicit formula for inverse matrices facilitates interpolation as a way to calculate the starting values for iterations in nonlinear regression with ratios of exponential functions.

Keywords: Compositional data, Hermite, nonlinear, restricted, explicit inverse, logistic.
AMS classification: 41A29, 41A30, 62J02, 65D05.

1. Introduction

We study the ratios of exponential functions as a model of dependence of proportions on some other independent variables. The origin of such proportions may range from categorical data as relative frequencies when absolute frequencies are known on the one hand to chemical composition on the other hand. Other examples on compositional data may be had in Aitchison (1986) or Pawlowsky-Glahn (2015).

Ratios of exponential functions are also known as logistic functions, Agresti (1990). We refer to the paper by Bukac (2023) Unfortunately, linear models have a serious disadvantage in taking on negative values or values greater than one. One way or another, this is one of the reasons why a logistic regression is used to eliminate this drawback. However and despite the disadvantages of the linear model, the ease of computation of regression coefficients in restricted linear regression makes it useful for obtaining the values and derivatives as a basis for obtaining good starting values for iterative processes.

First of all we derive explicit formulas for inverse matrices without which we could not move on. Only then do we form equations in which the values and derivatives of the ratios of exponential functions are set equal to preassigned values. It turns out that explicit solutions can be found.

We also discuss that the coefficients of the systems of equations form singular matrices and why the Gauss-Newton method does not work. Several approaches may be used to arrive at the minimal solution.

2. Helpful matrix results, explicit inverse

Our interpolation problems lead to certain types of matrices. Some matrices have the same rows, other times the matrices have the same columns.

Definition 2.1. *A matrix in which all the rows are the same is a matrix with repeated rows.*

If a matrix $\mathbf{A}$ with repeated rows is given, we study $\mathbf{A}-\mathbf{I}$, that is,

$$\mathbf{A}=\begin{pmatrix} a_1 & a_2 & \dots & a_n \\ a_1 & a_2 & \dots & a_n \\ \vdots & \vdots & \ddots & \vdots \\ a_1 & a_2 & \dots & a_n \end{pmatrix}, \quad \mathbf{A}-\mathbf{I}=\begin{pmatrix} a_1-1 & a_2 & \dots & a_n \\ a_1 & a_2-1 & \dots & a_n \\ \vdots & \vdots & \ddots & \vdots \\ a_1 & a_2 & \dots & a_n-1 \end{pmatrix} \tag{1}$$

Definition 2.2. *A matrix $\mathbf{A}$ in which all the columns are the same is a matrix with repeated columns,*

$$\mathbf{A}=\begin{pmatrix} a_1 & a_1 & \dots & a_1 \\ a_2 & a_2 & \dots & a_2 \\ \vdots & \vdots & \ddots & \vdots \\ a_n & a_n & \dots & a_n \end{pmatrix}$$

The following theorem 2.1 includes a statement regarding the singularity or non-singularity of the matrix **A-I**. It is easy to verify that $tr(\mathbf{A}) \neq 1$ which implies non-singularity allowing us to apply an explicite formula for the inverse.

Theorem 2.1. *If $n>0$, $tr(\mathbf{A}) \neq 1$, and $\mathbf{A}$ is a square matrix with n repeated rows $(a_1, a_2, \dots, a_n)$ or a square matrix with n repeated columns $(a_1, a_2, \dots, a_n)'$ then $(\mathbf{A}-\mathbf{I})^{-1}=\mathbf{A}/\big(tr(\mathbf{A})-1\big)-\mathbf{I}$. If $tr(\mathbf{A})=1$, then $\mathbf{A}-\mathbf{I}$ is singular.*

