MATHEMATICAL ENGINEERING TECHNICAL REPORTS

Perturbation Method for Determining Group of Invariance of Hierarchical Models

Tomonari SEI, Satoshi AOKI and Akimichi TAKEMURA

METR 2008–33 August 2008

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

Perturbation method for determining group of invariance of hierarchical models

Tomonari Sei
Graduate School of Information Science and Technology
University of Tokyo
Satoshi Aoki
Department of Mathematics and Computer Science
Kagoshima University
and
Akimichi Takemura
Graduate School of Information Science and Technology
University of Tokyo

August 2008

Abstract

We propose a perturbation method for determining the (largest) group of invariance of a toric ideal defined in Aoki and Takemura [2008b]. In the perturbation method, we investigate how a generic element in the row space of the configuration defining a toric ideal is mapped by a permutation of the indeterminates. Compared to the proof in Aoki and Takemura [2008b] which was based on stabilizers of a subset of indeterminates, the perturbation method gives a much simpler proof of the group of invariance. In particular, we determine the group of invariance for a general hierarchical model of contingency tables in statistics, under the assumption that the numbers of the levels of the factors are generic. We prove that it is a wreath product indexed by a poset related to the intersection poset of the maximal interaction effects of the model.

Key words and phrases: computational algebraic statistics, group action, stabilizer, sudoku, wreath product.

1 Introduction

Since the introduction of the notion of Markov basis by Diaconis and Sturmfels [1998], toric ideals associated with various statistical models have been intensively investigated by both statisticians and algebraists. In particular, statistical models for contingency tables have been rich sources for new developments (e.g. Aoki and Takemura [2003], Dobra and Sullivant [2004], Ohsugi and Hibi [2008]). The most important statistical model for contingency tables is the hierarchical model (e.g. Lauritzen [1996]), which describes interactions of factors in terms an abstract simplicial complex. The configuration and the toric ideal associated with a hierarchical model is highly symmetric. Therefore it is of considerable interest to determine the (largest) group of invariance of a general hierarchical model. The group of invariance is the set of permutations of the cells of contingency tables (or the indeterminates of a polynomial ring) which leaves the model (or, equivalently, the kernel of the configuration, or the row space of the configuration) invariant. Once the group of invariance is determined, a Markov basis (or equivalently a system of binomial generators of the toric ideal) can be very concisely described (Aoki and Takemura [2008a,b], Hara et al. [2007]) by a list of representative elements from the orbits of the group. Without the consideration of symmetry, Markov bases for statistical problems tend to be very large (e.g. Hemmecke and Malkin [2006]).

Given a particular statistical model it is often easy to guess a candidate group, under which the model is clearly invariant. However as shown in Aoki and Takemura [2008b] it is often difficult to prove that it is the largest group of invariance, i.e., every permutation outside the group does not leave the model invariant. In this paper we propose a perturbation method to determine the group of invariance. In this method, we look at a generic element of the model and check if a permutation maps the element to another element in the model. The candidate group is shown to be the largest group of invariance, if every permutation which maps a sufficiently generic element of the model into the model is necessarily an element of the candidate group. In order to show the effectiveness of this approach, we determine the group of invariance for a general hierarchical model of contingency tables, under the assumption that the numbers of the levels of the factors are generic. We prove that the group of invariance is a wreath product indexed by a poset related to the intersection poset of the maximal interaction effects of the hierarchical model.

In our proof we need to establish some basic facts on hierarchical models, which are not found in the existing statistical literature. These facts are of independent interest and we present them in Section 4.

The organization of the paper is as follows. In Section 2 we give preliminaries and present a perturbation lemma. In Section 3 we state our main theorem, which expresses the group of invariance of a hierarchical model as an intersection of wreath products of symmetric groups. In Section 4 we establish some basic facts on hierarchical models and in Section 5 we give a proof of the main theorem. In Section 6 we rewrite the group of invariance as a wreath product indexed by a poset related to the intersection poset of the maximal interaction effects of the hierarchical model. We conclude the paper by some discussions in Section 7.

2 Preliminaries and a perturbation lemma

In this section we summarize preliminary facts on hierarchical models for contingency tables, define the group of invariance and present a perturbation lemma, which is essential for our proofs. We mainly follow the notation and terminology of Lauritzen [1996].

2.1 Preliminaries on hierarchical models for contingency tables

A hierarchical model for m-factor contingency tables with numbers of levels $I_1, \dots, I_m$ is specified by an abstract simplicial complex. Let Δ be an abstract simplicial complex ([Kozlov, 2008, Section 2.1]) of subsets of a finite set $\{1, \dots, m\} = [m]$ of "factors". We denote the set of maximal simplices of Δ by $\operatorname{red}\Delta = \{D_1, \dots, D_K\}$. Maximal simplices are called maximal interaction effects of the model. For each factor $j \in [m]$, the set of "levels"of j is denoted by $\mathcal{I}_j = \{1, \dots, I_j\} = [I_j]$, where $I_j \geq 2$. The direct product of the set of levels $\mathcal{I} = \mathcal{I}_1 \times \cdots \times \mathcal{I}_m$ is the set of "cells" and its element $\boldsymbol{i} = (i_1, \dots, i_m)$ is a cell. A contingency table $x = (x(\boldsymbol{i}))_{\boldsymbol{i}\in\mathcal{I}}$ is a vector of nonnegative integers indexed by the cells. $x(\boldsymbol{i})$ is the frequency of the cell $\boldsymbol{i}$. In this paper, the symbol $A \subset B$ means that A is a subset of B. If A is a proper subset of B, then we write $A \subsetneq B$.

For a subset $D \subset [m]$ of factors, let $\mathcal{I}_D = \prod_{j\in D} \mathcal{I}_j$. A subvector of indices $\boldsymbol{i}_D = (i_j)_{j\in D} \in \mathcal{I}_D$ is called "a marginal cell". When a particular cell $\boldsymbol{i} = (i_1, \dots, i_m)$ is given, $\boldsymbol{i}_D$ is regarded as a subvector of $\boldsymbol{i}$, i.e., the projection of $\boldsymbol{i}$ onto the coordinates in D. For a contingency table x, its D-marginal table $x_D^+ = (x^+(\boldsymbol{i}_D))_{\boldsymbol{i}_D\in\mathcal{I}_D}$ is defined by

$$x^+(\boldsymbol{i}_D) = \sum_{\boldsymbol{j}\in\mathcal{I},\ \boldsymbol{j}_D=\boldsymbol{i}_D} x(\boldsymbol{j}).$$

Similar notation is used even when $x(\boldsymbol{i})$ is not necessarily a nonnegative integer.

