MATHEMATICAL ENGINEERING
TECHNICAL REPORTS

An Estimation Procedure for Contingency Table Models Based on the Nested Geometry

Yoshihiro HIROSE and Fumiyasu KOMAKI

METR 2012–12 June 2012

DEPARTMENT OF MATHEMATICAL INFORMATICS
GRADUATE SCHOOL OF INFORMATION SCIENCE AND TECHNOLOGY
THE UNIVERSITY OF TOKYO
BUNKYO-KU, TOKYO 113-8656, JAPAN

WWW page: http://www.keisu.t.u-tokyo.ac.jp/research/techrep/index.html

An Estimation Procedure for Contingency Table Models Based on the Nested Geometry

Yoshihiro HIROSE* and Fumiyasu KOMAKI

Department of Mathematical Informatics
Graduate School of Information Science and Technology
The University of Tokyo

June 4, 2012

Abstract

We propose a geometrical method for estimating the parameters of contingency tables. Our method–*bisector regression for contingency tables*–is based on a nested structure of models. The nested structure represents the variables that are independent. This means that a model includes smaller models allowing stronger independence, which also means that more parameters are eliminated in smaller models. Our method estimates parameters corresponding to the interactions of lower orders after those of higher orders are estimated or eliminated. Bisector regression generates a sequence of parameter estimates, each element of which represents a model and an estimate. The length of the sequence is much smaller than the total number of models. We describe the algorithm and show examples.

In this paper, contingency tables are considered. We introduce parametrization of multinomial distributions and propose an algorithm for estimating parameters. The proposed algorithm is bisector regression for contingency tables (BRCT). The main idea of BRCT comes from our previous works. In [6, 7], we proposed the bisector regression algorithm, which is an extension of least angle regression [4]. Least angle regression is an algorithm for parameter estimation, which is related to the l_1-regularization method (lasso, [3, 5, 11, 12]). In problems of contingency tables, our interest is to estimate parameters corresponding to interactions between factors. Factors, or random variables, are qualitative variables. Parameters are separated into groups depending on how many factors are involved. We apply the main idea of bisector regression for generalized linear regression ([6]) and Gaussian graphical models [7] to these parameter groups. We provide

*`hirose@stat.t.u-tokyo.ac.jp`

explanations for three cases: (a) two factors, (b) three factors, and (c) K factors. We first describe cases (a) and (b), and then state the algorithm for general case (c). We must distinguish the total number of factors from the number of factors involved with a parameter, especially for the general case.

The proposed algorithm BRCT is based on the geometry of dually flat space [1, 2, 9]. We consider a dually flat space of multinomial distributions. The natural parameter and expectation parameter are used as coordinate systems in this space. We estimate the natural parameter. BRCT is decided by the total number of factors K_1 and the number of factors used by estimated parameters K_2. These two numbers indicate the space where the $\mathrm{BRCT}(K_1, K_2)$ algorithm works. In case (a), we use the algorithm $\mathrm{BRCT}(2, 2)$. In case (b), the algorithms $\mathrm{BRCT}(3, 3)$ and $\mathrm{BRCT}(3, 2)$ are used. In the general case (c) in which the total number of factors is K_1, we use $\mathrm{BRCT}(K_1, K_2)$ for (a part of) $K_2 = K_1, K_1 - 1, \ldots, 2$.

BRCT generates a sequence of parameter estimates. Strictly speaking, each $\mathrm{BRCT}(K_1, K_2)$ generates a sequence. Each element of the sequence represents how variables are correlated. As shown in Section 2, $\mathrm{BRCT}(K_1, K_2)$ continuously connects to $\mathrm{BRCT}(K_1, K_2 - 1)$. This property helps us sequentially estimate parameters without any extra effort for combining algorithms. Furthermore, BRCT avoids the difficulty of combinations. The total number of combinations of independence is too large to consider when the number of factors is high. The length of a sequence generated by BRCT is the same as the total number of parameters, which is much smaller than the total number of models. BRCT helps us narrow down the candidates efficiently.

In Section 1, we consider multinomial distributions and introduce a parametrization for these. The natural parameter and expectation parameter are used in our method. The natural parameter is separated into groups depending on the number of indices. In Section 2, we propose the algorithm BRCT to estimate parameters and select interactions simultaneously. Each parameter group is estimated separately. We do not deal with all parameters equally. In Section 3, the results of our method are shown for some datasets. We give the conclusion in Section 4.

1 Introduction

We consider contingency tables and multinomial distributions. First, we explain them in the case of two factors. The natural parameter and expectation parameter are introduced. Second, we consider the case of three factors. Finally, the case of K factors is considered. Parameters in the case of K factors are confusing, and therefore, we present parameters in the case of two factors first.

We consider contingency tables of two factors X_1 and X_2, and suppose

Table 1: Notations for Two-Factor Contingency Tables. X_1: factor with $m+1$ levels, X_2: factor with $n+1$ levels, y_{ij}: the number of observations of cell (i, j).

$X_1 \backslash X_2$	0	1	$\ldots$	n	total
0	y_{00}	y_{01}	$\cdots$	y_{0n}	y_{0+}
1	y_{10}	y_{11}	$\cdots$	y_{1n}	y_{1+}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
m	y_{m0}	y_{m1}	$\cdots$	y_{mn}	y_{m+}
total	y_{+0}	y_{+1}	$\cdots$	y_{+n}	N

that they have $m+1$ levels, $X_1 = 0, 1, \ldots, m$, and $n+1$ levels, $X_2 = 0, 1, \ldots, n$, respectively (Table 1). In this case, a multinomial distribution is given as

$$f(y|\, p) = \frac{N!}{y_{00}!y_{01}! \ldots y_{mn}!} p_{00}^{y_{00}} p_{01}^{y_{01}} \ldots p_{mn}^{y_{mn}},$$

where N is the total number of observations, y_{ij} is the number of observations of cell (i, j) with constraint $\sum_{i=0}^{m} \sum_{j=0}^{n} y_{ij} = N$, and p_{ij} is the probability of cell (i, j) with constraint $\sum_{i=0}^{m} \sum_{j=0}^{n} p_{ij} = 1$.

The logarithm of the probability distribution is

$$\begin{aligned}
\log f(y|\, p) &= \sum_{i=0}^{m} \sum_{j=0}^{n} y_{ij} \log p_{ij} + \log \frac{N!}{y_{00}!y_{01}! \ldots y_{mn}!} \\
&= \sum_{i=1}^{m} \sum_{j=1}^{n} y_{ij} \log p_{ij} + \sum_{i=1}^{m} y_{i0} \log p_{i0} \\
&\quad + \sum_{j=1}^{n} y_{0j} \log p_{0j} + y_{00} \log p_{00} + \log \frac{N!}{y_{00}!y_{01}! \ldots y_{mn}!} \\
&= \sum_{i=1}^{m} \sum_{j=1}^{n} y_{ij} \log \frac{p_{ij} p_{00}}{p_{i0} p_{0j}} + \sum_{i=1}^{m} \Big(\sum_{j=0}^{n} y_{ij} \Big) \log \frac{p_{i0}}{p_{00}} \\
&\quad + \sum_{j=1}^{n} \Big(\sum_{i=0}^{m} y_{ij} \Big) \log \frac{p_{0j}}{p_{00}} + N \log p_{00} + \log \frac{N!}{y_{00}!y_{01}! \ldots y_{mn}!}.
\end{aligned}$$

We introduce the natural parameter as follows:

$$\theta_{X_1}^{i} = \log \frac{p_{i0}}{p_{00}}, \; \theta_{X_2}^{j} = \log \frac{p_{0j}}{p_{00}}, \; \theta_{X_1 X_2}^{ij} = \log \frac{p_{ij} p_{00}}{p_{i0} p_{0j}}$$

for $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n$. Let

$$\begin{aligned}
\theta &= (\theta_{X_1}; \theta_{X_2}; \theta_{X_1 X_2}) \\
&= \left(\theta_{X_1}^{1}, \theta_{X_1}^{2}, \ldots, \theta_{X_1}^{m}; \theta_{X_2}^{1}, \theta_{X_2}^{2}, \ldots, \theta_{X_1}^{n}; \theta_{X_1 X_2}^{11}, \theta_{X_1 X_2}^{12}, \ldots, \theta_{X_1 X_2}^{mn} \right),
\end{aligned}$$

where θ_{X_1} is an m-dimensional vector, θ_{X_2} is an n-dimensional vector, and $\theta_{X_1X_2}$ is an mn-dimensional vector. We apply the main idea of bisector regression to $\theta_{X_1X_2}$.

The logarithm of f is represented by the natural parameter θ as follows:

$$\begin{aligned}\log f(y|\,\theta) = &\sum_{i=1}^{m}\sum_{j=1}^{n} y_{ij}\theta^{ij}_{X_1X_2} + \sum_{i=1}^{m}\Big(\sum_{j=0}^{n} y_{ij}\Big)\theta^{i}_{X_1} \\ &+ \sum_{j=1}^{n}\Big(\sum_{i=0}^{m} y_{ij}\Big)\theta^{j}_{X_2} - \psi(\theta) + \log\frac{N!}{y_{00}!y_{01}!\dots y_{mn}!},\end{aligned}$$

where ψ is a convex function of θ, the potential function, and it is defined as

$$\begin{aligned}\psi(\theta) &= -N\log p_{00} \\ &= N\log\left(1+\sum_{i=1}^{m}\exp(\theta^{i}_{X_1})+\sum_{j=1}^{n}\exp(\theta^{j}_{X_2})+\sum_{i=1}^{m}\sum_{j=1}^{n}\exp\left(\theta^{i}_{X_1}+\theta^{j}_{X_2}+\theta^{ij}_{X_1X_2}\right)\right).\end{aligned}$$

It is not difficult to prove the second equality. In fact, from the definition of the natural parameter, we have

$$\begin{aligned}p_{i0} &= p_{00}\exp(\theta^{i}_{X_1}),\ p_{0j} = p_{00}\exp(\theta^{j}_{X_2}), \\ p_{ij} &= \frac{p_{i0}p_{0j}}{p_{00}}\exp(\theta^{ij}_{X_1X_2}) = p_{00}\exp\left(\theta^{i}_{X_1}+\theta^{j}_{X_2}+\theta^{ij}_{X_1X_2}\right)\end{aligned}$$

for $i = 1, 2, \dots, m$ and $j = 1, 2, \dots, n$. The condition $\sum_{i=0}^{m}\sum_{j=0}^{n} p_{ij} = 1$ leads to

$$p_{00}\left(1+\sum_{i=1}^{m}\exp(\theta^{i}_{X_1})+\sum_{j=1}^{n}\exp(\theta^{j}_{X_2})+\sum_{i=1}^{m}\sum_{j=1}^{n}\exp\left(\theta^{i}_{X_1}+\theta^{j}_{X_2}+\theta^{ij}_{X_1X_2}\right)\right) = 1.$$