Proof. Let $\mathbf{A}$ be a matrix with repeated rows such as the one in (1). We consider the matrix $\mathbf{AA}$ (it is not $\mathbf{A'A}$), an ij-th element of which is $a_j \sum_{k=1}^n a_k$, therefore $\mathbf{AA}=(\sum_{k=1}^n a_k)\mathbf{A}=tr(\mathbf{A})\mathbf{A}$. Now we calculate

$$\big(\mathbf{A}+(1-tr(\mathbf{A}))\mathbf{I}\big)\big(\mathbf{A}-\mathbf{I}\big)=\mathbf{AA}+(1-tr(\mathbf{A}))\mathbf{A}-\mathbf{A}-(1-tr(\mathbf{A}))\mathbf{I}=$$

$$tr(\mathbf{A})\mathbf{A}+\mathbf{A}-tr(\mathbf{A})\mathbf{A}-\mathbf{A}-(1-tr(\mathbf{A}))\mathbf{I}=(tr(\mathbf{A})-1)\mathbf{I}.$$

This proves the first part of the theorem when $tr(\mathbf{A}) \neq 1$ and $\mathbf{A}$ has repeated rows. To prove the second part, we add the elements in each of the rows of $\mathbf{A}-\mathbf{I}$ in (1) to check that their sum is zero. It follows that the sum of all the n columns in $\mathbf{A}-\mathbf{I}$ is a zero vector which means they are linearly dependent.

If the matrix $\mathbf{A}$ has repeated columns, then its transpose $\mathbf{A}'$ has repeated rows. If $tr(\mathbf{A})=1$, then $\mathbf{A}'-\mathrm{I}$ is singular and so is $\mathbf{A}-\mathrm{I}$.

If $tr(\mathbf{A}) \neq 1$, then $(\mathbf{A}'-\mathbf{I})^{-1}=\mathbf{A}'/\big(tr(\mathbf{A})-1\big)-\mathbf{I}$ and $(\mathbf{A}-\mathbf{I})^{-1}=\big((\mathbf{A}'-\mathbf{I})^{-1}\big)'=\big(\mathbf{A}'/\big(tr(\mathbf{A})-1\big)-\mathbf{I}\big)'=\mathbf{A}/\big(tr(\mathbf{A})-1\big)-\mathbf{I}$.

Theorem 2.2. *Let $\mathbf{j}=(1,1,\dots,1)'$ be a column vector consisting of n ones. If $\mathbf{A}$ is a square matrix with repeated rows $(a_1, a_2, \dots, a_n)$ and $tr(\mathbf{A}) \neq 1$ then*

$$(\mathbf{A}-\mathbf{I})^{-1}\mathbf{j}=\frac{\mathbf{j}}{tr(\mathbf{A})-1}.$$

Proof.

$$(\mathbf{A}-\mathbf{I})^{-1}\mathbf{j}=\frac{\mathbf{Aj}}{tr(\mathbf{A})-1}-\mathbf{Ij}=\frac{tr(\mathbf{A})}{tr(\mathbf{A})-1}\mathbf{j}-\mathbf{j}=\frac{tr(\mathbf{A})-tr(\mathbf{A})+1}{tr(\mathbf{A})-1}\mathbf{j}=\frac{\mathbf{j}}{tr(\mathbf{A})-1}.$$

3. Calculation of values and derivatives

Iterative minimization requires good starting parameters that may be based on the values and derivatives of functions at some specific point. The methods of obtaining these values and derivatives may vary. Our criterion is the overall sum of squares of deviations, our options depend on the type of data. It is obvious that the use of restricted linear regression is easy to implement to all situations. It is described in the appendix.

We may also consider other approaches. If we have T observations on each of M variables $z_{t1}, \ldots, z_{tM}$ for each $t = 1, \ldots, T$ and K variables $x_{t1}, \ldots, x_{tK}$. We define the $T \times M$ regressand matrix $\mathbf{Z}$ as $\mathbf{Z} = (\mathbf{z_1}, \ldots, \mathbf{z_m}, \ldots, \mathbf{z_M})$. We may use the M classical regression models and minimize the sum of squares

$$\sum_{t=1}^{T} \left(\sum_{k=1}^{K} x_{tk} b_{km} - z_{tm} \right)^2$$

for each $m = 1, 2, \ldots, M$. If, for all $m = 1, 2, \ldots, M$ we have $\sum_{k=1}^{K} x_k b_{km} > 0$ for x_k, $k = 1, 2, \ldots, K$ in some common domain of definition. But, when we want to approximate the proportions, we can form the ratios

$$\frac{\sum_{k=1}^{K} x_k b_{km}}{\sum_{d=1}^{M} \sum_{k=1}^{K} x_k b_{km}}.$$

When the functions are defined this way, the sum of proportions is one.