Fix $\mathcal{I}$ and a hierarchical model Δ with $\operatorname{red}\Delta = \{D_1, \dots, D_K\}$. Write $\nu = \sum_{k=1}^K |\mathcal{I}_{D_k}|$ and $p = |\mathcal{I}|$. For each $\boldsymbol{i} = (i_1, \dots, i_m) \in \mathcal{I}$ consider the following vector (cf. Ohsugi and Hibi [2008])

$$\mathbf{e}^{(1)}(\boldsymbol{i}_{D_1}) \oplus \mathbf{e}^{(2)}(\boldsymbol{i}_{D_2}) \oplus \cdots \oplus \mathbf{e}^{(K)}(\boldsymbol{i}_{D_K}) \ \in\ \mathbb{Z}^\nu$$

where $\mathbf{e}^{(k)}(\boldsymbol{i}_{D_k})$ is a unit coordinate vector of dimension $|\mathcal{I}_{D_k}|$ with 1 at the position $\boldsymbol{i}_{D_k}$ and 0 everywhere else. The configuration A_Δ for Δ is the set of p vectors

$$A_\Delta = \left\{\mathbf{e}^{(1)}(\boldsymbol{i}_{D_1}) \oplus \cdots \oplus \mathbf{e}^{(K)}(\boldsymbol{i}_{D_K})\right\}_{\boldsymbol{i}\in\mathcal{I}}.$$

In this paper we regard A_Δ as a $\nu \times p$ integral matrix representing a linear map from $\mathbb{Q}^p$ to $\mathbb{Q}^\nu$. The matrix A_Δ can also be expressed by Kronecker products of identity matrices and vectors consisting of 1's ([Takemura and Aoki, 2004, Section 2.1]). We also assume that the domain $\mathbb{Q}^p$ is equipped with the standard inner product and we identify $\mathbb{Q}^p$ with its dual space by the standard inner product.

Let $\{u_{\boldsymbol{i}}\}_{\boldsymbol{i}\in\mathcal{I}}$ be the set of indeterminates indexed by the cells and let $\{t^{(1)}_{\boldsymbol{i}_{D_1}}\}_{\boldsymbol{i}_{D_1}\in\mathcal{I}_{D_1}} \cup \cdots \cup \{t^{(K)}_{\boldsymbol{i}_{D_K}}\}_{\boldsymbol{i}_{D_K}\in\mathcal{I}_{D_K}}$ denote the set of indeterminates indexed by the rows of A_Δ. The toric ideal

I_{A_Δ} is the kernel of the polynomial homomorphism π_Δ defined by $\pi_\Delta(u_{\boldsymbol{i}}) = t^{(1)}_{\boldsymbol{i}_{D_1}} \times \cdots \times t^{(K)}_{\boldsymbol{i}_{D_K}}$. The structure of the toric ideal is much more difficult than the kernel of matrix A_Δ. However we will define the invariance property of I_{A_Δ} in terms of the invariance property of the kernel of A_Δ.

As we discuss in Section 2.2 we are interested in the kernel of A_Δ and the linear space spanned by the rows of A_Δ. In the following we denote the kernel of A_Δ and the the linear space spanned by the rows of A_Δ by $\ker A_\Delta$ and $r(A_\Delta)$, respectively. Note that $\ker A_\Delta$ and $r(A_\Delta)$ are orthogonal complements to each other: $r(A_\Delta) = (\ker A_\Delta)^\perp$.

In statistical theory, $r(A_\Delta)$ corresponds to a log-linear model of cell probabilities, where the canonical parameter vector of the exponential family is specified to lie in the linear space $r(A_\Delta)$. We use the single term "model" for Δ, $r(A_\Delta)$ and $\ker A_\Delta$ because they correspond to each other.

The explicit form of $\ker A_\Delta$ and $r(A_\Delta)$ are well known in the literature on contingency tables (e.g. Lauritzen [1996]). $\ker A_\Delta$ is written as

$$\ker A_\Delta = \{y \mid y^+(\boldsymbol{i}_D) = 0, \forall \boldsymbol{i}_D \in \mathcal{I}_D, \forall D \in \operatorname{red}\Delta\}. \tag{1}$$

For $D \subset [m]$, let $\theta_D : \mathcal{I}_D \to \mathbb{Q}$ denote a function defined on the set of marginal cells $\mathcal{I}_D$. Then extend the domain of θ_D to $\mathcal{I}$ by $\theta_D(\boldsymbol{i}) = \theta_D(\boldsymbol{i}_D)$. We call θ_D a function (or a table) depending only on the marginal cell $\boldsymbol{i}_D$. Let $L_D = \{\theta_D\} \subset \mathbb{Q}^p$ denote the linear space of these tables. Then

$$r(A_\Delta) = \sum_{D \in \operatorname{red}\Delta} L_D, \tag{2}$$

where the summation on the right-hand side denotes the subspace spanned by $\{L_D\}_{D\in\operatorname{red}\Delta}$. Note that if $E \in \Delta$, then $L_E \subset L_D$ for some $D \in \operatorname{red}\Delta$. Therefore the right-hand is spanned by L_E, $E \in \Delta$.

2.2 The group of invariance of a toric ideal

Now we give a definition of the group of invariance of a toric ideal.

Let $S_{\mathcal{I}}$ denote the symmetric group on $\mathcal{I}$, i.e. an element $g \in S_{\mathcal{I}}$ is a permutation of the cells of $\mathcal{I}$. Then $g \in S_{\mathcal{I}}$ acts (from the left) on the $|\mathcal{I}|$-dimensional rational vector space $\mathbb{Q}^{|\mathcal{I}|} = \{(y(\boldsymbol{i}))_{\boldsymbol{i}\in\mathcal{I}}\}$ by the permutation of components: $(gy)(\boldsymbol{i}) = y(g^{-1}(\boldsymbol{i}))$. Similarly g acts on the set of indeterminates $\{u_{\boldsymbol{i}}\}_{\boldsymbol{i}\in\mathcal{I}}$. If we regard g as a linear map from $\mathbb{Q}^{|\mathcal{I}|}$ to itself, then it is represented by a permutation matrix. We denote the permutation matrix also by g. Note that g is orthogonal. For a given subspace $L \subset \mathbb{Q}^{|\mathcal{I}|}$, let $G_L = \{g \in S_{|\mathcal{I}|} \mid gL = L\}$ denote the set-wise stabilizer of L.

Let A be a $\nu \times p$ rational matrix as in the previous subsection. The symmetric group S_p acts on the set of columns of A and on $\mathbb{Q}^p$. In Aoki and Takemura [2008b] we defined the *group of invariance* for A as the set-wise stabilizer $G_{\ker A} \subset S_p$ of $\ker A$. From the viewpoint of toric ideal, the group of invariance is the set of permutations of the indeterminates, which leaves the toric ideal invariant. Let $r(A) \subset \mathbb{Q}^p$ denote the linear space spanned by the rows of A. By Proposition 1 of Aoki and Takemura [2008b], we have $G_{\ker A} = G_{r(A)}$.

Our objective is to understand $G_{\ker A_\Delta} = G_{r(A_\Delta)}$ of a hierarchical model Δ.

2.3 A perturbation lemma

Here we present the following lemma.