The expectation parameter η corresponding to the natural parameter θ is defined as

$$\begin{aligned}\eta^{X_1}_{i} &= \mathrm{E}\Big[\sum_{j=0}^{n} y_{ij}\Big] = N\sum_{j=0}^{n} p_{ij}, \\ \eta^{X_2}_{j} &= \mathrm{E}\Big[\sum_{i=0}^{m} y_{ij}\Big] = N\sum_{i=0}^{m} p_{ij}, \\ \eta^{X_1X_2}_{ij} &= \mathrm{E}\Big[y_{ij}\Big] = Np_{ij}\end{aligned}$$

for $i = 1, 2, \dots, m$ and $j = 1, 2, \dots, n$. The potential function ϕ of the expectation parameter η is given by

$$\begin{aligned}\phi(\eta) &= -NH(p) \\ &= N\sum_{i=0}^{m}\sum_{j=0}^{n} p_{ij}\log p_{ij},\end{aligned}$$

where $H(p)$ is the entropy in information theory. The cell probabilities p_{ij} are represented by the expectation parameter η. In fact, we have

$$\begin{aligned}
p_{00} &= 1 - \sum_{i=1}^{m} \frac{\eta_i^{X_1}}{N} - \sum_{j=1}^{n} \frac{\eta_j^{X_2}}{N} + \sum_{i=1}^{m}\sum_{j=1}^{n} \frac{\eta_{ij}^{X_1X_2}}{N},\\
p_{i0} &= \frac{\eta_i^{X_1}}{N} - \sum_{j=1}^{n} \frac{\eta_{ij}^{X_1X_2}}{N},\\
p_{0j} &= \frac{\eta_j^{X_2}}{N} - \sum_{i=1}^{m} \frac{\eta_{ij}^{X_1X_2}}{N},\\
p_{ij} &= \frac{\eta_{ij}^{X_1X_2}}{N}.
\end{aligned}$$

Therefore, ϕ is a function of η.

Next, we consider contingency tables with three factors. Suppose that three factors–X_1, X_2, and X_3–have m_1, m_2, and m_3 levels, respectively. The natural parameter θ is defined by

$$\begin{aligned}
&\theta_{X_1}^{i_1} = \log \frac{p_{i_1 00}}{p_{000}},\ \theta_{X_2}^{i_2} = \log \frac{p_{0 i_2 0}}{p_{000}},\ \theta_{X_3}^{i_3} = \log \frac{p_{00 i_3}}{p_{000}},\\
&\theta_{X_1X_2}^{i_1i_2} = \log \frac{p_{i_1 i_2 0} p_{000}}{p_{i_1 00} p_{0 i_2 0}},\ \theta_{X_1X_3}^{i_1i_3} = \log \frac{p_{i_1 0 i_3} p_{000}}{p_{i_1 00} p_{00 i_3}},\ \theta_{X_2X_3}^{i_2i_3} = \log \frac{p_{0 i_2 i_3} p_{000}}{p_{0 i_2 0} p_{00 i_3}},\\
&\theta_{X_1X_2X_3}^{i_1i_2i_3} = \log \frac{p_{i_1 i_2 i_3} p_{i_1 00} p_{0 i_2 0} p_{00 i_3}}{p_{i_1 i_2 0} p_{i_1 0 i_3} p_{0 i_2 i_3} p_{000}}
\end{aligned}$$

for $i_1 = 1, 2, \ldots, m_1$, $i_2 = 1, 2, \ldots, m_2$, and $i_3 = 1, 2, \ldots, m_3$. The logarithm of the probability function f is represented with respect to the natural parameter as follows:

$$\begin{aligned}
\log f(y|\,\theta) = &\sum_{i_1=1}^{m_1}\sum_{i_2=1}^{m_2}\sum_{i_3=1}^{m_3} y_{i_1i_2i_3}\theta_{X_1X_2X_3}^{i_1i_2i_3} + \sum_{i_1=1}^{m_1}\sum_{i_2=1}^{m_2}\Big(\sum_{i_3=0}^{m_3} y_{i_1i_2i_3}\Big)\theta_{X_1X_2}^{i_1i_2}\\
&+ \sum_{i_1=1}^{m_1}\sum_{i_3=1}^{m_3}\Big(\sum_{i_2=0}^{m_2} y_{i_1i_2i_3}\Big)\theta_{X_1X_3}^{i_1i_3} + \sum_{i_2=1}^{m_2}\sum_{i_3=1}^{m_3}\Big(\sum_{i_1=0}^{m_1} y_{i_1i_2i_3}\Big)\theta_{X_2X_3}^{i_2i_3}\\
&+ \sum_{i_1=1}^{m_1}\Big(\sum_{i_2=0}^{m_2}\sum_{i_3=0}^{m_3} y_{i_1i_2i_3}\Big)\theta_{X_1}^{i_1} + \sum_{i_2=1}^{m_2}\Big(\sum_{i_1=0}^{m_1}\sum_{i_3=0}^{m_3} y_{i_1i_2i_3}\Big)\theta_{X_2}^{i_2}\\
&+ \sum_{i_3=1}^{m_3}\Big(\sum_{i_1=0}^{m_1}\sum_{i_2=0}^{m_2} y_{i_1i_2i_3}\Big)\theta_{X_3}^{i_3} - \psi(\theta) + \log \frac{N!}{y_{000}!y_{001}!\cdots y_{m_1m_2m_3}!},
\end{aligned}$$

where $\psi(\theta) = -N \log p_{000}$. Similar to the case of two factors, all $p_{i_1i_2i_3}$ can be represented by the natural parameter θ:

$$
\begin{aligned}
&p_{i_100} = p_{000}\exp(\theta^{i_1}_{X_1}),\ p_{0i_20} = p_{000}\exp(\theta^{i_2}_{X_2}),\ p_{00i_3} = p_{000}\exp(\theta^{i_3}_{X_3}),\\
&p_{i_1i_20} = p_{000}\exp\left(\theta^{i_1}_{X_1} + \theta^{i_2}_{X_2} + \theta^{i_1i_2}_{X_1X_2}\right),\ p_{i_10i_3} = p_{000}\exp\left(\theta^{i_1}_{X_1} + \theta^{i_3}_{X_3} + \theta^{i_1i_3}_{X_1X_3}\right),\\
&p_{0i_2i_3} = p_{000}\exp\left(\theta^{i_2}_{X_2} + \theta^{i_3}_{X_3} + \theta^{i_2i_3}_{X_2X_3}\right),\\
&p_{i_1i_2i_3} = p_{000}\exp\left(\theta^{i_1}_{X_1} + \theta^{i_2}_{X_2} + \theta^{i_3}_{X_3} + \theta^{i_1i_2}_{X_1X_2} + \theta^{i_1i_3}_{X_1X_3} + \theta^{i_2i_3}_{X_2X_3}\right),\\
&p_{000} = \Big\{1 + \sum\exp(\theta^{i_1}_{X_1}) + \sum\exp(\theta^{i_2}_{X_2}) + \sum\exp(\theta^{i_3}_{X_3})\\
&\qquad\qquad + \exp\left(\theta^{i_1}_{X_1} + \theta^{i_2}_{X_2} + \theta^{i_1i_2}_{X_1X_2}\right) + \exp\left(\theta^{i_1}_{X_1} + \theta^{i_3}_{X_3} + \theta^{i_1i_3}_{X_1X_3}\right)\\
&\qquad\qquad\quad + \exp\left(\theta^{i_2}_{X_2} + \theta^{i_3}_{X_3} + \theta^{i_2i_3}_{X_2X_3}\right)\\
&\qquad\qquad\qquad + \exp\left(\theta^{i_1}_{X_1} + \theta^{i_2}_{X_2} + \theta^{i_3}_{X_3} + \theta^{i_1i_2}_{X_1X_2} + \theta^{i_1i_3}_{X_1X_3} + \theta^{i_2i_3}_{X_2X_3}\right)\Big\}^{-1}.
\end{aligned}
$$

The expectation parameter η corresponding to the natural parameter θ is given by

$$
\begin{aligned}
&\eta^{X_1}_{i_1} = N\sum_{i_2=0}^{m_2}\sum_{i_3=0}^{m_3} p_{i_1i_2i_3},\ \eta^{X_2}_{i_2} = N\sum_{i_1=0}^{m_1}\sum_{i_3=0}^{m_3} p_{i_1i_2i_3},\ \eta^{X_3}_{i_3} = N\sum_{i_1=0}^{m_1}\sum_{i_2=0}^{m_2} p_{i_1i_2i_3},\\
&\eta^{X_1X_2}_{i_1i_2} = \sum_{i_3=0}^{m_3} p_{i_1i_2i_3},\ \eta^{X_1X_3}_{i_1i_3} = \sum_{i_2=0}^{m_2} p_{i_1i_2i_3},\ \eta^{X_2X_3}_{i_2i_3} = \sum_{i_1=0}^{m_1} p_{i_1i_2i_3},\\
&\eta^{X_1X_2X_3}_{i_1i_2i_3} = p_{i_1i_2i_3}
\end{aligned}
$$

for $i_1 = 1, 2, \ldots, m_1$, $i_2 = 1, 2, \ldots, m_2$, and $i_3 = 1, 2, \ldots, m_3$.