This approach may become inconvenient because such functions may become negative. Another serious trouble is that the sample sizes may vary making this model useless. This type of model is of no use when the proportions only are known. Proportions of chemical substances fall in this category.

We could also form the ratios of exponential functions and use the parameters as the starting values. Such a model has the same drawbacks as the ratios of linear functions except that the exponential functions are always positive. Exponential functions are nonlinear in parameters but we could, of course, linearise them by taking logarithms.

4. Joint logistic functions

It turns out to be useful to introduce a parameter a_m to represent the constant term rather than saying that $x_{t1} = 1$ for $t = 1, 2, \ldots, T$.

Definition 4.1. *Let $M > 1, m = 1 \ldots M$. We define joint logistic functions as*

$$F_m(\mathbf{a},\mathbf{B},\mathbf{x}) = \frac{\exp(a_m + x_1 b_{1m} + \ldots + x_K b_{Km})}{\sum_{d=1}^{M} \exp(a_d + x_1 b_{1d} + \ldots + x_K b_{Kd})} = \frac{\exp(a_m + \mathbf{x}'\mathbf{b}_\mathrm{m})}{\sum_{d=1}^{M} \exp(a_d + \mathbf{x}'\mathbf{b}_\mathrm{d})}$$

where $\mathbf{x} = (x_1, x_2, \ldots, x_K)'$ are independent variables, $\mathbf{a} = (\mathrm{a}_1, \mathrm{a}_2, \ldots \mathrm{a}_\mathrm{M})'$ and

$$\mathbf{B} = \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1M} \\ b_{21} & b_{22} & \dots & b_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ b_{K1} & b_{K2} & \dots & b_{KM} \end{pmatrix}$$

are a vector and matrix of parameters respectively.

The m-th column $\mathbf{b}_\mathrm{m}$ of the matrix $\mathbf{B}$ appears in the numerator as $\mathbf{x}'\mathbf{b}_\mathrm{m}$ to comply with the usual notation in regression analysis.

These functions are positive for any argument as opposed to linear models. The sum of such functions is $\sum_{m=1}^{M} F_m(\mathbf{a},\mathbf{B},\mathbf{x}) = 1$ which is the property intrinsically included in the definition of these functions unlike the linear models. It is obvious that for $M = 1$ we get the usual logistic function.

We may limit ourselves to the simpler exponential functions of $a_m + b_m x$ instead of the general terms $a_m + \sum_{k=1}^{K} x_k b_{km}$. The results we obtain may be easily generalized for the latter general case.

We may now represent the individual functions, for $m = 1, 2, \dots, M$, in a simpler form as

$$F_m(x) = \frac{\exp(a_m + b_m x)}{\sum_{d=1}^{M} \exp(a_d + b_d x)}.$$

We can easily see that $F_m(x) > 0$ and $\sum_{d=1}^{M} F_m(x) = 1$ for any x. We obtain the derivative of this sum as $\sum_{d=1}^{M} F_m'(x) = 0$ because it is a derivative of a constant. All the higher derivatives of $\sum_{d=1}^{M} F_m(x)$ are equal to zero as well.

If both the numerator and denominator are multiplied by the same non-zero real number the value of the fraction stays the same. We see that a function of this type may be represented by distinct parameters

Theorem 4.1. *Let α be a real number. Let $a_1, a_2, \dots, a_M, b_1, b_2, \dots, b_M$ represent the functions of x,*

$$\frac{\exp(a_m + b_m x)}{\sum_{d=1}^{M} \exp(a_d + b_d x)} \qquad \textit{for} \quad m = 1, 2, \dots, M.$$

Then $\alpha + a_1, \alpha + a_2, \dots, \alpha + a_M, b_1, b_2, \dots, b_M$ represent the same functions.