Lemma 1. *(Perturbation lemma) Let n, b be positive integers. There exist n positive integers $(Y_l)_{l=1}^n$, such that*

$$\{-b, -b+1, \ldots, b-1, b\}^n \ni (c_l)_{l=1}^n \mapsto \sum_{l=1}^n c_l Y_l \tag{3}$$

is injective. Furthermore we can choose n vectors $Y^{(j)} = (Y_l^{(j)})_{l=1}^n$, $j = 1, \ldots, n$, such that (3) is injective for each j and they constitute a basis of the vector space $\mathbb{Q}^n$.

Proof. Let $Y_l^{(j)} = (2b+j)^{l-1}$, $(l, j \in [n])$. By the uniqueness of the base $2b+j$ expression of positive integers, the map $(c_l)_{l=1}^n \mapsto \sum_{l=1}^n c_l Y_l^{(j)}$ is injective. Furthermore $Y_l^{(j)}$, $j = 1, \ldots, n$, are linearly independent in view of the van der Monde determinant. □

In view of the above lemma, we define a generic contingency table belonging to $r(A_\Delta)$ for a given set of cells $\mathcal{I}$ and a hierarchical model Δ with $\operatorname{red}\Delta = \{D_1, \ldots, D_K\}$.

Definition 1. *For $n = \nu = \sum_{k=1}^K |\mathcal{I}_{D_k}|$ and $b = |\mathcal{I}|$ choose $(Y_l)_{l=1}^n$ such that (3) is injective. Decompose $(Y_l)_{l=1}^n$ into subvectors of sizes $|\mathcal{I}_{D_k}|$, $k = 1, \ldots, K$, as $(Y_l)_{l=1}^n = \big((\theta_{D_1}(\boldsymbol{i}_{D_1}))_{\boldsymbol{i}_{D_1} \in \mathcal{I}_{D_1}}, \ldots, (\theta_{D_K}(\boldsymbol{i}_{D_K}))_{\boldsymbol{i}_{D_K} \in \mathcal{I}_{D_K}}\big)$ and define*

$$x(\boldsymbol{i}) = \theta_{D_1}(\boldsymbol{i}_{D_1}) + \cdots + \theta_{D_K}(\boldsymbol{i}_{D_K}), \quad \boldsymbol{i} \in \mathcal{I}.$$

We call this x a generic element of $r(A_\Delta)$.

Note that an element g of the group of invariance $G_{r(A_\Delta)}$ has to map a generic element x of $r(A_\Delta)$ into $r(A_\Delta)$. This fact helps us to determine $G_{r(A_\Delta)}$.

3 Group of invariance of hierarchical models

In this section we first consider a candidate group for the group of invariance $G_{\ker A_\Delta}$ and then present our main theorem, which states that the candidate group is indeed the group of invariance, provided that the number of levels I_j, $j \in [m]$, are generic.

For $D \subset [m]$ consider a simplicial complex Δ^D, which consists of all subsets of D, and let $\ker A_{\Delta^D} = L_D^\perp = \{y \mid y^+(\boldsymbol{i}_D) = 0, \forall \boldsymbol{i}_D \in \mathcal{I}_D\}$. Then $\ker A_\Delta = \bigcap_{D \in \operatorname{red}\Delta} \ker A_{\Delta^D}$ by (1). Let $G_D = G_{\ker A_{\Delta^D}}$ denote the group of invariance for Δ^D. Then it is easily seen that

$$\bigcap_{D \in \operatorname{red}\Delta} G_D \subset G_{\ker A_\Delta}. \tag{4}$$

Therefore we can take $\bigcap_{D\in\mathrm{red}\,\Delta} G_D$ as a candidate group for the group of invariance $G_{\ker A_\Delta}$. As we will present an example of sudoku in Section 6, in general the inclusion in (4) is strict. However if the number of levels I_j, $j \in [m]$, are generic, then the inclusion in (4) is in fact an equality.

Before stating our main theorem, we prove that $G_D = G_{\ker A_{\Delta^D}}$ is a wreath product of symmetric groups. Let $S_{\mathcal{I}_D}$ denote the symmetric group acting on the set of D-marginal cells and $S_{\mathcal{I}_{D^C}}$ denote the symmetric group acting on the set of D^C-marginal cells, where D^C is the complement of D. Let $(S_{\mathcal{I}_{D^C}})^{\mathcal{I}_D}$ denote the set of all functions from $\mathcal{I}_D$ to $S_{\mathcal{I}_{D^C}}$. Then the wreath product $S_{\mathcal{I}_{D^C}} \,\mathrm{wr}\, S_{\mathcal{I}_D}$ is a set $W = S_{\mathcal{I}_D} \times (S_{\mathcal{I}_{D^C}})^{\mathcal{I}_D}$. The operation of W as a subgroup of $S_{\mathcal{I}}$ is defined by its action to $\mathcal{I}$, where $g = (h, \tilde{h}) \in W$ acts on $\boldsymbol{i} \in \mathcal{I}$ by $(g\boldsymbol{i})_D = h\boldsymbol{i}_D$ and $(g\boldsymbol{i})_{D^C} = \tilde{h}(\boldsymbol{i}_D)\boldsymbol{i}_{D^C}$. Then we have the following proposition.

Proposition 1. *The group of invariance G_D for the hierarchical model Δ^D is given by the wreath product $S_{\mathcal{I}_{D^C}} \,\mathrm{wr}\, S_{\mathcal{I}_D}$.*

Proof. For notational simplicity, we prove the proposition for the case of $m = 2$ and $D = \{1\}$ and write $\boldsymbol{i}$ as (i, j). We denote $S_{\mathcal{I}_D}$ and $S_{\mathcal{I}_{D^C}}$ by S_{I_1} and S_{I_2}, respectively. The proof for a general case is totally the same by the consideration of a "pseudofactor" (see Section 4 for details on pseudofactors).

First we show that $S_{I_2} \,\mathrm{wr}\, S_{I_1} \subset G_D$. Let $x \in r(A_{\Delta^D}) = L_D$. Then $x(i, j) = \theta(i)$. Let $g \in S_{I_2} \,\mathrm{wr}\, S_{I_1}$. Then $g(i, j) = (h(i), \tilde{h}_i(j))$, where $h \in S_{I_1}$ and $\tilde{h}_i \in S_{I_2}$ for each $i \in [I_1]$. Then

$$(gx)(i, j) = x(g^{-1}(i, j)) = \theta(g^{-1}(i, j)_1) = \theta(h^{-1}(i)),$$

where the subscript "1" in $g^{-1}(i, j)_1$ denotes the first component. Therefore $gx \in L_D$.