We consider the case of K factors: an $(m_1+1)\times(m_2+1)\times\cdots\times(m_K+1)$-contingency table. Let $a \in \{1, 2, \ldots, K\}$ and $i_a \in \{0, 1, \ldots, m_a\}$, where the latter is an index of factors, indicating the level of factor X_a. Before introducing the parameters, we prepare the notations. For $l \leq h \leq K$, define

$$
V_l(i_{a_1}, i_{a_2}, \ldots, i_{a_h}) = \left\{\ p_{i'_1i'_2\cdots i'_K}\ \middle|\ \begin{array}{l} h-l \text{ indices are decided by } (i_{a_1}, i_{a_2}, \ldots, i_{a_h}),\\ l \text{ elements of } (i'_{a_1}, i'_{a_2}, \ldots, i'_{a_h}) \text{ are } 0,\\ i'_a = 0 \text{ for } a \notin \{a_1, a_2, \ldots, a_h\} \end{array} \right\}.
$$

For example, when $K = 3, h = 2, l = 1, a_1 = 1$, and $a_2 = 3$, we have $V_l(i_{a_1}, i_{a_2}) = V_1(i_1, i_3) = \{p_{i_100}, p_{00i_3}\}$. The natural parameter θ is defined as

$$
\theta^{i_{a_1}i_{a_2}\ldots i_{a_q}}_{X_{a_1}X_{a_2}\ldots X_{a_q}} = \log\frac{\left(\prod_{p^{(q)}\in V_0(i_{a_1}, i_{a_2}, \ldots, i_{a_q})} p^{(q)}\right)\left(\prod_{p^{(q-2)}\in V_2(i_{a_1}, i_{a_2}, \ldots, i_{a_q})} p^{(q-2)}\right)\cdots}{\left(\prod_{p^{(q-1)}\in V_1(i_{a_1}, i_{a_2}, \ldots, i_{a_q})} p^{(q-1)}\right)\left(\prod_{p^{(q-3)}\in V_3(i_{a_1}, i_{a_2}, \ldots, i_{a_q})} p^{(q-3)}\right)\cdots},
$$

for $q = 1, 2, \ldots, K$ and $1 \leq i_a \leq m_a\,(a = 1, \ldots, K)$. For example, when $K = q = 3$, we have

$$
\theta^{i_1i_2i_3}_{X_1X_2X_3} = \log\frac{p_{i_1i_2i_3}p_{i_100}p_{0i_20}p_{00i_3}}{p_{i_1i_20}p_{i_10i_3}p_{0i_2i_3}p_{000}},
$$

which is the natural parameter in the case of three factors. The expectation parameter η corresponding to the natural parameter θ is given by

$$\begin{aligned}\eta_{i_{a_1}i_{a_2}\dots i_{a_q}}^{X_{a_1}X_{a_2}\dots X_{a_q}} &= \mathrm{E}\left[\sum_{a\notin\{a_1,a_2,\dots,a_q\}}\sum_{i_a=0}^{m_a} y_{i_1i_2\dots i_K}\right]\\ &= N\sum_{a\notin\{a_1,a_2,\dots,a_q\}}\sum_{i_a=0}^{m_a} p_{i_1i_2\dots i_K}\end{aligned}$$

for $q = 1, 2, \dots, K$.

2 Proposed Algorithm

First, we provide an algorithm for two-factor contingency tables. The simplest model that we assume is the independent model: for all $i = 1, 2, \dots, m$ and $j = 1, 2, \dots, n$,

$$p_{ij}p_{00} = p_{i0}p_{0j},$$

which is equivalent to

$$\theta_{X_1X_2}^{ij} = \log\frac{p_{ij}p_{00}}{p_{i0}p_{0j}} = 0$$

for all $i = 1, 2, \dots, m$ and $j = 1, 2, \dots, n$. Let S denote the dually flat space of all multinomial distributions. A submodel N is defined by

$$N = \left\{\theta |\, \theta_{X_1X_2}^{ij} = 0,\, 1 \le i \le m,\, 1 \le j \le n\right\};$$

this is the independent model (Figure 1). Let $\hat{\theta}_{\mathrm{MLE}}$ and $\hat{\theta}_0$ denote the MLE of S and the MLE of the submodel N, respectively. We define M by the m-flat subspace that intersects orthogonally with N at $\hat{\theta}_0$. It is known that M includes both $\hat{\theta}_{\mathrm{MLE}}$ and $\hat{\theta}_0$. Points in M can be represented as

$$((\hat{\eta}_0)_i^{X_1}, (\hat{\eta}_0)_j^{X_2}; \theta_{X_1X_2}^{ij}),$$

where $(\hat{\eta}_0)_i^{X_1}$ and $(\hat{\eta}_0)_j^{X_2}$ are a part of the m-affine coordinate of $\hat{\theta}_0$. Our algorithm BRCT(2) works within M. We apply our method to only $\theta_{X_1X_2}^{ij}$ and fix the $\eta_i^{X_1}$-coordinate and $\eta_j^{X_2}$-coordinate. As a notation in the case of two factors, $((\hat{\eta}_0)_i^{X_1}, (\hat{\eta}_0)_j^{X_2}; \theta_{X_1X_2}^{ij})$ is represented by θ. A sequence of parameter estimates, $\hat{\theta}_{(0)}, \hat{\theta}_{(1)}, \dots, \hat{\theta}_{(mn)}$, is generated by our algorithm.

Before describing the algorithm, we introduce some submodels. Let

$$\begin{aligned}M\left(I\right) &= \left\{\theta \mid \theta_{X_1X_2}^{i'j'} = 0,\, (i',j') \notin I\right\},\\ M\left((i,j),s,I\right) &= \left\{\theta \mid \theta_{X_1X_2}^{ij} = s,\, \theta_{X_1X_2}^{i'j'} = 0,\, (i',j') \notin I\right\}\end{aligned}$$

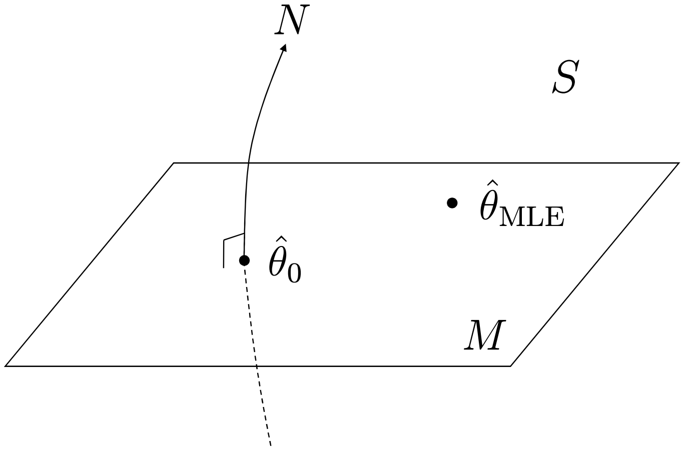

Figure 1: Submodels for BRCT(2). S: space of all multinomial distributions, $\hat{\theta}_{\text{MLE}}$: MLE of S, N: independent model defined by $N = \{\theta | \, \theta^{ij}_{X_1X_2} = 0, \, 1 \le i \le m, \, 1 \le j \le n\}$, $\hat{\theta}_0$: MLE of the submodel N, M: m-flat subspace that intersects orthogonally with N at $\hat{\theta}_0$. M is known to include both $\hat{\theta}_{\text{MLE}}$ and $\hat{\theta}_0$. BRCT(2) works within M, and it estimates $\theta^{ij}_{X_1X_2}$ for $1 \le i \le m$, $1 \le j \le n$. Points in M can be represented as $((\hat{\eta}_0)^{X_1}_i, (\hat{\eta}_0)^{X_2}_j; \theta^{ij}_{X_1X_2})$, where $(\hat{\eta}_0)^{X_1}_i$ and $(\hat{\eta}_0)^{X_2}_j$ are m-affine coordinates of $\hat{\theta}_0$.

for $I \subseteq \{(i', j') | \, 1 \le i' \le m, \, 1 \le j' \le n\}$, $(i, j) \in I$, and $s \in \mathbf{R}$ (Figure 2).

The algorithm **BRCT(2)** (BRCT with two factors) is given as follows. Steps 2 to 6 are iterated.

```
BRCT(2)

   input: observation $y_{ij}$ of each cell $(i, j)$
  output: parameter estimates $\hat{\theta}_{(0)}, \hat{\theta}_{(1)}, \ldots, \hat{\theta}_{(mn)}$
      1. Let $\hat{\theta}_{(0)} := \hat{\theta}_{\text{MLE}}$, $I := \{(i, j) | \, 1 \le i \le m, \, 1 \le j \le n\}$, and $k := 0$.
      2. Calculate the MLE $\bar{\theta}^{-ij}_{(k)}$ of the model $M \, (I \setminus \{(i, j)\})$ for $(i, j) \in I$.
      3. Find $t^* := \min_{(i,j) \in I} D\left(\hat{\theta}_{(k)} \mid \bar{\theta}^{-ij}_{(k)}\right)$ and
         $(i^*, j^*) := \arg\min_{(i,j) \in I} D\left(\hat{\theta}_{(k)} \mid \bar{\theta}^{-ij}_{(k)}\right)$.
      4. For $(i, j) \in I$, calculate $s^*_{ij}$ and $\tilde{\theta}^{-ij}_{(k)} \in l^{-ij}_{(k)}$ satisfying both
         $D(\hat{\theta}_{(k)} \mid \tilde{\theta}^{-ij}_{(k)}) = t^*$ and $\tilde{\theta}^{-ij}_{(k)} \in M((i, j), s^*_{ij}, I)$.
      5. Let $\hat{\theta}^{ij}_{(k+1)} := s^*_{ij}$ for $(i, j) \in I$ and $\hat{\theta}^{ij}_{(k+1)} := 0$ for $(i, j) \notin I$.
      6. If $k + 1 < mn - 1$, then go to step 2 with $k := k + 1, I :=$
         $I \setminus \{(i^*, j^*)\}$. If $k + 1 = mn - 1$, then go to step 7.
```

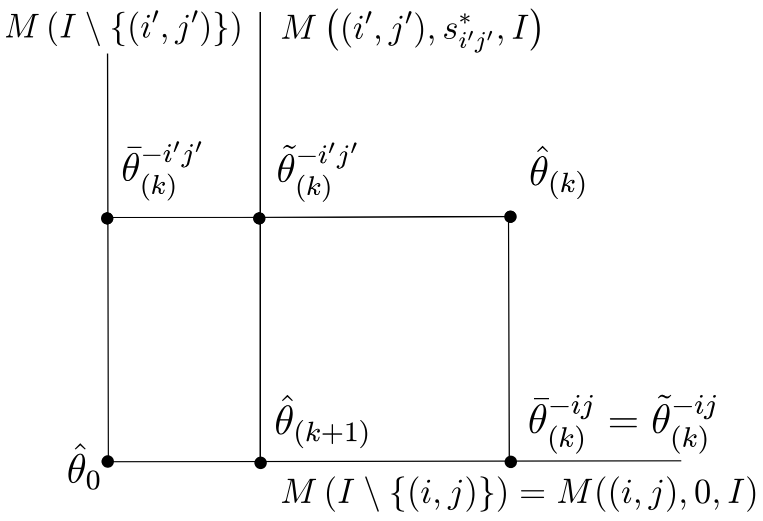
$M(I \setminus \{(i', j')\})$
$M\left((i', j'), s^*_{i'j'}, I\right)$
$\bar{\theta}^{-i'j'}_{(k)}$
$\tilde{\theta}^{-i'j'}_{(k)}$
$\hat{\theta}_{(k)}$
$\hat{\theta}_{(k+1)}$
$\bar{\theta}^{-ij}_{(k)} = \tilde{\theta}^{-ij}_{(k)}$
$\hat{\theta}_0$
$M(I \setminus \{(i, j)\}) = M((i, j), 0, I)$

Figure 2: Update of an Estimator in BRCT(2). $\hat{\theta}_{(k)}$: kth estimate of θ, $\hat{\theta}_{(k+1)}$: $(k+1)$th estimate of θ, $\hat{\theta}_0$: the MLE of the model N, $M((i,j),s,I) = \left\{\theta \,\middle|\, \theta^{ij} = s,\ \theta^{i'j'} = 0\ ((i',j') \notin I)\right\}$, $\bar{\theta}^{-ij}_{(k)}$: the MLE of the submodel $M(I \setminus \{(i,j)\})$, $\tilde{\theta}^{-ij}_{(k)}$: the projection of $\hat{\theta}_{(k)}$ to the submodel $M((i,j), s^*_{ij}, I)$. In this figure, $M(I \setminus \{(i,j)\})$ is nearer to $\hat{\theta}_{(k)}$ than $M(I \setminus \{(i',j')\})$. The divergence to $\tilde{\theta}^{-i'j'}_{(k)}$ from $\hat{\theta}_{(k)}$ is the same as the divergence to $\bar{\theta}^{-ij}_{(k)}$. $\hat{\theta}_{(k+1)}$ is defined as the intersection of $M((i,j),0,I)$ and $M((i',j'), s^*_{i'j'}, I)$.