Proof.

$$\frac{\exp(\alpha + a_m + b_m x)}{\sum_{d=1}^{M} \exp(\alpha + a_d + b_d x)} = \frac{\exp(\alpha)\exp(a_m + b_m x)}{\sum_{d=1}^{M} \exp(\alpha)\exp(a_d + b_d x)} = \frac{\exp(a_m + b_m x)}{\sum_{d=1}^{M} \exp(a_d + b_d x)}.$$

Theorem 4.2. *Let $a_1, a_2, \dots, a_M, b_1, b_2, \dots, b_M$ be the parameters of the functions of x*

$$\frac{\exp(a_m + b_m x)}{\sum_{d=1}^{M} \exp(a_d + b_d x)} \qquad \textit{for} \quad m = 1, 2, \dots, M.$$

Let C be a real number. Then there is an α such that $\sum_{d=1}^{M} \exp(\alpha+a_d+b_dC) = 1$ and $\alpha + a_1, \alpha + a_2, \ldots, \alpha + a_M, b_1, b_2, \ldots, b_M$ represent the same functions of x.

Proof. Let C be fixed. For any α we may write $\sum_{d=1}^{M} \exp(\alpha + a_d + b_dC) = \exp(\alpha) \sum_{d=1}^{M} \exp(a_d + b_dC)$. The equality is satisfied if we pick an α such that

$$1/\exp(\alpha) = \sum_{d=1}^{M} \exp(a_d + b_dC).$$

One direct application of this theorem 4.2 is in the minimization of the sum of squares of differences. Even though a local minimum is calculated and its value is unique, the parameters are not unique. If we allow the parameters $a_1, a_2, \ldots, a_M$ to be arbitrarily large, we may get an overflow if the argument of the exponential function is too large. This problem may be taken care of by checking the size of the argument beforehand and making it smaller with a convenient choice of α.

We may also use the above theorem 4.2 and make the denominator equal to one at some fixed point C_1. We are mentioning this special choice of α at this moment but it will be used shortly.

5. Joint logistic interpolation

Depending on the type of data at hand we may use the restricted linear regression or ratios of linear regressions to obtain estimates of values $V_m > 0$, $\sum V_m = 1$, and derivatives D_m, where $\sum D_m = 0$, at some center point C at which all the values V_m are positive. The values and derivatives may be obtained in other ways but it is this context for which the restricted regression has been presented.

We assume that $b_1, b_2, \ldots, b_M$ are known and fixed. To find out what the values of $a_1, a_2, \ldots, a_M$ look like is easy.

Theorem 5.1. *Let $M > 1$. Let $V_m > 0$, for $m = 1, 2, \ldots, M$, $\sum_{m=1}^{M} V_m = 1$. Let $b_1, b_2, \ldots, b_M$ be fixed. Then the equations $\exp(a_m+b_mC) = V_m$ are satisfied if $a_m = \ln V_m - b_mC$ and $\sum \exp(a_m + b_mC) = 1$.*

Proof. We solve for a_m in $\exp(a_m + b_mC) = V_m$, which is trivial. Also $\sum V_m = 1$ implies that $\sum \exp(a_m + b_mC) = 1$.

We denote the derivatives with respect to x of $F_m(x)$, defined in the definition 4.1, as $F'_m(x)$ and calculate them as

$$F'_m(x) = \frac{\exp(a_m + b_mx)\big(b_m \sum_{d=1}^{M} \exp(a_d + b_dx) - \sum_{d=1}^{M} b_d \exp(a_d + b_dx)\big)}{\big(\sum_{d=1}^{M} \exp(a_d + b_dx)\big)^2}$$

Due to theorem 4.2, without losing generality, we may assume $\sum_{d=1}^{M} \exp(a_d + b_d C) = 1$ and obtain $F_m(C) = \exp(a_m + b_m C)$ for $m = 1, 2, \ldots, M$, and use the notation $\exp(a_m + b_m C) = V_m$ for $m = 1, 2, \ldots, M$.