We now show the converse $G_D \subset S_{I_2} \,\mathrm{wr}\, S_{I_1}$. In order to show this we assume that $x \in L_D$ is generic, i.e. $\theta(i)$'s are distinct. Suppose that $(gx)(i, j) = \theta(g^{-1}(i, j)_1) \in L_D$. Then $g^{-1}(i, j)_1$ does not depend on j. Therefore we can write $g^{-1}(i, j) = (h(i), \bar{h}(i, j))$. Since g is a bijection, h is a bijection and $j \mapsto \bar{h}(i, j)$ is a bijection for each i. Therefore $g^{-1} \in S_{I_2} \,\mathrm{wr}\, S_{I_1}$. □

Now we state the main theorem of this paper.

Theorem 1. *Consider a hierarchical model Δ. Assume that $|\mathcal{I}_D|$, $D \in \mathrm{red}\,\Delta$, are distinct and $I_j > 2$ except for at most one $j \in [m]$. Then the group of invariance $G_{\ker A_\Delta}$ is given by*

$$G_{\ker A_\Delta} = \bigcap_{D\in\mathrm{red}\,\Delta} \left(S_{\mathcal{I}_{D^C}} \,\mathrm{wr}\, S_{\mathcal{I}_D}\right). \tag{5}$$

A proof of this theorem is given in Section 5 after we establish several important facts on hierarchical models in Section 4. As seen from the statement of Theorem 1, it seems that the case of two-level factors $I_j = 2$ needs a special consideration, although the requirements on the levels in Theorem 1 may be too restrictive. We discuss these points again in Section 7. We will give some examples of Theorem 1 in Section 6 after rewriting the right-hand side of (5).

4 Some basic facts on hierarchical models

In this section we establish basic facts on hierarchical models. In particular we are interested in the behavior of a hierarchical model, when a maximal simplex is deleted from $\operatorname{red}\Delta$. This is because for our proof of Theorem 1 we employ the induction on the number $K = |\operatorname{red}\Delta|$ of maximal interaction effects in Δ.

Let $E \subset [m]$. We first define "incremental subspaces" of L_E by

$$N_E = L_E \cap \left(\sum_{j \in E} L_{E \setminus \{j\}} \right)^{\perp} \tag{6}$$

if $E \neq \emptyset$, and $N_\emptyset = L_\emptyset$. Recall that L_E is the linear space of tables depending only on the marginal cell $\boldsymbol{i}_D$ and that $\sum_{j \in E} L_{E \setminus \{j\}}$ is the subspace spanned by $\{L_{E \setminus \{j\}}\}_{j \in E}$. The following lemma is easily proved and well known in statistical analysis of variance (ANOVA).

Lemma 2. *Let E and F be subsets of $[m]$.*

(1) $N_E = L_E \cap (\sum_{j \in E} L_{[m] \setminus \{j\}})^{\perp}$.

(2) If $E \neq F$, then $N_E \perp N_F$.

(3) $L_E = \sum_{F \subset E} N_F$.

(4) For any simplicial complex Δ, $r(A_\Delta) = \sum_{F \in \Delta} N_F$ and $\ker A_\Delta = \sum_{F \notin \Delta} N_F$.

(5) The orthogonal projection π_{N_E} onto N_E is given by

$$(\pi_{N_E} x)(\boldsymbol{i}) = (\pi_{N_E} x)(\boldsymbol{i}_E) = \sum_{F \subset E} \frac{(-1)^{|E \setminus F|}}{|\mathcal{I}_{F^C}|} x^+(\boldsymbol{i}_F)$$

for all $x \in \mathbb{Q}^{\mathcal{I}}$. Recall that $|E \setminus F|$ is the cardinality of $E \setminus F$.

Let $D \in \operatorname{red}\Delta$ be a maximal simplex. As in the beginning of Section 3 let Δ^D denote the simplicial complex consisting of all subsets of D. Note that $r(A_{\Delta^D}) = L_D$. Now we define $\Delta_{\setminus D}$ by "deleting the maximal interaction effects D from $\operatorname{red}\Delta$", i.e. by

$$\operatorname{red}\Delta_{\setminus D} = (\operatorname{red}\Delta) \setminus D.$$

We have the following proposition.

Proposition 2. *Let $D \in \operatorname{red}\Delta$. Then*

$$r(A_\Delta) \cap \ker A_{\Delta_{\setminus D}} = r(A_{\Delta^D}) \cap \ker A_{\Delta_{\setminus D}} = \sum_{E \in \Delta^D \setminus \Delta_{\setminus D}} N_E. \tag{7}$$

Proof. By Lemma 2, we have $r(A_\Delta) = \sum_{E\in\Delta} N_E$ and $\ker A_{\Delta_{\setminus D}} = \sum_{E\notin\Delta_{\setminus D}} N_E$. Therefore the equalities follow from the relation $\Delta \setminus \Delta_{\setminus D} = \Delta^D \setminus \Delta_{\setminus D}$. □

We next define a partial difference operator. For $j \in [m]$ and $x = (x(\boldsymbol{i}))_{\boldsymbol{i}\in\mathcal{I}}$ define

$$(\partial_j x)(\boldsymbol{i}) = x(\boldsymbol{i}) - x(i_1, \ldots, i_{j-1}, 1, i_{j+1}, \ldots, i_m), \quad \boldsymbol{i} = (i_1, \ldots, i_m).$$

For $E \subset [m]$ define $\partial_E = \prod_{j\in E} \partial_j$. Note that for two subsets $D, E \subset [m]$, $E \not\subset D$, we have

$$\partial_E \theta_D = 0, \qquad \forall \theta_D \in L_D. \tag{8}$$

It is obvious that for any $D, E \subset [m]$ and $\theta_D \in L_D$, we have $\partial_E \theta_D \in L_D$.

Concerning the partial difference operator ∂_E we have the following proposition.

Proposition 3. *For all $E \subset [m]$, $\ker(\partial_E) = \sum_{F\not\supset E} N_F$.*

Proof. We first show that the subspace $\ker \partial_j$ is equal to $L_{[m]\setminus\{j\}}$. Let $x \in \ker \partial_j$. Then $x(\boldsymbol{i}) = x(i_1, \ldots, i_{j-1}, 1, i_{j+1}, \ldots, i_m)$ and therefore $x \in L_{[m]\setminus\{j\}}$. Conversely, if $x \in L_{[m]\setminus\{j\}}$, then $\partial_j x = 0$. Therefore we see that $\ker \partial_j = L_{[m]\setminus\{j\}}$. Since the operators $\{\partial_j\}_{j\in[m]}$ are mutually commutable projectors (and therefore simultaneously diagonalizable), we have $\ker \partial_E = \sum_{j\in E} \ker \partial_j$. Therefore, by using Lemma 2,

$$\ker \partial_E = \sum_{j\in E} \ker \partial_j = \sum_{j\in E} L_{[m]\setminus\{j\}} = \sum_{j\in E} \sum_{F\subset[m]\setminus\{j\}} N_F = \sum_{F\not\supset E} N_F.$$

The last equality comes from the fact that $F \not\supset E$ if and only if $F \subset [m] \setminus \{j\}$ for some $j \in E$. □

Combining Lemma 2 and Proposition 3, we have the following proposition. We will use the proposition with $\Delta' = \Delta_{\setminus D}$ in the proof of the main theorem.