7. Let $\hat{\theta}_{(mn)} := \hat{\theta}_0$. Stop the algorithm.

Note that in step 5, $\hat{\theta}^{i^*j^*}_{(k+1)} = 0$ for $(i^*, j^*) \in I$ that was obtained in step 3. This fact indicates that one element of θ becomes 0 in each iteration and that our method selects covariance models sequentially.

Next, we propose an algorithm for three-factor contingency tables. We consider an $m_1 \times m_2 \times m_3$-contingency table. A part of the natural parameter is estimated first, and the remainder is estimated by an analogy of BRCT(2) thereafter. We estimate $\theta^{i_1 i_2 i_3}_{X_1 X_2 X_3}$ by bisector regression under the condition that complementary elements of the expectation parameter, $\eta^{X_1}_{i_1}, \eta^{X_2}_{i_2}, \eta^{X_3}_{i_3}, \eta^{X_1 X_2}_{i_1 i_2}, \eta^{X_1 X_3}_{i_1 i_3}, \eta^{X_2 X_3}_{i_2 i_3}$, are fixed at the values of the MLE. Note that two types of elements of the natural parameters, $\theta^{i_1 i_2}_{X_1 X_2}$ and $\theta^{i_1 i_2 i_3}_{X_1 X_2 X_3}$, are not dealt with equally. The former type, $\theta^{i_1 i_2}_{X_1 X_2}$, is estimated by an analogy of BRCT(2) after the latter type, $\theta^{i_1 i_2 i_3}_{X_1 X_2 X_3}$, is estimated. We define some submodels, similar to N and M in the case of two factors. A submodel $N^{[3]}$ is defined by

$$N^{[3]} = \left\{\theta \middle| \, \theta^{i_1 i_2 i_3}_{X_1 X_2 X_3} = 0,\ 1 \leq i_1 \leq m_1,\ 1 \leq i_2 \leq m_2,\ 1 \leq i_3 \leq m_3 \right\}.$$

Let $\hat{\theta}_{\mathrm{MLE}}$ and $\hat{\theta}_0^{[3]}$ denote the MLE of S and the MLE of the submodel $N^{[3]}$, respectively. We define $M^{[3]}$ as the m-flat subspace that intersects orthogonally with $N^{[3]}$ at $\hat{\theta}_0^{[3]} \in N^{[3]}$. It is known that $M^{[3]}$ includes both $\hat{\theta}_{\mathrm{MLE}}$ and $\hat{\theta}_0^{[3]}$, similar to the case of two factors. Points in $M^{[3]}$ can be represented as

$$\left((\hat{\eta}_0^{[3]})_{i_1}^{X_1}, (\hat{\eta}_0^{[3]})_{i_2}^{X_2}, (\hat{\eta}_0^{[3]})_{i_3}^{X_3}; (\hat{\eta}_0^{[3]})_{i_1 i_2}^{X_1 X_2}, (\hat{\eta}_0^{[3]})_{i_1 i_3}^{X_1 X_3}, (\hat{\eta}_0^{[3]})_{i_2 i_3}^{X_2 X_3}; \theta_{X_1 X_2 X_3}^{i_1 i_2 i_3} \right),$$

where the η-part, $(\hat{\eta}_0^{[3]})_{i_1}^{X_1}, (\hat{\eta}_0^{[3]})_{i_2}^{X_2}, (\hat{\eta}_0^{[3]})_{i_3}^{X_3}, (\hat{\eta}_0^{[3]})_{i_1 i_2}^{X_1 X_2}, (\hat{\eta}_0^{[3]})_{i_1 i_3}^{X_1 X_3}, (\hat{\eta}_0^{[3]})_{i_2 i_3}^{X_2 X_3}$, is a part of the m-affine coordinate of $\hat{\theta}_0^{[3]}$. Our algorithm BRCT(3) first works within $M^{[3]}$. We apply bisector regression to $\theta_{X_1 X_2 X_3}^{i_1 i_2 i_3}$ under the condition that a part of the expectation parameter, $\eta_{i_1}^{X_1}, \eta_{i_2}^{X_2}, \eta_{i_3}^{X_3}, \eta_{i_1 i_2}^{X_1 X_2}, \eta_{i_1 i_3}^{X_1 X_3}, \eta_{i_2 i_3}^{X_2 X_3}$, is fixed. Let $\theta^{[3]}$ denote $(\theta_{X_1 X_2 X_3}^{i_1 i_2 i_3})$. A sequence of estimates of $\theta^{[3]}$, $\hat{\theta}_{(0)}, \hat{\theta}_{(1)}, \ldots, \hat{\theta}_{(d_3)}$, is generated by BRCT(3), where $d_3 = \prod_i m_i = m_1 m_2 m_3$ is the length of $\theta^{[3]}$. Strictly speaking, parameter estimates $\hat{\theta}_{(k)}$ should be written as $\hat{\theta}_{(k)}^{[3]}$ so that they can be distinguished from $\hat{\theta}_{(k)}$ in the case of two factors. However, the superscript [3] is omitted for simplicity. We define other submodels $M^{[3]}(I)$ and $M^{[3]}((i_1, i_2, i_3), s, I)$ in $M^{[3]}$ as

$$\begin{aligned} M^{[3]}(I) &= \left\{ \theta \mid \theta_{X_1 X_2 X_3}^{i'_1 i'_2 i'_3} = 0,\ (i'_1, i'_2, i'_3) \notin I \right\}, \\ M^{[3]}((i_1, i_2, i_3), s, I) &= \left\{ \theta \mid \theta_{X_1 X_2 X_3}^{i_1 i_2 i_3} = s,\ \theta_{X_1 X_2 X_3}^{i'_1 i'_2 i'_3} = 0,\ (i'_1, i'_2, i'_3) \notin I \right\} \end{aligned}$$

for $I \subseteq \{(i'_1, i'_2, i'_3) \mid 1 \le i'_1 \le m_1,\ 1 \le i'_2 \le m_2,\ 1 \le i'_3 \le m_3\}$, $(i_1, i_2, i_3) \in I$, and $s \in \mathbf{R}$.

We provide an algorithm for the setting where contingency tables have three factors. The algorithm **BRCT(3)** (BRCT with three factors) is given as follows. Steps 2 to 6 are iterated, and run BRCT$(3, 2)$ after the iterations. The algorithm BRCT$(3, 2)$ is an analogy of BRCT(2) for the case of three factors, and it is explained after BRCT(3) is described.

BRCT(3)

```
 input: observation $y_{i_1 i_2 i_3}$ of each cell $(i_1, i_2, i_3)$
output: parameter estimates $\hat{\theta}_{(0)}, \hat{\theta}_{(1)}, \ldots, \hat{\theta}_{(d_3)}$
    1. Let $\hat{\theta}_{(0)} := \hat{\theta}_{\mathrm{MLE}}$, $I := \{(i_1, i_2, i_3) \mid 1 \le i_1 \le m_1,\ 1 \le i_2 \le m_2,\ 1 \le i_3 \le m_3\}$, and $k := 0$.
    2. Calculate the MLE $\bar{\theta}_{(k)}^{-i_1 i_2 i_3}$ of the model $M^{[3]}(I \setminus \{(i_1, i_2, i_3)\})$
       for all $(i_1, i_2, i_3) \in I$.
    3. Find $t^* := \min_{(i_1, i_2, i_3) \in I} D\left(\hat{\theta}_{(k)} \mid \bar{\theta}_{(k)}^{-i_1 i_2 i_3}\right)$, and
       $(i_1^*, i_2^*, i_3^*) := \arg\min_{(i_1, i_2, i_3) \in I} D\left(\hat{\theta}_{(k)} \mid \bar{\theta}_{(k)}^{-i_1 i_2 i_3}\right)$.
```

4. For all $(i_1, i_2, i_3) \in I$, calculate $s^*_{i_1 i_2 i_3}$ and $\tilde{\theta}^{-i_1 i_2 i_3}_{(k)} \in l^{-i_1 i_2 i_3}_{(k)}$ satisfying both $D(\hat{\theta}_{(k)} \mid \tilde{\theta}^{-i_1 i_2 i_3}_{(k)}) = t^*$ and $\tilde{\theta}^{-i_1 i_2 i_3}_{(k)} \in M((i_1, i_2, i_3), s^*_{i_1 i_2 i_3}, I)$.
5. Let $\hat{\theta}^{i_1 i_2 i_3}_{(k+1)} := s^*_{i_1 i_2 i_3}$ for $(i_1, i_2, i_3) \in I$ and $\hat{\theta}^{i_1 i_2 i_3}_{(k+1)} := 0$ for $(i_1, i_2, i_3) \notin I$.
6. If $k + 1 < d_3 - 1$, then go to step 2 with $k := k + 1, I := I \setminus \{(i^*_1, i^*_2, i^*_3)\}$. If $k + 1 = d_3 - 1$, then go to step 7.
7. Let $\hat{\theta}_{(d_3)} := \hat{\theta}^{[3]}_0$. Run **BRCT(3, 2)** with $\hat{\theta}^{[3]}_0$ as the start point.

Note that other elements of the natural parameter are decided by the elements of the natural and expectation parameters that may be estimated or fixed.