The derivatives are $F'_m(C) = \exp(a_m + b_m C)\big(b_m - \sum_{d=1}^{M} b_d \exp(a_d + b_d C)\big)$ yielding equalities $V_m\big(b_m - \sum_{d=1}^{M} b_d V_d\big) = D_m$, where $D_m = F'_m(C)$, for $m = 1, 2, \ldots, M$. Because $V_m > 0$ for each $m = 1, 2, \ldots, M$, these equalities are equivalent to

$$\sum_{d=1}^{M} b_d V_d - b_m = -D_m / V_m \quad \text{for} \quad m = 1, 2, \ldots, M.$$

Now the question arises whether for any given $V_m > 0$ and D_m, $m = 1, 2, \ldots, M$, there are parameters $a_1, a_2, \ldots, a_M, b_1, b_2, \ldots, b_M$ that represent functions in accordance with definition 4.1.

Theorem 5.2. *Let $M > 1$. Let $V_m > 0$, for $m = 1, 2, \ldots, M$, $\sum_{m=1}^{M} V_m = 1$, and D_m are given then the system of linear equations for $x_1, x_2, \ldots, x_M$,*

$$\sum_{d=1}^{M} x_d V_d - x_m = -\frac{D_m}{V_m} \qquad \textit{for} \quad m = 1, 2, \ldots, M,$$

is of rank $M - 1$.

Proof. It is easy to see that the sum of columns of the matrix of this system of equations is equal to a zero column because $\sum V_m = 1$ showing the matrix of the system has rank less than M. We consider the submatrix consisting of the first $M - 1$ rows and $M - 1$ columns. From $\sum_{m=1}^{M} V_m = 1$ and $V_M > 0$ it follows that $\sum_{m=1}^{M-1} V_m \neq 1$ and we may use theorem 2.1 to show this submatrix has an inverse.

Definition 5.1 *Let $M > 1$. Let $V_m > 0$, for $m = 1, 2, \ldots, M$, $\sum_{m=1}^{M} V_m = 1$, and D_m are given. Consider the system of M linear equations for $x_1, x_2, \ldots, x_M$,*

$$\sum_{d=1}^{M} x_d V_d - x_m = -\frac{D_m}{V_m} \qquad \textit{for} \quad m = 1, 2, \ldots, M.$$

The system of the first $M - 1$ equations will be called a reduced system. The matrix of the reduced system will be denoted by $\mathbf{W}$ of type $(M - 1) \times M$.

Theorem 5.3. *Let $M > 1$. Let the values $V_m > 0$, for $m = 1, 2, \ldots, M$, $\sum_{m=1}^{M} V_m = 1$, and derivatives D_m at some point C be given. Then V_m and D_m may be interpolated with joint logistic functions,*

$$F_m(x) = \frac{\exp(a_m + b_m x)}{\sum_{d=1}^{M} \exp(a_d + b_d x)} \qquad \textit{for} \quad m = 1, 2, \ldots, M,$$

by solving the system of linear equations for $b_1, b_2, \ldots, b_M$,

$$\sum_{d=1}^{M} b_d V_d - b_m = -\frac{D_m}{V_m} \qquad \textit{for} \quad m = 1, 2, \ldots, M,$$

if and only if $\sum_{d=1}^{M} D_d = 0$.

Proof. The left hand side of the reduced system may be writen as $\mathbf{A} - \mathbf{I}$ of type $((M-1) \times (M-1))$ where $\mathbf{A}$ is a matrix with repeated rows. Since $\sum_{m=1}^{M} V_m = 1$ and $V_m > 0$ for each m, we have $\sum_{m=1}^{M-1} V_m \neq 1$. It means that $\mathbf{A} - \mathbf{I}$ has an inverse, $(\mathbf{A} - \mathbf{I})^{-1} = \mathbf{A}/(\sum_{K=1}^{M-1} a_k - 1) - \mathrm{I}$, due to theorem 2.1.