Proposition 4. *Let Δ and Δ' be two simplicial complexes such that $\Delta \supset \Delta'$. Then $x \in r(A_{\Delta'})$ if and only if $x \in r(A_\Delta)$ and $\partial_E x = 0$ for all $E \in \Delta \setminus \Delta'$.*

Proof. The statement is equivalent to $r(A_{\Delta'}) = r(A_\Delta) \cap (\cap_{E\in\Delta\setminus\Delta'} \ker \partial_E)$. The left-hand side is $\sum_{F\in\Delta'} N_F$. The right-hand side is

$$\left(\sum_{F\in\Delta} N_F\right) \cap \left(\bigcap_{E\in\Delta\setminus\Delta'} \sum_{F\not\supset E} N_F\right) = \sum_{F\in\Delta''} N_F,$$

where $\Delta'' = \{F \in \Delta \mid F \not\supset E, \ \forall E \in \Delta \setminus \Delta'\}$. It is sufficient to prove that $\Delta' = \Delta''$. Let $F \in \Delta'$. Clearly $F \in \Delta$. Now assume that there exists some $E \in \Delta \setminus \Delta'$ such that $F \supset E$. Then, since $F \in \Delta'$ and $F \supset E$, we have $E \in \Delta'$. This contradicts to $E \in \Delta \setminus \Delta'$. Therefore $F \not\supset E$ for any $E \in \Delta \setminus \Delta'$. Conversely, suppose that $F \in \Delta$ and $F \not\supset E$ for any $E \in \Delta \setminus \Delta'$. Then, since $F \in \Delta$ and $F \notin \Delta \setminus \Delta'$, we have $F \in \Delta \setminus (\Delta \setminus \Delta') = \Delta'$. □

Note that in the proof of Proposition 1, we treated the combination of factors in D as a single factor and the combination of factors in D^C as another single factor. This identification is well known in design of experiments as a "pseudofactor" (e.g. Monod and Bailey [1992]). As the last topic of this section we fully discuss the notion of a pseudofactor and a natural partial order induced on the set of pseudofactors from the hierarchical model. The resulting poset plays an essential role in the next section.

For each $i \in [m]$ let

$$(\mathrm{red}\,\Delta)(i) = \{D \in \mathrm{red}\,\Delta \mid i \in D\} \tag{9}$$

denote the set of maximal interaction effects containing i. If $(\mathrm{red}\,\Delta)(i) = (\mathrm{red}\,\Delta)(j)$ we say that i, j belong to the same pseudofactor and denote this as $i \overset{\Delta}{\sim} j$. The relation $\overset{\Delta}{\sim}$ is an equivalence relation and $[m]$ is partitioned into disjoint equivalence classes. We call each equivalence class a pseudofactor. In the framework of this paper, we can replace a pseudofactor by a single factor, although we do not do this in this paper. Let $\mathcal{P}$ denote the set of pseudofactors, i.e. $\mathcal{P} = [m]/\overset{\Delta}{\sim}$. For $\rho \in \mathcal{P}$ let

$$(\mathrm{red}\,\Delta)(\rho) = (\mathrm{red}\,\Delta)(i), \qquad i \in \rho.$$

Now we introduce a partial order onto $\mathcal{P}$ by

$$\rho \geq \rho' \;\Leftrightarrow\; (\mathrm{red}\,\Delta)(\rho) \supset (\mathrm{red}\,\Delta)(\rho').$$

With this partial order $\mathcal{P}$ becomes a partially ordered set (poset). We call this poset the "sieve poset", induced by the simplicial complex Δ.

The sieve poset induced by Δ is related to the intersection poset. The intersection poset $\mathcal{Q}$ of $\mathrm{red}\,\Delta$ is the set of intersections of $\mathrm{red}\,\Delta$, that is, $\mathcal{Q} = \{\cap_{D\in S} D \mid S \subset \mathrm{red}\,\Delta\}$. The order of $\mathcal{Q}$ is the reverse inclusion order: $\cap_{D\in S} D \leq \cap_{D\in S'} D$ if $S \subset S'$. We assume $[m] \in \mathcal{Q}$ just for convenience. We show that there is an injective homomorphism from $\mathcal{P}$ into $\mathcal{Q}$. In fact, the following lemma holds.

Lemma 3. *Let $V(\rho) = \cup_{\rho' \geq \rho} \rho'$. Then $V(\rho) = \cap_{D\in(\mathrm{red}\,\Delta)(\rho)} D$. Furthermore V is an injective homomorphism from $\mathcal{P}$ into $\mathcal{Q}$.*

Proof. Let $i \in V(\rho)$. Then there exists some $\rho' \geq \rho$ such that $i \in \rho'$. This means $(\mathrm{red}\,\Delta)(i) = (\mathrm{red}\,\Delta)(\rho') \supset (\mathrm{red}\,\Delta)(\rho)$. Therefore $i \in \cap_{D\in(\mathrm{red}\,\Delta)(\rho)} D$. The converse is similarly proved. Next we prove that V is homomorphic and injective. If $\rho' \geq \rho$ then $(\mathrm{red}\,\Delta)(\rho') \supset (\mathrm{red}\,\Delta)(\rho)$ and therefore $V(\rho') = \cap_{D\in(\mathrm{red}\,\Delta)(\rho')} D \geq \cap_{D\in(\mathrm{red}\,\Delta)(\rho)} D = V(\rho)$ from the definition of the order of $\mathcal{Q}$. If $\rho' \neq \rho$, then $V(\rho') = \cup_{\rho'' \geq \rho'} \rho'' \neq \cup_{\rho'' \geq \rho} \rho'' = V(\rho)$. □

We remark that V is not surjective in general. For example, let $m = 3$ and $\mathrm{red}\,\Delta = \{\{1,2\},\{2,3\},\{1,3\}\}$. Then $\mathcal{P} = \{\{1\},\{2\},\{3\}\}$ with a trivial order (i.e. no two distinct elements are comparable) and $\mathcal{Q} = \{\emptyset,\{1\},\{2\},\{3\},\{1,2\},\{2,3\},\{1,3\},\{1,2,3\}\}$. The homomorphism is $V(\{i\}) = \{i\}$ for $i \in \{1,2,3\}$. Thus V is not surjective. In other words, the poset $\mathcal{Q}$ has the same amount of information as $\mathrm{red}\,\Delta$ because $\mathrm{red}\,\Delta = \mathrm{red}(\mathcal{Q}\setminus\{[m]\})$, but the poset $\mathcal{P}$ loses the information as the example shows. For description of the group of invariance, we only need the sieve poset rather than the intersection poset.