The algorithm **BRCT(3, 2)** (BRCT with 3 factors and 2-indexed parameters) is a three-factor version of BRCT(2). BRCT(2) and BRCT(3) are thus denoted as BRCT(2, 2) and BRCT(3, 3), respectively. Let $d_{[3,2]} = \sum_{1 \leq \tau(1) < \tau(2) \leq 3} m_{\tau(1)} m_{\tau(2)} = m_1 m_2 + m_1 m_3 + m_2 m_3$ denote the number of parameters to be estimated by BRCT(3, 2). We define some submodels in $N^{[3]}$, which are similar to $N^{[3]}$ and $M^{[3]}$ of BRCT(3). A submodel $N^{[3,2]}$ is defined by

$$N^{[3,2]} = \Big\{ \theta \big| \, \theta^{i_1 i_2 i_3}_{X_1 X_2 X_3} = 0, \, \theta^{i_1 i_2}_{X_1 X_2} = 0, \, \theta^{i_1 i_3}_{X_1 X_3} = 0, \, \theta^{i_2 i_3}_{X_2 X_3} = 0, \\ 1 \leq i_1 \leq m_1, \, 1 \leq i_2 \leq m_2, \, 1 \leq i_3 \leq m_3 \Big\}.$$

Let $\hat{\theta}^{[3,2]}_{\mathrm{MLE}}$ and $\hat{\theta}^{[3,2]}_0$ denote the MLE of the model $N^{[3]}$ and the model $N^{[3,2]}$, respectively. Recall that $\hat{\theta}^{[3]}_0$ is the MLE of the model $N^{[3]}$ too, that is, $\hat{\theta}^{[3,2]}_{\mathrm{MLE}} = \hat{\theta}^{[3]}_0$. We define $M^{[3,2]}$ as the m-flat subspace of $N^{[3]}$ that intersects orthogonally with $N^{[3,2]}$ at $\hat{\theta}^{[3,2]}_0 \in N^{[3,2]}$. It is known that $M^{[3,2]}$ includes both $\hat{\theta}^{[3,2]}_{\mathrm{MLE}}$ and $\hat{\theta}^{[3,2]}_0$. Points in $M^{[3,2]}$ can be represented as

$$\left((\hat{\eta}^{[3,2]}_0)^{X_1}_{i_1}, (\hat{\eta}^{[3,2]}_0)^{X_2}_{i_2}, (\hat{\eta}^{[3,2]}_0)^{X_3}_{i_3}; \theta^{i_1 i_2}_{X_1 X_2}, \theta^{i_1 i_3}_{X_1 X_3}, \theta^{i_2 i_3}_{X_2 X_3}; 0 \right).$$

The model $M^{[3,2]}$ is the model that BRCT(3, 2) works within. Let $\theta^{[3,2]}$ denote $(\theta^{i_1 i_2}_{X_1 X_2}, \theta^{i_1 i_3}_{X_1 X_3}, \theta^{i_2 i_3}_{X_2 X_3})$. A sequence of estimates of $\theta^{[3,2]}$– $\hat{\theta}_{(0)}, \hat{\theta}_{(1)}, \ldots, \hat{\theta}_{(d_3)}$– is generated by BRCT(3, 2). Strictly speaking, parameter estimates $\hat{\theta}_{(k)}$ should be written as $\hat{\theta}^{[3,2]}_{(k)}$, but the superscript $[3, 2]$ is omitted for simplicity. We define the submodels of $N^{[3]}$, named $M^{[3]}(I)$ and $M^{[3]}((i_1, i_2), s, I)$, as

$$\begin{aligned}M^{[3,2]}\left(I\right) = \Big\{\theta \mid\ & \theta_{X_1X_2}^{i_1'i_2'} = 0,\, \theta_{X_1X_3}^{i_1'i_3'} = 0,\, \theta_{X_2X_3}^{i_2'i_3'} = 0, \\ & \qquad\qquad (i_1', i_2') \notin I,\, (i_1', i_3') \notin I,\, (i_2', i_3') \notin I\Big\}, \\ M^{[3,2]}\left((i_1, i_2), s, I\right) = \Big\{\theta \mid\ & \theta_{X_1X_2}^{i_1i_2} = s,\, \theta_{X_1X_2}^{i_1'i_2'} = 0,\, \theta_{X_1X_3}^{i_1'i_3'} = 0,\, \theta_{X_2X_3}^{i_2'i_3'} = 0, \\ & \qquad\qquad (i_1', i_2') \notin I,\, (i_1', i_3') \notin I,\, (i_2', i_3') \notin I\Big\},\end{aligned}$$

respectively, for $I \subseteq \Big\{(i_1', i_2'), (i_1', i_3'), (i_2', i_3') | \, 1 \le i_1' \le m,\, 1 \le i_2' \le m_2,\, 1 \le i_3' \le m_3\Big\}$, $(i_1, i_2) \in I$, and $s \in \mathbf{R}$. The submodels $M^{[3,2]}\left((i_1, i_3), s, I\right)$ and $M^{[3,2]}\left((i_2, i_3), s, I\right)$ are defined in the same way as $M^{[3,2]}\left((i_1, i_2), s, I\right)$. In the definitions of both $M^{[3,2]}\left(I\right)$ and $M^{[3,2]}\left((i_1, i_2), s, I\right)$, the condition $\theta_{X_1X_2X_3}^{i_1i_2i_3} = 0$ is omitted. Note that $M^{[3,2]}\left(I\right)$ and $M^{[3,2]}\left((i_1, i_2), s, I\right)$ are subspaces of $N^{[3]}$.

The algorithm $\mathrm{BRCT}(3, 2)$ is described as follows.

```
BRCT(3,2)

 input: observation $y_{i_1i_2i_3}$ of each cell $(i_1, i_2, i_3)$
output: parameter estimates $\hat{\theta}_{(0)}, \hat{\theta}_{(1)}, \ldots, \hat{\theta}_{(d_{[3,2]})}$

    1. Let $\hat{\theta}_{(0)} := \hat{\theta}_{\mathrm{MLE}}^{[3,2]}$, $I = I^{[3,2]} := \{(i_{\tau(1)}, i_{\tau(2)}) | \, 1 \le \tau(1) < \tau(2) \le 3\}$, and $k := 0$.
    2. Calculate the MLE $\bar{\theta}_{(k)}^{-i_{\tau(1)}i_{\tau(2)}}$ of the model $M^{[3,2]}\left(I \setminus \{(i_{\tau(1)}, i_{\tau(2)})\}\right)$ for all $(i_{\tau(1)}, i_{\tau(2)}) \in I$.
    3. Find $t^* := \min_{(i_{\tau(1)}, i_{\tau(2)}) \in I} D\left(\hat{\theta}_{(k)} \mid \bar{\theta}_{(k)}^{-i_{\tau(1)}i_{\tau(2)}}\right)$ and $(i_{\tau(1)}^*, i_{\tau(2)}^*) := \arg\min_{(i_{\tau(1)}, i_{\tau(2)}) \in I} D\left(\hat{\theta}_{(k)} \mid \bar{\theta}_{(k)}^{-i_{\tau(1)}i_{\tau(2)}}\right)$.
    4. For $(i_{\tau(1)}, i_{\tau(2)}) \in I$, calculate $s_{i_{\tau(1)}i_{\tau(2)}}^*$ and $\tilde{\theta}_{(k)}^{-i_{\tau(1)}i_{\tau(2)}} \in l_{(k)}^{-i_{\tau(1)}i_{\tau(2)}}$ satisfying both $D(\hat{\theta}_{(k)} \mid \tilde{\theta}_{(k)}^{-i_{\tau(1)}i_{\tau(K_2)}}) = t^*$ and $\tilde{\theta}_{(k)}^{-i_{\tau(1)}i_{\tau(K_2)}} \in M^{[3,2]}((i_{\tau(1)}, i_{\tau(2)}), s_{i_{\tau(1)}i_{\tau(2)}}^*, I)$.
    5. Let $\hat{\theta}_{(k+1)}^{i_{\tau(1)}i_{\tau(2)}} := s_{i_{\tau(1)}i_{\tau(2)}}^*$ for $(i_{\tau(1)}, i_{\tau(2)}) \in I$ and $\hat{\theta}_{(k+1)}^{i_{\tau(1)}i_{\tau(2)}} := 0$ for $(i_{\tau(1)}, i_{\tau(2)}) \notin I$.
    6. If $k + 1 < d_{[3,2]} - 1$, then go to step 2 with $k := k + 1$, $I := I \setminus \{(i_{\tau(1)}^*, i_{\tau(2)}^*)\}$. If $k + 1 = d_{[3,2]} - 1$, then go to step 7.
    7. Let $\hat{\theta}_{(d_{[3,2]})} := \hat{\theta}_0^{[3,2]}$. Stop the algorithm.
```

In step 1, $I^{[3,2]}$ is strictly given by $I^{[3,2]} := \{(i_{\tau(1)}, i_{\tau(2)}) | \, 1 \le i_{\tau(1)} \le m_{\tau(1)},\, 1 \le i_{\tau(2)} \le m_{\tau(2)},\, 1 \le \tau(1) < \tau(2) \le 3\}$. Note that $(\tau(1), \tau(2)) \in \{(1, 2), (1, 3), (2, 3)\}$.

In the case of three factors, BRCT$(3,2)$ is completed after $d_{[3,2]}$ iterations of steps 2 to 6, and thus, BRCT(3) is completed.

Finally, we provide an algorithm for the general case where contingency tables have K_1 factors. We consider an $m_1 \times \cdots \times m_{K_1}$-contingency table. For $K_2 \leq K_1$, we propose an algorithm **BRCT(K_1, K_2)** (BRCT with K_1 factors and K_2-indexed parameters), which is followed by BRCT(K_1, K_2-1) unless $K_2 = 2$. Let $d_{[K_1,K_2]} = \sum_{1 \leq \tau(1) < \cdots < \tau(K_2) \leq K_1} m_{\tau(1)} \ldots m_{\tau(K_2)}$ denote the number of parameters to be estimated by BRCT(K_1, K_2). We define some submodels. A submodel $N^{[K_1,K_2]}$ is defined by

$$\begin{aligned} N^{[K_1,K_2]} = &\left\{\theta |\, \theta^{i_{\tau(1)}\cdots i_{\tau(K_2)}}_{X_{\tau(1)}\cdots X_{\tau(K_2)}} = 0,\, 1 \leq \tau(1) < \cdots < \tau(K_2) \leq K_1\right\} \\ &\cap \left\{\theta |\, \theta^{i_{\tau(1)}\cdots i_{\tau(K_2+1)}}_{X_{\tau(1)}\cdots X_{\tau(K_2+1)}} = 0,\, 1 \leq \tau(1) < \cdots < \tau(K_2+1) \leq K_1\right\} \\ &\cdots \\ &\cap \left\{\theta |\, \theta^{i_1 \cdots i_{K_1}}_{X_1 \cdots X_{K_1}} = 0\right\}, \end{aligned}$$