We may pick any b_M and calculate an explicite solution for $b_1, b_2, \ldots, b_{M-1}$ by left multiplying the matrix $\mathbf{W}$ by $(\mathbf{A} - \mathbf{I})^{-1} = \mathbf{A}/(\sum_{K=1}^{M-1} a_k - 1) - \mathrm{I}$. we also left multiply the column consisting of the $(M-1)$ components forming the right hand side of the system of equations.

Now we consider the function

$$F_M(x) = \frac{\exp(a_M + b_M x)}{\sum_{d=1}^{M} \exp(a_d + b_d x)}$$

and its derivative with respect to x. Since $\sum_{m=1}^{M} F_m(x) = 1$ for all x, The sum of derivatives is zero, $\sum_{m=1}^{M} F'_m(C) = 0$. It means the values and derivatives of $F_M(x)$ at C satisfy the M-th equation. If, on the other hand, $\sum_{m=1}^{M} D_m \neq 0$, the values and derivatives of $F_M(x)$ do not satisfy the system of all M equations. This finishes the proof for any b_M we pick.

The following theorem is more specific and may give a better understanding.

Theorem 5.4. *Let* $M > 1$. *Let* $V_m > 0$, *for* $m = 1, 2, \ldots, M$, $\sum_{m=1}^{M} V_m = 1$, *and* D_m *satisfy* $\sum_{m=1}^{M} D_m = 0$. *We form the system of linear equations*

$$V_m \sum_{d=1}^{M} x_d V_d - x_m V_m = -D_m \quad \textit{for} \quad m = 1, 2, \ldots, M.$$

If $b_1, b_2, \ldots, b_M$ *satisfy the first* $M-1$ *equations, then the* M*-th eqation is satisfied as well.*

Proof. We form the sum of the first $M-1$ equations as

$$\sum_{m=1}^{M-1} V_m \sum_{d=1}^{M} b_d V_d - \sum_{m=1}^{M-1} b_m V_m = -\sum_{m=1}^{M-1} D_m.$$

We use the following formulas:

$$1 = \sum_{m=1}^{M} V_m = V_M + \sum_{m=1}^{M-1} \quad \text{iff} \quad \sum_{m-1}^{M-1} V_m = 1 - V_M,$$

$$\sum_{m=1}^{M} b_m V_m = b_M V_M + \sum_{m=1}^{M-1} b_m V_m \quad \text{iff} \quad \sum_{m=1}^{M-1} b_m V_m = \sum_{m=1}^{M} b_m V_m - b_M V_M,$$

$$0 = \sum_{m=1}^{M} D_m = D_M + \sum_{m=1}^{M-1} \quad \text{iff} \quad D_M = -\sum_{m=1}^{M-1} D_m.$$

Now the sum of the first $M-1$ equations may be rewritten as

$$(1 - V_M) \sum_{d=1}^{M} b_d V_d + b_M V_M - \sum_{m=1}^{M} b_m V_m = D_M,$$

$$\sum_{d=1}^{M} b_d V_d - V_M \sum_{d=1}^{M} b_d V_d + b_M V_M - \sum_{m=1}^{M} b_m V_m = D_M.$$

Since $\sum_{d=1}^{M} b_d V_d$ and $\sum_{m=1}^{M} b_m V_m$ are the same number, we write

$$-V_M \sum_{d=1}^{M} b_d V_d + b_M V_M = D_M$$

or

$$V_M \sum_{d=1}^{M} b_d V_d - b_M V_M = -D_M,$$

which is the required M-th equation.

Theorem 5.5. *Let $M > 1$. Let the values $V_m > 0$, for $m = 1, 2, \ldots, M$, $\sum_{m=1}^{M} V_m = 1$, and derivatives D_m be given and $\sum_{d=1}^{M} D_d = 0$. We define*

$$s_m = \frac{D_m}{V_m} - \frac{D_M}{V_M},$$

for $m = 1, 2, \ldots M$. Then the solution for which the sum of squares $\sum_{m=1}^{M} b_m^2$ is minimal is given by $b_m = s_m - \bar{s}$ where $\bar{s} = \sum_{m=1}^{M} s_m / M$.