5 A proof of the main theorem

Now we employ induction on $K = |\operatorname{red}\Delta|$. The theorem is true for $K = 1$ by Proposition 1. Therefore assume that the theorem holds for $K-1$. Throughout the proof we choose $D \in \operatorname{red}\Delta$ such that $|\mathcal{I}_D| = \min_{F\in\operatorname{red}\Delta} |\mathcal{I}_F|$. We consider deleting D from $\operatorname{red}\Delta$.

Let $x = \sum_{F\in\operatorname{red}\Delta} \theta_F$ be a generic element of $r(A_\Delta)$ (Definition 1). List the values of θ_F as $\alpha_{\boldsymbol{i}_F} = \theta_F(\boldsymbol{i}_F) = \theta_F(\boldsymbol{i})$, $\boldsymbol{i}_F \in \mathcal{I}_F$. Then $x(\boldsymbol{i})$ can be written as

$$x(\boldsymbol{i}) = \sum_{F\in\operatorname{red}\Delta} \sum_{\boldsymbol{j}_F\in\mathcal{I}_F} \chi_{\boldsymbol{j}_F}(\boldsymbol{i})\alpha_{\boldsymbol{j}_F}, \qquad \chi_{\boldsymbol{j}_F}(\boldsymbol{i}) = \begin{cases} 1 & \text{if } \boldsymbol{i}_F = \boldsymbol{j}_F, \\ 0 & \text{otherwise.} \end{cases} \tag{10}$$

In view of (4) it suffices to show that any $g \in G_{r(A_\Delta)}$ belongs to the right-hand side of (5). Fix an arbitrary $g \in G_{\ker A_\Delta} = G_{r(A_\Delta)}$ and let $y = gx$. Then $y \in r(A_\Delta)$ and y can be written as $y = \sum_{F\in\operatorname{red}\Delta} \eta_F$. Note that at this point we do not have any relation between θ_F's and η_F's. Fix an arbitrary $E \in \Delta \setminus \Delta_{\setminus D}$ and take the partial difference with respect to E.

$$\partial_E y = \partial_E \eta_D \tag{11}$$

by (8). The right-hand side $\partial_E\eta_D$ depends only on $\boldsymbol{i}_D$. Now consider the left-hand side $\partial_E y$. $(\partial_E y)(\boldsymbol{i})$ is a linear combination of $2^{|E|}$ $y(\boldsymbol{j})$'s with the coefficient 1 for $2^{|E|-1}$ terms and -1 for other $2^{|E|-1}$ terms. Now $y(\boldsymbol{j}) = x(g^{-1}(\boldsymbol{j})) = (x \circ g^{-1})(\boldsymbol{j})$. We substitute $x(g^{-1}(\boldsymbol{j}))$ by the right-hand side of (10) and take the linear combination. Then $(\partial_E y)(\boldsymbol{i})$ is written as

$$(\partial_E y)(\boldsymbol{i}) = \sum_{F\in\operatorname{red}\Delta} \sum_{\boldsymbol{j}_F\in\mathcal{I}_F} Q_{\boldsymbol{j}_F}(\boldsymbol{i})\alpha_{\boldsymbol{j}_F}, \tag{12}$$

where

$$Q_{\boldsymbol{j}_F}(\boldsymbol{i}) = (\partial_E(\chi_{\boldsymbol{j}_F} \circ g^{-1}))(\boldsymbol{i}) \in \{-2^{|E|-1}, \ldots, 2^{|E|-1}\}.$$

Since we have taken generic $\alpha_{\boldsymbol{j}_F}$'s, by the perturbation lemma, $Q_{\boldsymbol{j}_F}(\boldsymbol{i})$ is uniquely determined by $(\partial_E y)(\boldsymbol{i})$ for each $\boldsymbol{i}$ and for each $F \in \operatorname{red}\Delta$ and $\boldsymbol{j}_F$. However recall by (11) that $(\partial_E y)(\boldsymbol{i})$ only depends on $\boldsymbol{i}_D$. This implies that $Q_{\boldsymbol{j}_F}(\boldsymbol{i})$ also depends only on $\boldsymbol{i}_D$ for each $\boldsymbol{j}_F$. More precisely, if we take the $\boldsymbol{i}_D$ marginal of (12), then we have

$$(\partial_E y)^+(\boldsymbol{i}_D) = |\mathcal{I}_{D^C}|(\partial_E y)(\boldsymbol{i}) = \sum_{F\in\operatorname{red}\Delta} \sum_{\boldsymbol{j}_F\in\mathcal{I}_F} Q^+_{\boldsymbol{j}_F}(\boldsymbol{i}_D)\alpha_{\boldsymbol{j}_F}.$$

Therefore by uniqueness we see that $Q_{\boldsymbol{j}_F}(\boldsymbol{i}) = Q^+_{\boldsymbol{j}_F}(\boldsymbol{i}_D)/|\mathcal{I}_{D^C}|$ depends only on $\boldsymbol{i}_D$.

Now we claim that $Q_{\boldsymbol{j}_F}(\boldsymbol{i}) = 0$ for all $\boldsymbol{j}_F$, $F \neq D$, and for all $\boldsymbol{i} \in \mathcal{I}$. For readability, we state this as a lemma and give a proof. Recall that $E \in \Delta \setminus \Delta_{\setminus D}$ is arbitrarily fixed and the following lemma holds for any such E.

Lemma 4. $Q_{\boldsymbol{j}_F}(\boldsymbol{i}) = 0$ *for all* $\boldsymbol{j}_F \in \mathcal{I}_F$, $F \in (\operatorname{red}\Delta) \setminus \{D\}$, *and for all* $\boldsymbol{i} \in \mathcal{I}$.

Proof. Suppose that there exists some $\boldsymbol{i}^0$ and some $\boldsymbol{j}_F$, such that $Q_{\boldsymbol{j}_F}(\boldsymbol{i}^0) \neq 0$. Then, because $Q_{\boldsymbol{j}_F}(\boldsymbol{i}^0)$ only depends on $\boldsymbol{i}^0_D$, for this $\boldsymbol{j}_F$ we have

$$|\{\boldsymbol{i} \mid Q_{\boldsymbol{j}_F}(\boldsymbol{i}) \neq 0\}| \ \geq\ |\{\boldsymbol{i} \mid \boldsymbol{i}_D = \boldsymbol{i}^0_D\}| = |\mathcal{I}_{D^C}| = \frac{|\mathcal{I}|}{|\mathcal{I}_D|}.$$

Write

$$\mathcal{I}_{|\boldsymbol{i}^0_D} = \{\boldsymbol{i} \mid \boldsymbol{i}_D = \boldsymbol{i}^0_D\} = \{(\boldsymbol{i}_{D^C}, \boldsymbol{i}^0_D)\}_{\boldsymbol{i}_{D^C} \in \mathcal{I}_{D^C}}.$$

The D^C-component $\boldsymbol{i}_{D^C}$ of the elements of $\mathcal{I}_{|\boldsymbol{i}^0_D}$ are all distinct.