where the condition that $1 \leq i_a \leq m_a$ for $a = 1, 2, \ldots, K_1$ is omitted for simplicity. We define $S^{[K_1,K_2]} = N^{[K_1,K_2+1]}$ and $S^{[K_1,K_1]} = S$, where S is the dually flat space of all multinomial distributions corresponding to the contingency table (Figure 3). Let $\hat{\theta}^{[K_1,K_2]}_{\mathrm{MLE}}$ and $\hat{\theta}^{[K_1,K_2]}_{0}$ denote the MLE of the model $S^{[K_1,K_2]}$ and the model $N^{[K_1,K_2]}$, respectively. Note that $\hat{\theta}^{[K_1,K_2]}_{\mathrm{MLE}} = \hat{\theta}^{[K_1,K_2+1]}_{0}$ because of the definitions of $S^{[K_1,K_2]}$ and $N^{[K_1,K_2]}$. We define $M^{[K_1,K_2]}$ by the m-flat subspace of $S^{[K_1,K_2]}$ that intersects orthogonally with $N^{[K_1,K_2]}$ at $\hat{\theta}^{[K_1,K_2]}_{0} \in N^{[K_1,K_2]}$. It is known that $M^{[K_1,K_2]}$ includes both $\hat{\theta}^{[K_1,K_2]}_{\mathrm{MLE}}$ and $\hat{\theta}^{[K_1,K_2]}_{0}$. Points in $M^{[K_1,K_2]}$ can be represented as

$$\left((\hat{\eta}^{[K_1,K_2]}_{0})^{X_{\tau(1)}}_{i_{\tau(1)}};\, \ldots;\, (\hat{\eta}^{[K_1,K_2]}_{0})^{X_{\tau(1)}\cdots X_{\tau(K_2-1)}}_{i_{\tau(1)}\cdots i_{\tau(K_2-1)}};\, \theta^{i_{\tau(1)}\cdots i_{\tau(K_2)}}_{X_{\tau(1)}\cdots X_{\tau(K_2)}}; 0; \ldots; 0\right),$$

where $\hat{\eta}^{[K_1,K_2]}_{0}$ is the m-affine coordinate of $\hat{\theta}^{[K_1,K_2]}_{0}$. The algorithm BRCT$(K_1, K_2)$ works within $M^{[K_1,K_2]}$. Let $\theta^{[K_1,K_2]}$ denote $\left(\theta^{i_{\tau(1)}\cdots i_{\tau(K_2)}}_{X_{\tau(1)}\cdots X_{\tau(K_2)}}\right)$. A sequence of estimates of $\theta^{[K_1,K_2]}$, $\hat{\theta}_{(0)}, \hat{\theta}_{(1)}, \ldots, \hat{\theta}_{(d_{[K_1,K_2]})}$, is generated by BRCT$(K_1, K_2)$. Strictly speaking, parameter estimates $\hat{\theta}_{(k)}$ should be written as $\hat{\theta}^{[K_1,K_2]}_{(k)}$. However, for simplicity, the superscript $[K_1, K_2]$ is omitted. We define the submodels of $S^{[K_1,K_2]}$, named $M^{[K_1,K_2]}(I)$ and $M^{[K_1,K_2]}\big((i_{\tau(1)}, \ldots, i_{\tau(K_2)}), s, I\big)$, as

$$\begin{aligned} &M^{[K_1,K_2]}(I) = \left\{\theta \mid \theta^{i_{\tilde{\tau}(1)}\cdots i_{\tilde{\tau}(K_2)}}_{X_{\tilde{\tau}(1)}\cdots X_{\tilde{\tau}(K_2)}} = 0, (i_{\tilde{\tau}(1)}, \ldots, i_{\tilde{\tau}(K_2)}) \notin I\right\}, \\ &M^{[K_1,K_2]}\left((i_{\tau(1)}, \ldots, i_{\tau(K_2)}), s, I\right) \\ &\qquad = \left\{\theta \mid \theta^{i_{\tau(1)}\cdots i_{\tau(K_2)}}_{X_{\tau(1)}\cdots X_{\tau(K_2)}} = s,\, \theta^{i_{\tilde{\tau}(1)}\cdots i_{\tilde{\tau}(K_2)}}_{X_{\tilde{\tau}(1)}\cdots X_{\tilde{\tau}(K_2)}} = 0, (i_{\tilde{\tau}(1)}, \ldots, i_{\tilde{\tau}(K_2)}) \notin I\right\}, \end{aligned}$$

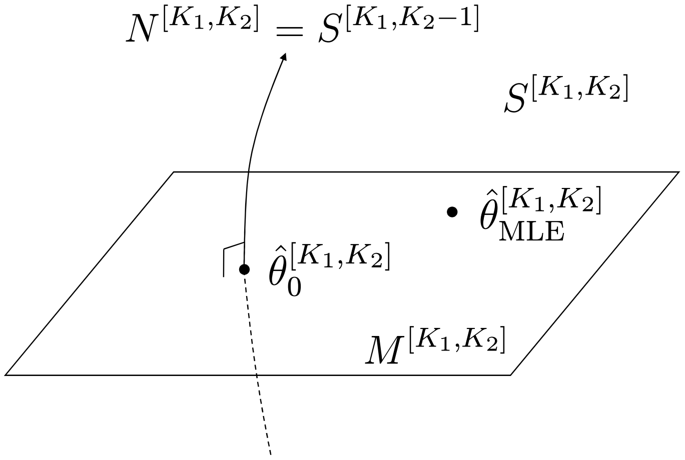
$N^{[K_1,K_2]} = S^{[K_1,K_2-1]}$
$S^{[K_1,K_2]}$
$\hat{\theta}^{[K_1,K_2]}_{\mathrm{MLE}}$
$\hat{\theta}^{[K_1,K_2]}_{0}$
$M^{[K_1,K_2]}$

Figure 3: Submodels for BRCT(K_1, K_2). $N^{[K_1,K_2]} = \Big\{\theta |\, \theta^{i_{\tau(1)}\cdots i_{\tau(K_2)}}_{X_{\tau(1)}\cdots X_{\tau(K_2)}} = 0,\, 1 \le i_{\tau(1)} < \cdots < i_{\tau(K_2)} \le K_1\Big\} \cap \Big\{\theta |\, \theta^{i_{\tau(1)}\cdots i_{\tau(K_2+1)}}_{X_{\tau(1)}\cdots X_{\tau(K_2+1)}} = 0,\, 1 \le i_{\tau(1)} < \cdots < i_{\tau(K_2-1)} \le K_1\Big\} \cap \ldots \cap \Big\{\theta |\, \theta^{i_1\ldots i_{K_1}}_{X_1\ldots X_{K_1}} = 0\Big\}$, $S^{[K_1,K_2]} = N^{[K_1,K_2+1]}$, $\hat{\theta}^{[K_1,K_2]}_{\mathrm{MLE}}$: MLE of the model $S^{[K_1,K_2]}$, $\hat{\theta}^{[K_1,K_2]}_{0}$: MLE of the submodel $N^{[K_1,K_2]}$, $M^{[K_1,K_2]}$: m-flat subspace that intersects orthogonally with $N^{[K_1,K_2]}$ at $\hat{\theta}^{[K_1,K_2]}_{0}$. $M^{[K_1,K_2]}$ is known to include both $\hat{\theta}^{[K_1,K_2]}_{\mathrm{MLE}}$ and $\hat{\theta}^{[K_1,K_2]}_{0}$. BRCT$(K_1, K_2)$ works within $M^{[K_1,K_2]}$, and it estimates $\theta^{i_{\tau(1)}\cdots i_{\tau(K_2)}}_{X_{\tau(1)}\cdots X_{\tau(K_2)}}$.

respectively, for $I \subseteq \Big\{(i_{\tilde{\tau}(1)}, \ldots, i_{\tilde{\tau}(K_2)}) \Big|\, 1 \le \tilde{\tau}(1) < \cdots < \tilde{\tau}(K_2) \le K_1,\, 1 \le i_{\tilde{\tau}(1)} \le m_{\tilde{\tau}(1)}, \ldots, 1 \le i_{\tilde{\tau}(K_2)} \le m_{\tilde{\tau}(K_2)}\Big\}$, $1 \le \tau(1) < \cdots < \tau(K_2) \le K_1$, $(i_{\tau(1)}, \ldots, i_{\tau(K_2)}) \in I$, and $s \in \mathbf{R}$. In the definitions of both $M^{[K_1,K_2]}(I)$ and $M^{[K_1,K_2]}\big((i_{\tau(1)}, \ldots, i_{\tau(K_2)}), s, I\big)$, the condition that $1 \le \tilde{\tau}(1) < \cdots < \tilde{\tau}(K_2) \le K_1$ is omitted. Note again that $M^{[K_1,K_2]}(I)$ and $M^{[K_1,K_2]}\big((i_{\tau(1)}, \ldots, i_{\tau(K_2)}), s, I\big)$ are defined as the submodels of $S^{[K_1,K_2]}$.

The algorithm BRCT(K_1, K_2) is given as follows. Steps 2 to 6 are iterated, generating the sequence $\hat{\theta}_{(0)}, \hat{\theta}_{(1)}, \ldots, \hat{\theta}_{(d_{[K_1,K_2]})}$. Run BRCT$(K_1, K_2 - 1)$ after the iterations unless $K_2 = 2$.

BRCT(K_1, K_2)

input: observation $y_{i_1\ldots i_{K_1}}$ of each cell $(i_1, \ldots, i_{K_1})$

output: parameter estimates $\hat{\theta}_{(0)}, \hat{\theta}_{(1)}, \ldots, \hat{\theta}_{(d_{[K_1,K_2]})}$

1. Let $\hat{\theta}_{(0)} := \hat{\theta}^{[K_1,K_2]}_{\mathrm{MLE}}$, $I = I^{[K_1,K_2]} := \{(i_{\tau(1)}, \ldots, i_{\tau(K_2)}) |\, 1 \le \tau(1) < \cdots < \tau(K_2) \le K_1\}$, and $k := 0$.