Proof. The M-th column of the reduced matrix $\mathbf{W}$ may be written as $V_M \mathbf{j}$ where V_M is the value repeated $M-1$ times in the M-th column of $\mathbf{W}$ and $\mathbf{j}$ is a column vector, $\mathbf{j} = (1, 1, \ldots, 1)'$, consisting of $M-1$ ones. Let $\mathbf{A}$ be a $(M-1) \times (M-1)$ submatrix consisting of the first $M-1$ columns of the reduced matrix $\mathbf{W}$. From theorem 2.2 it follows that $(\mathbf{A} - \mathbf{I})^{-1} V_M \mathbf{j} = V_M \mathbf{j}$.

We can also show how to multiply the right hand side of the reduced system of equations, that is, the column vector consisting of $M-1$ components, $(-D_1/V_1, -D_2/V_2, \ldots, -D_{M-1}/V_{M-1})'$, by the inverse $(\mathbf{A} - \mathrm{I})^{-1}$ of the square matrix $\mathbf{A} - \mathrm{I}$ consisting of the first $M-1$ columns of $\mathbf{W}$

$$\left[\begin{pmatrix} V_1 & V_2 & \ldots & V_{M-1} \\ \vdots & \vdots & \ddots & \vdots \\ V_1 & V_2 & \ldots & V_{M-1} \end{pmatrix} \frac{1}{\sum_{d=1}^{M-1} V_d - 1} - \mathrm{I} \right] \begin{pmatrix} -D_1/V_1 \\ \vdots \\ -D_{M-1}/V_{M-1} \end{pmatrix} =$$

$$\begin{pmatrix} D_1/V_1 - D_M/V_M \\ \vdots \\ D_{M-1}/V_{M-1} - D_M/V_M \end{pmatrix}$$

Now it is much clearer how the affine subspace of solutions is formed. If we define $s_m = D_m/V_m - D_M/V_M$, then $s_M = D_M/V_M - D_M/V_M = 0$. We may pick any parameter t and set $b_M = t$. Since $s_M = 0$, we may write $b_M = t + s_M$. For $m = 1, 2, \dots, M-1$, we get $b_m = t + s_m$.

We may introduce one more condition, such as the minimization of the distance of a solution, $b_1, b_2, \dots, b_M$, from the origin, the only purpose of such an approach being to obtain reasonable starting values for the iteration process. We minimize

$$\sum_{m=1}^{M} b_m^2 = \sum_{m=1}^{M} (t + s_m)^2$$

with respect to t and obtain the minimizing t_{min} as $t_{min} = -\sum_{m=1}^{M} s_m/M = -\bar{s}$ and then the minimizing $b_m = s_m - \bar{s}$.

If we introduce a parameter t to replace x_M to avoid confusion and set the M-th column of the reduced matrix $\mathbf{W}$ equal to $tV_M\mathbf{j}$ where tV_M is the value repeated $M-1$ times in the M-th column of $\mathbf{W}$ and $\mathbf{j}$ is a column vector, $\mathbf{j} = (1, 1, \dots, 1)'$, consisting of $M-1$ ones. From theorem 2.2 it follows that $(\mathbf{A} - \mathbf{I})^{-1} tV_M\mathbf{j} = tV_M\mathbf{j}$. Geometrically it means that the lines representing the solution of the system of equations are all parallel to each other independently of what fixed values of C were chosen to calculate the values and derivatives of the functions.