For $\boldsymbol{i} \in \mathcal{I}_{|\boldsymbol{i}^0_D}$, consider $Q_{\boldsymbol{j}_F}(\boldsymbol{i}_{D^C}, \boldsymbol{i}^0_D) = (\partial_E(\chi_{\boldsymbol{j}_F} \circ g^{-1}))(\boldsymbol{i}_{D^C}, \boldsymbol{i}^0_D)$, which is a sum of $2^{|E|}$ terms of the form $\pm\chi_{\boldsymbol{j}_F}(g^{-1}(\boldsymbol{i}'))$. Since the operator ∂_E only touches indices $i_j, j \in E \subset D$, we note that these terms $\chi_{\boldsymbol{j}_F}(g^{-1}(\boldsymbol{i}'))$ have the common index $\boldsymbol{i}_{D^C}$, i.e., $\boldsymbol{i}'_{D^C} = \boldsymbol{i}_{D^C}$. Therefore we can write

$$Q_{\boldsymbol{j}_F}(\boldsymbol{i}_{D^C}, \boldsymbol{i}^0_D) = \sum_{\boldsymbol{j}_D \in \mathcal{I}_D} \beta_{\boldsymbol{j}_D} \chi_{\boldsymbol{j}_F}(g^{-1}(\boldsymbol{i}_{D^C}, \boldsymbol{j}_D)), \tag{13}$$

where $\beta_{\boldsymbol{j}_D} \in \{-1, 0, 1\}$. It is important to note that the sets of cells $\{g^{-1}(\boldsymbol{i}_{D^C}, \boldsymbol{j}_D)\}_{\boldsymbol{j}_D \in \mathcal{I}_D}$ are mutually disjoint for different values of $\boldsymbol{i}_{D^C}$, because g^{-1} is a bijection on $\mathcal{I}$.

Now if $Q_{\boldsymbol{j}_F}(\boldsymbol{i}_{D^C}, \boldsymbol{i}^0_D) \neq 0$, there exists at least one non-zero term on the right-hand side of (13). Therefore for each $\boldsymbol{i}_{D^C}$ there exists $\boldsymbol{j}_D$ such that $\chi_{\boldsymbol{j}_F}(g^{-1}(\boldsymbol{i}_{D^C}, \boldsymbol{j}_D)) = 1$. By the disjointness noted above, it follows that

$$|\{\boldsymbol{i}' \mid \chi_{\boldsymbol{j}_F}(g^{-1}(\boldsymbol{i}')) = 1\}| \ \geq\ |\mathcal{I}_{D^C}|.$$

On the other hand, by the definition of $\chi_{\boldsymbol{j}_F}$, we have

$$|\{\boldsymbol{i}' \mid \chi_{\boldsymbol{j}_F}(g^{-1}(\boldsymbol{i}')) = 1\}| = |\{\boldsymbol{i}' \mid \chi_{\boldsymbol{j}_F}(\boldsymbol{i}') = 1\}| = |\mathcal{I}_{F^C}|.$$

Combining the above results we have

$$\frac{|\mathcal{I}|}{|\mathcal{I}_F|} = |\mathcal{I}_{F^C}| \geq |\mathcal{I}_{D^C}| = \frac{|\mathcal{I}|}{|\mathcal{I}_D|} \qquad \text{or} \qquad |\mathcal{I}_F| \leq |\mathcal{I}_D|$$

However we have assumed that $|\mathcal{I}_D|$ is the (unique) minimum among $|\mathcal{I}_F|$, $F \in \operatorname{red}\Delta$. Therefore $F = D$. □

From the above lemma, we have

$$(\partial_E y)(\boldsymbol{i}) = \sum_{\boldsymbol{j}_D \in \mathcal{I}_D} Q_{\boldsymbol{j}_D}(\boldsymbol{i}) \alpha_{\boldsymbol{j}_D}. \tag{14}$$

We have shown (14) for generic x. However, since (14) is an algebraic relation and all generic tables span $r(A_\Delta)$ by the perturbation lemma, (14) holds for all $x \in r(A_\Delta)$. Now in (10) set $\alpha_{\boldsymbol{j}_D} = 0$, $\forall \boldsymbol{j}_D \in \mathcal{I}_D$. Namely let $x = \sum_{F \in \operatorname{red}\Delta, F \neq D} \theta_F$ be any element of $r(A_{\Delta_{\setminus D}})$. Then $\partial_E y = \partial_E(x \circ g^{-1}) = 0$ for all $E \in \Delta \setminus \Delta_{\setminus D}$. Therefore $y \in r(A_{\Delta_{\setminus D}})$ by

Proposition 4. This means that $g \in G_{r(A_\Delta)}$ has to map every $x \in r(A_{\Delta_{\setminus D}})$ into $r(A_{\Delta_{\setminus D}})$. In other words, $g \in G_{r(A_{\Delta_{\setminus D}})}$. By induction assumption we have shown

$$g \in \bigcap_{F \in \mathrm{red}\,\Delta, F \neq D} \big(S_{\mathcal{I}_{F^C}} \,\mathrm{wr}\, S_{\mathcal{I}_F}\big).$$

Now it remains to show that $g \in S_{\mathcal{I}_{D^C}} \,\mathrm{wr}\, S_{\mathcal{I}_D}$. By assumption g maps $r(A_\Delta)$ into itself. We have shown that g maps $r(A_{\Delta_{\setminus D}})$ into itself. Since g is orthogonal as a linear map, it follows that g maps the subspace $M = r(A_\Delta) \cap r(A_{\Delta_{\setminus D}})^\perp$ into itself. By Proposition 2, we obtain

$$M = r(A_\Delta) \cap r(A_{\Delta_{\setminus D}})^\perp = r(A_{\Delta^D}) \cap \ker A_{\Delta_{\setminus D}} = \sum_{E \in \Delta^D \setminus \Delta_{\setminus D}} N_E.$$

Recall that N_E is the incremental subspace defined by (6). Note that $N_D \subset M \subset r(A_{\Delta^D})$. We claim that there exists a table ϕ_D in M such that $\phi_D(\boldsymbol{i}_D)$, $\boldsymbol{i}_D \in \mathcal{I}_D$, are all distinct. We state this as a lemma and give a proof.