2. For all $(i_{\tau(1)}, \ldots, i_{\tau(K_2)}) \in I$, calculate the MLE $\bar{\theta}_{(k)}^{-i_{\tau(1)}\cdots i_{\tau(K_2)}}$ of the model $M^{[K_1,K_2]}\left(I \setminus \{(i_{\tau(1)}, \ldots, i_{\tau(K_2)})\}\right)$.
3. Find $t^* := \min_{(i_{\tau(1)},\ldots,i_{\tau(K_2)})\in I} D\left(\hat{\theta}_{(k)} \mid \bar{\theta}_{(k)}^{-i_{\tau(1)}\cdots i_{\tau(K_2)}}\right)$ and $(i^*_{\tau(1)}, \ldots, i^*_{\tau(K_2)}) := \arg\min_{(i_{\tau(1)},\ldots,i_{\tau(K_2)})\in I} D\left(\hat{\theta}_{(k)} \mid \bar{\theta}_{(k)}^{-i_{\tau(1)}\cdots i_{\tau(K_2)}}\right)$.
4. For $(i_{\tau(1)}, \ldots, i_{\tau(K_2)}) \in I$, calculate $s^*_{i_{\tau(1)}\cdots i_{\tau(K_2)}}$ and $\tilde{\theta}_{(k)}^{-i_{\tau(1)}\cdots i_{\tau(K_2)}} \in l_{(k)}^{-i_{\tau(1)}\cdots i_{\tau(K_2)}}$ satisfying both $D(\hat{\theta}_{(k)} \mid \tilde{\theta}_{(k)}^{-i_{\tau(1)}\cdots i_{\tau(K_2)}}) = t^*$ and $\tilde{\theta}_{(k)}^{-i_{\tau(1)}\cdots i_{\tau(K_2)}} \in M^{[K_1,K_2]}((i_{\tau(1)}, \ldots, i_{\tau(K_2)}), s^*_{i_{\tau(1)}\cdots i_{\tau(K_2)}}, I)$.
5. Let $\hat{\theta}_{(k+1)}^{i_{\tau(1)}\cdots i_{\tau(K_2)}} := s^*_{i_{\tau(1)}\cdots i_{\tau(K_2)}}$ for $(i_{\tau(1)}, \ldots, i_{\tau(K_2)}) \in I$ and $\hat{\theta}_{(k+1)}^{i_{\tau(1)}\cdots i_{\tau(K_2)}} := 0$ for $(i_{\tau(1)}, \ldots, i_{\tau(K_2)}) \notin I$.
6. If $k + 1 < d_{[K_1,K_2]} - 1$, then go to step 2 with $k := k + 1$, $I := I \setminus \{(i^*_{\tau(1)}, \ldots, i^*_{\tau(K_2)})\}$. If $k + 1 = d_{[K_1,K_2]} - 1$, then go to step 7.
7. Let $\hat{\theta}_{(d_K)} := \hat{\theta}_0^{[K_1,K_2]}$. If $K_2 = 2$, stop the algorithm. If $K_2 > 2$, run $\mathbf{BRCT}(K_1, K_2 - 1)$ with $\hat{\theta}_0^{[K_1,K_2]}$ as the start point.

In step 1, $I^{[K_1,K_2]}$ is strictly defined by $I^{[K_1,K_2]} := \{(i_{\tau(1)}, \ldots, i_{\tau(K_2)}) | \, 1 \leq \tau(1) < \cdots < \tau(K_2) \leq K_1,\ 1 \leq i_a \leq m_a\ (1 \leq a \leq K_1)\}$. Note that we have $\hat{\theta}_0^{[K_1,K_2]} = \hat{\theta}_{\mathrm{MLE}}^{[K_1,K_2-1]}$, which means that our algorithm is connected naturally with respect to K_2. In step 5, $\hat{\theta}_{(k+1)}^{i^*_{\tau(1)}\cdots i^*_{\tau(K_2)}} = 0$ for $(i^*_{\tau(1)}, \ldots, i^*_{\tau(K_2)}) \in I$ that was obtained in step 3. This fact indicates that one element of θ becomes 0 in each iteration

When we have an $m_1 \times \cdots \times m_K$-contingency table, we can apply BRCT(K, K) to the contingency table. After BRCT(K, K) is finished, BRCT($K, K-1$) starts next. The algorithms BRCT(K, K), BRCT($K, K-1$), $\ldots$, BRCT(K, 2) then run in turn. We estimate the parameters with more indices earlier, and estimate all parameters to be estimated by our method. Parameters with K indices vanish first and parameters with less indices become zero in turn.

Another choice for the case of K factors is to apply BRCT(K, K') directly to the contingency table, where K' satisfies $K' < K$. This choice is equivalent to the first choice, that is, applying BRCT(K, K) to the contingency table, given that algorithms before BRCT($K, K' + 1$) have already been finished. If we are not interested in the parameters with more indices than K', we can fix such parameters as zero in advance. It is possible for us to start our algorithm at any point unless we want to start it in the middle of the BRCT(K, K'') algorithm for a certain K''.

Table 2: Dataset 1; A Record of Voting in England [13]. The factors X_1 and X_2 have three levels. X_1: factor indicating the party for which a constituent voted in 1966, X_2: factor indicating the party for which a constituent voted in 1970, 175 observations. Level 1: Conservative, Level 2: Labor, Level 0: Liberal. The dimension of θ is 8, and $\theta_{X_1X_2}$ is a 4-dimensional, θ_{X_1} is a 2-dimensional, and θ_{X_2} is a 2-dimensional parameter.

$X_1\backslash X_2$	0	1	2	total
0	13	12	3	28
1	1	68	1	70
2	5	12	60	77
total	19	92	64	175

3 Example

The results of our method are shown for some datasets. We consider only the case of two factors in this section. We used the software R [10] for computing the algorithm. Figures show changes in the values of the parameters in our algorithm. Furthermore, the values of AIC are shown.

3.1 Dataset 1

The first dataset is shown in Table 2. This dataset is a record of voting in England [13]. The factors X_1 and X_2 have three levels. The factor X_1 indicates the party for which a constituent voted in 1966. Level 1 means Conservative, level 2, Labor, and level 0, Liberal. Similarly, the factor X_2 indicates the party for which a constituent voted in 1970. The parameters to be estimated are $\theta^{11}_{X_1X_2}, \theta^{12}_{X_1X_2}, \theta^{21}_{X_1X_2}$, and $\theta^{22}_{X_1X_2}$. Other parameters, $\theta^1_{X_1}, \theta^2_{X_1}, \theta^1_{X_2}$, and $\theta^2_{X_2}$, are calculated from the condition that $\eta^{X_1}_1, \eta^{X_1}_2, \eta^{X_2}_1$, and $\eta^{X_2}_2$ are fixed at the values of the MLE.

The result of the algorithm is shown in Figure 4. The horizontal axis indicates the square root of the divergence from the estimates to the origin. The origin corresponds to the MLE for the independent model. The vertical axis indicates the values of $\theta^{ij}_{X_1X_2}$ $(i, j = 1, 2)$. Line 1 in the figure corresponds to $\theta^{11}_{X_1X_2}$, Line 2, to $\theta^{12}_{X_1X_2}$, Line 3, to $\theta^{21}_{X_1X_2}$, and Line 4, to $\theta^{22}_{X_1X_2}$. The first estimator, the MLE for the full model, is represented on the right-hand side of Figure 4. The algorithm starts from the right-hand side and proceeds to the left-hand side in the figure. Each vertical line corresponds to an estimate. The algorithm ends when the estimator reaches the origin. Five estimates, including both ends of the figure, are obtained. It is clear that one $\theta^{ij}_{X_1X_2}$ becomes zero at each iteration. The vertical line indicates how many elements are not zero. For example, after first iteration,

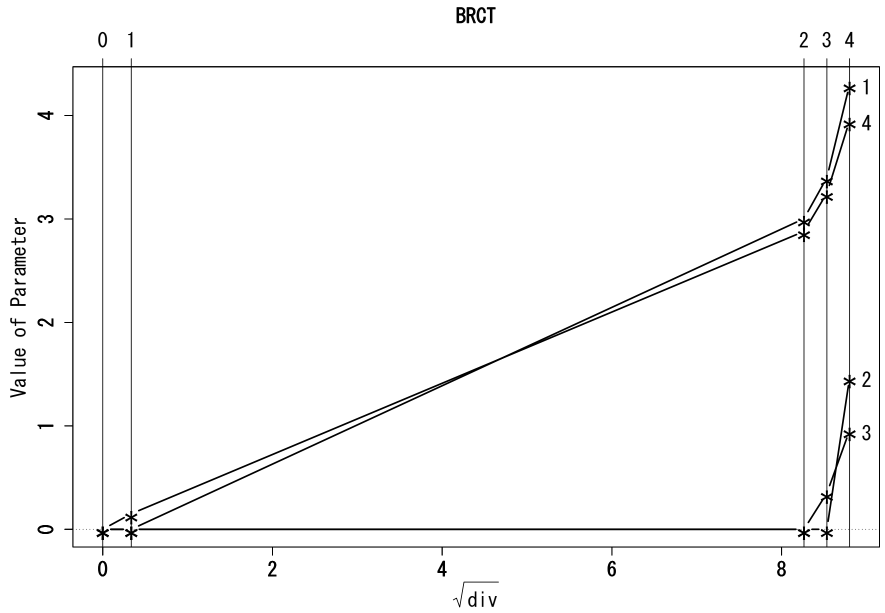

Figure 4: The Result of BRCT(2) for Dataset 1 (Table 2). Four elements of $\theta_{X_1X_2}$ are illustrated. The horizontal axis indicates the square root of the divergence from an estimate to the origin, which corresponds to the Euclidean distance (l_2-norm) in Euclidean space. The vertical axis indicates the values of θ^i $(i = 1, 2, \ldots, 4)$. The first estimate is the rightmost one, which is the MLE of S. The last estimate, fourth estimate, is the leftmost one, which is the MLE of the independent model N. The algorithm proceeds from right to left in this figure. At each iteration, one element of $\theta_{X_1X_2}$ becomes zero. The vertical line indicates how many elements are not zero.

$\theta^{12}_{X_1X_2}$, corresponding to line 2, is zero and others are not zero. According to BRCT(2), $\theta^{ij}_{X_1X_2}$ become zero in the following sequence: $\theta^{12}_{X_1X_2}$, $\theta^{21}_{X_1X_2}$, $\theta^{11}_{X_1X_2}$, and $\theta^{22}_{X_1X_2}$.

3.2 Dataset 2

The second dataset is shown in Table 3. Dataset 2 appeared on [8] as a contingency table for a simulation study. The factor X_1 has four levels, $X_1 = 0, 1, 2, 3$, and the factor X_2 has three levels, $X_2 = 0, 1, 2$. The parameters to be estimated are $\theta^{11}_{X_1X_2}, \theta^{12}_{X_1X_2}, \theta^{21}_{X_1X_2}, \theta^{22}_{X_1X_2}, \theta^{31}_{X_1X_2}$, and $\theta^{32}_{X_1X_2}$. Other parameters, $\theta^1_{X_1}, \theta^2_{X_1}, \theta^3_{X_1}, \theta^1_{X_2}$, and $\theta^2_{X_2}$, are calculated from the fact that $\eta^{X_1}_1, \eta^{X_1}_2, \eta^{X_1}_3, \eta^{X_2}_1$, and $\eta^{X_2}_2$ are fixed at the values of the MLE.

The result of the algorithm is shown in Figure 5. Seven estimates, including both ends of the figure, are obtained. According to BRCT(2), $\theta^{ij}_{X_1X_2}$ become zero in the following sequence: $\theta^{21}_{X_1X_2}$, $\theta^{11}_{X_1X_2}$, $\theta^{32}_{X_1X_2}$, $\theta^{12}_{X_1X_2}$, $\theta^{22}_{X_1X_2}$,

Table 3: Dataset 2; A Contingency Table for a Simulation Study [8]. X_1: factor having four levels, X_2: factor having three levels, 1123 observations. The dimension of θ is 11, and $\theta_{X_1X_2}$ is a 6-dimensional, θ_{X_1} is a 3-dimensional, and θ_{X_2} is a 2-dimensional parameter.