Note. We may generalize this approach to functions of more than one variable by considering partial derivatives. The left hand side of the equations is the same but the right hand side depends on with respect to which parameter we calculate the partial derivatives. The parameters $a_1, a_2, \dots, a_M$ are calculated afterwards. In the case of more than one variables it is easy to generalize theorem 5.1 as

Theorem 5.6. *Let $M > 1$. Let $\mathbf{B}$ be fixed. Let $\mathbf{c} = (\mathrm{c}_1, \mathrm{c}_2, \dots \mathrm{c}_\mathrm{K})'$ be a center point at which all the values $V_m > 0$ and $\sum_{m=1}^{M} V_m = 1$. Then the equations $\exp(a_m + \sum_{i=1}^{K} b_{im}c_i) = V_m$ are satisfied by $a_m = \ln V_m - \sum_{i=1}^{K} b_{im}c_i$ and $\sum \exp(a_m + \sum_{i=1}^{K} b_{im}c_i) = 1$.*

Proof. We solve for a_m in $\exp(a_m + \sum_{i=1}^{K} b_{im}c_i) = V_m$, which is trivial. Also $\sum V_m = 1$ implies that $\sum \exp(a_m + \sum_{i=1}^{K} b_{im}c_i) = 1$.

6. Numerical example

Our approach was originally designed to study concentrations of certain substances. Since this example would not be obvious, we took the data about the cause of death in the US. The number of inhabitants may be different each year but the proportions may follow some other pattern.

http://www.cdc.gov/nchs/nvss/mortality/lcwk9.htm
is the address from which only a small part of data was used.

The most prominent causes of death are a heart disease and cancer, other specific causes have frequency less than six percent. The data are presented in the table below. Coefficients of the restricted linear regression follow. We indicate that we used the year minus 2000. Interpolation at the mean, $C = 2004 - 2000 = 4$, of the independent variable allows us to calculate the starting values for the iteration process.

The calculated values are displayed next followed by the parameters obtained by the numerical minimization of the overall sum of squares. It is interesting to see that the overall sum of squares decreases by 4.5 percent when we minimize the overall sum of squares to calculate the coefficients of the logistic regression.

Original data

Year	Heart	Cancer	Other
1999	0.30327	0.22987	0.46686
2000	0.29578	0.23009	0.47412
2001	0.28969	0.22911	0.48120
2002	0.28526	0.22802	0.48672
2003	0.27984	0.22746	0.49270
2004	0.27214	0.23099	0.49688
2005	0.26638	0.22848	0.50515
2006	0.26037	0.23075	0.50888
2007	0.25423	0.23221	0.51356
2008	0.24957	0.22878	0.52165
2009	0.24591	0.23293	0.52116

Restricted linear regression

Coeff	Heart	Cancer	Other
a	0.29627	0.22894	0.47479
b	-0.005830	0.0002348	0.005595

y=a+b(Year-2000)

Overall sum of squares 7.038E-05

Terms exp(a+b(Year-2000))
obtained by interpolation

Coeff	Heart	Cancer	Other
a	-1.2251	-1.4864	-0.75595
b	-0.018331	0.0040491	0.014282

Overall sum of squares 6.947E-05

We can see that interpolation alone even happens to yield an overall sum of squares of differences which is smaller than the one obtained with the restricted linear regression.

7. Least squares method, case study

Interpolation gives us reasonable starting values for the purpose of minimization of the sum of squares of differences. Unfortunately, the matrix that is calculated, if we want to use the Gauss-Newton method, is singular. It means we cannot calculate the solution of the system of equation to improve the existing values of parameters. Now it is time to stop writing the paper. It will be continued later.

References

Agresti, A., 1990, *Categorical Data Analysis,* John Wiley and Sons, New York, 313.

Aitchison, J., 1986, *The Statistical Analysis of Compositional Data,* Chapman and Hall; reprinted in 2003 by The Blackburn Press.

Bukac, J., 2023, *Linear compositional regression,* Vixra: 2308.1083.

Goldberger, A.S., 1965, *Econometric Theory,* second printing John Wiley and Sons, New York London Sydney, 255-258.

Pawlowsky-Glahn, V., Egozcue, J.J., Tolosana-Delgado, R., 2015, *Modeling and Analysis of Compositional Data,* Wiley.