$X_1 \backslash X_2$	0	1	2	total
0	88	91	84	263
1	107	115	92	314
2	96	97	82	275
3	85	100	86	271
total	376	403	344	1123

and $\theta^{31}_{X_1X_2}$.

3.3 AIC-Type Selection

We present the values of AIC for the datasets. The AIC for dataset 1 is shown in Figure 6. The horizontal axis indicates an estimate's number, which means the order of estimates generated by the BRCT algorithm. These numbers also mean how far an estimate is from the origin; a smaller number means that the estimate is farther from the origin. In addition, an estimate's number indicates the number of zeros that the estimate has. Note that an estimate's number differs from the number on the top of Figure 4. The zeroth estimate is the MLE of the full model. The fourth estimate is the MLE of the independence model. The vertical axis indicates the AIC values of the estimates. The minimum AIC is achieved by the zeroth estimate, which is the MLE of the full model.

The AIC for dataset 2 is shown in Figure 7. The horizontal axis indicates an estimate's number, which means the order of estimates generated by the BRCT algorithm. The minimum is achieved by the fifth estimate, which has five zeros.

4 Conclusion

We considered contingency tables and multinomial distributions and introduced natural and expectation parameters for estimation. We proposed BRCT, and provided explanations for three cases: two factors, three factors, and K factors. The BRCT(2) algorithm was proposed for the cases of two factors, and algorithms BRCT(3) and BRCT$(3, 2)$ were proposed for the cases of three factors. In the general case, we proposed the algorithm BRCT(K_1, K_2) for efficiently estimating parameters. These algorithms are based on the information geometry of dually flat spaces. The main idea

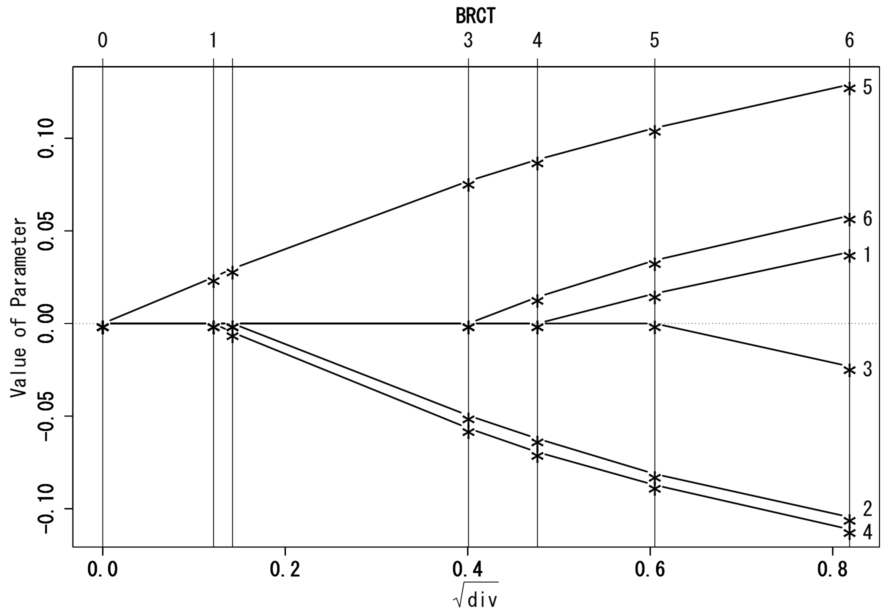

Figure 5: The Result of BRCT(2) for Dataset 2 (Table 3). Six elements of $\theta_{X_1X_2}$ are illustrated. The horizontal axis indicates the square root of the divergence from an estimate to the origin, which corresponds to the Euclidean distance (l_2-norm) in Euclidean space. The vertical axis indicates the values of θ^i $(i = 1, 2, \ldots, 6)$. The first estimate is the rightmost one, which is the MLE of S. The last estimate, sixth estimate, is the leftmost one, which is the MLE of the independent model N. The algorithm proceeds from right to left in this figure. At each iteration, one element of $\theta_{X_1X_2}$ becomes zero. The vertical line indicates how many elements are not zero.

of BRCT(K_1, K_2) came from bisector regression for generalized linear regression [6]. However, we dealt with parameters that are separated into groups depending on type, or the number of indices, while our previous works dealt equivalently with all parameters to be estimated. For contingency tables, we run BRCT(K_1, K_2), BRCT$(K_1, K_2 - 1)$, ..., BRCT$(K_1, 2)$, which means that we estimate parameters with many indices first. Algorithms BRCT(K_1, K_2) connect continuously and we did not need additional effort for combining these algorithms. This fact reflected the nested structure of the models. After these algorithms finish, we obtain the last estimate, the MLE of the independent model. A sequence of parameter estimates is generated, the length of which is the number of parameters to be estimated by our method. The number of our candidates is $d_{[K_1,K_2]} + d_{[K_1,K_2-1]} + \cdots + d_{[K_1,2]}$, which is much smaller than the total number of all possible submodels. We also showed the results of our method for two datasets. AIC was applied for model selection. In the future, we intend to provide an efficient code for

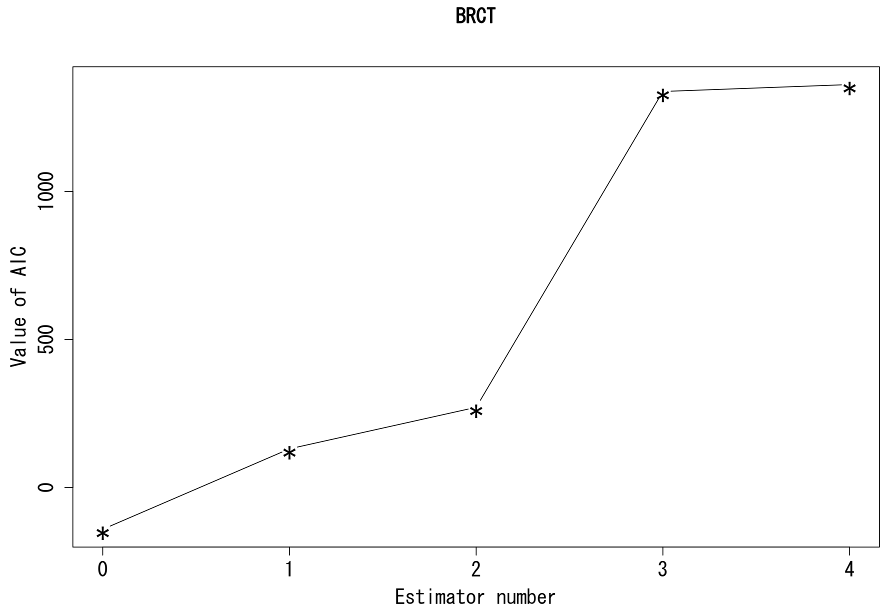

Figure 6: Value of AIC by BRCT for Dataset 1.

contingency tables with three or more factors. A stable and efficient code will help us conduct various simulations. It will also help readers analyze datasets of interest.

References

[1] S. Amari (1985). *Differential-Geometrical Methods in Statistics.* Springer Lecture Notes in Statistics, vol. 28, Springer.

[2] S. Amari and H. Nagaoka (2000). *Methods of Information Geometry.* Translations of Mathematical Monographs, vol. 191, Oxford University Press.

[3] S. S. Chen, D. L. Donoho, and M. A. Saunders (1998). Atomic Decomposition by Basis Pursuit, *SIAM Journal on Scientific Computing*, vol. 20, 33–61.

[4] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani (2004). Least Angle Regression (with discussion), *The Annals of Statistics*, vol. 32, 407–499.

[5] T. Hastie, R. Tibshirani, and J. Friedman (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd Edition. Springer, New York.

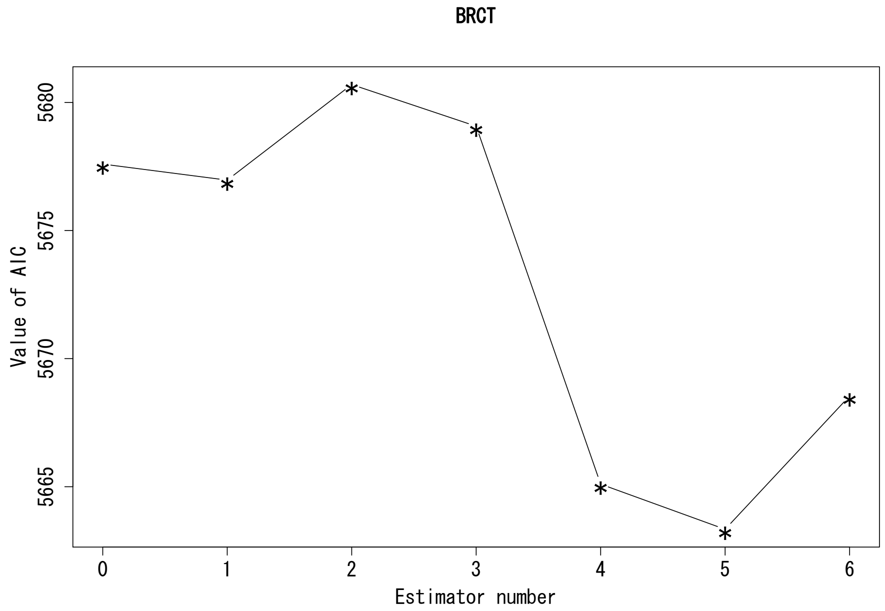

Figure 7: Value of AIC by BRCT for Dataset 2.

[6] Y. Hirose and F. Komaki (2010). An Extension of Least Angle Regression Based on the Information Geometry of Dually Flat Spaces, *Journal of Computational and Graphical Statistics*, vol. 19, 1007–1023.

[7] Y. Hirose and F. Komaki (2011). Edge Selection Based on the Geometry of Dually Flat Spaces for Gaussian Graphical Models, *Mathematical Engineering Technical Report*, METR 2011-35, University of Tokyo.

[8] C. Hirotsu (2007). Row-wise Multiple Comparisons in a Two-way Contingency Table, *Japanese Journal of Applied Statistics*, vol. 36, 1–7.

[9] R. Kass and P. Vos (1997). *Geometrical Foundations of Asymptotic Inference.* John Wiley, New York.

[10] R Development Core Team (2012). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, http://www.R-project.org/.

[11] R. Tibshirani (1996). Regression Shrinkage and Selection via the Lasso, *Journal of the Royal Statistical Society, Series B*, vol. 58, 267–288.

[12] R. Tibshirani (2011). Regression Shrinkage and Selection via the Lasso: A Retrospective, *Journal of the Royal Statistical Society, Series B*, vol. 73, 273–282.

[13] G. J. G. Upton (1977). *The Analysis of Cross-Tabulated Data.* John Wiley & Sons, New York.