Lemma 5. *There exists a table ϕ_D in M such that $\phi_D(\boldsymbol{i}_D)$, $\boldsymbol{i}_D \in \mathcal{I}_D$, are all distinct.*

Proof. Consider a generic element θ_D of L_D. Let π_{N_D} denote the orthogonal projection to N_D and put $\phi_D = \pi_{N_D}\theta_D$. By Lemma 2, the following expression for $\phi_D(\boldsymbol{i}_D)$ holds.

$$\phi_D(\boldsymbol{i}_D) = \sum_{E \subset D} (-1)^{|D \setminus E|} \frac{1}{|\mathcal{I}_{E^C}|} \theta_D^+(\boldsymbol{i}_E).$$

Recall that $\theta_D^+(\boldsymbol{i}_E) = \sum_{\boldsymbol{j} \in \mathcal{I}, \boldsymbol{j}_E = \boldsymbol{i}_E} \theta_D(\boldsymbol{j}_D)$. Multiplying each side by $|\mathcal{I}|$, we have

$$\begin{aligned} |\mathcal{I}|\phi_D(\boldsymbol{i}_D) &= \sum_{E \subset D} (-1)^{|D \setminus E|} |\mathcal{I}_E| \theta_D^+(\boldsymbol{i}_E) \\ &= \sum_{\boldsymbol{j}_D \in \mathcal{I}_D} C(\boldsymbol{i}_D, \boldsymbol{j}_D) \theta_D(\boldsymbol{j}_D), \end{aligned}$$

where

$$\begin{aligned} C(\boldsymbol{i}_D, \boldsymbol{j}_D) &= \sum_{E \subset \mathrm{eq}(\boldsymbol{i}_D, \boldsymbol{j}_D)} (-1)^{|D \setminus E|} |\mathcal{I}_E||\mathcal{I}_{D^C}| \\ &= |\mathcal{I}_{D^C}|(-1)^{|D \setminus \mathrm{eq}(\boldsymbol{i}_D, \boldsymbol{j}_D)|} \prod_{j \in \mathrm{eq}(\boldsymbol{i}_D, \boldsymbol{j}_D)} (I_j - 1) \end{aligned}$$

and $\mathrm{eq}(\boldsymbol{i}_D, \boldsymbol{j}_D) = \{j \in D \mid i_j = j_j\}$. Note that $C(\boldsymbol{i}_D, \boldsymbol{j}_D) \in \{-|\mathcal{I}|, \ldots, |\mathcal{I}|\}$. For given $\boldsymbol{i}_D$ and $\boldsymbol{i}'_D$, if $C(\boldsymbol{i}_D, \boldsymbol{j}_D) \neq C(\boldsymbol{i}'_D, \boldsymbol{j}_D)$ for some $\boldsymbol{j}_D$, then $\phi_D(\boldsymbol{i}_D) \neq \phi_D(\boldsymbol{i}'_D)$ because θ_D is generic. Therefore it is sufficient to prove that if $\boldsymbol{i}_D \neq \boldsymbol{i}'_D$, then there exists some $\boldsymbol{j}_D \in \mathcal{I}_D$ such that $C(\boldsymbol{i}_D, \boldsymbol{j}_D) \neq C(\boldsymbol{i}'_D, \boldsymbol{j}_D)$. Since I_j is greater than 2 except for at most one $j \in [m]$, we can show that $C(\boldsymbol{i}_D, \boldsymbol{j}_D) = |\mathcal{I}_{D^C}| \prod_{j \in D}(I_j - 1)$ if and only if $\boldsymbol{i}_D = \boldsymbol{j}_D$. Thus $C(\boldsymbol{i}_D, \boldsymbol{i}_D) \neq C(\boldsymbol{i}'_D, \boldsymbol{i}_D)$ whenever $\boldsymbol{i}_D \neq \boldsymbol{i}'_D$. This proves the lemma. □

We have proved that there exists $\phi_D \in N_D \subset M$ such that $\phi_D(\boldsymbol{i}_D)$, $\boldsymbol{i}_D \in \mathcal{I}_D$, are all distinct. Since $g\phi_D \in M \subset r(A_{\Delta^D})$, the same proof as in Proposition 1 shows that $g \in S_{\mathcal{I}_{D^C}} \operatorname{wr} S_{\mathcal{I}_D}$.

This completes the proof of Theorem 1.

6 The wreath product indexed by the sieve poset

Although (5) gives a form of the group of invariance, it is not yet sufficiently explicit to write down the group of invariance for a given hierarchical model. We can employ the notion of a wreath product of a partially ordered set of actions to describe the group of invariance more explicitly. The notion of a wreath product of an partially ordered set of actions has been defined by many authors (Holland [1969], Wells [1976], Silcock [1977], Bailey et al. [1983]). We follow a succinct definition in Section 7 of Wells [1976].

The poset we use is the sieve poset $(\mathcal{P}, \leq)$ defined in Section 4. Recall that $\mathcal{P}$ is a partition of $[m]$ and each class $\rho \in \mathcal{P}$ has $(\operatorname{red}\Delta)(\rho) = (\operatorname{red}\Delta)(i) = \{D \in \operatorname{red}\Delta \mid i \in D\}$ for $i \in \rho$. The order relation $\rho \leq \rho'$ on $\mathcal{P}$ is defined by $(\operatorname{red}\Delta)(\rho) \subset (\operatorname{red}\Delta)(\rho')$. Recall that $V(\rho) = \cup_{\rho' \geq \rho}\rho'$. We also define the ancestor set of ρ by

$$A(\rho) = \cup_{\rho' > \rho}\rho' = V(\rho) \setminus \rho.$$

If $A(\rho) = \emptyset$, then we let $\mathcal{I}_{A(\rho)}$ be a 1-element set, say $\{1\}$.

Definition 2 (Wells [1976])**.** *The wreath product of the symmetric groups $(S_{\mathcal{I}_\rho})_{\rho\in\mathcal{P}}$ indexed by the poset $\mathcal{P}$ is defined by $W = \prod_{\rho\in\mathcal{P}}(S_{\mathcal{I}_\rho})^{\mathcal{I}_{A(\rho)}}$, where $(S_{\mathcal{I}_\rho})^{\mathcal{I}_{A(\rho)}}$ is the set of all functions from $\mathcal{I}_{A(\rho)}$ to $S_{\mathcal{I}_\rho}$. The action of $w = (w_\rho)_{\rho\in\mathcal{P}} \in W$ on $\mathcal{I}$ is defined by*

$$(w\boldsymbol{i})_\rho = w_\rho(\boldsymbol{i}_{A(\rho)})\boldsymbol{i}_\rho.$$

In the above definition, we use the parentheses for evaluating functions (such as $w_\rho(\boldsymbol{i}_{A(\rho)})$) and do not use them for action (such as $w\boldsymbol{i}$).

For example, if $\operatorname{red}\Delta = \{D\}$ and $\emptyset \subsetneq D \subsetneq [m]$, then $\mathcal{P} = \{D, D^C\}$ with the order relation $D > D^C$. In this case, the wreath product of $(S_{\mathcal{I}_\rho})_{\rho\in\mathcal{P}}$ is the usual wreath product $S_{\mathcal{I}_{D^C}} \operatorname{wr} S_{\mathcal{I}_D}$ because $\mathcal{I}_{A(D)} = \{1\}$ and $\mathcal{I}_{A(D^C)} = \mathcal{I}_D$.

The following lemma by Bailey et al. [1983] is useful.

Lemma 6 (Theorem B of Bailey et al. [1983])**.** *The wreath product is characterized as follows.*

$$\prod_{\rho\in\mathcal{P}}(S_{\mathcal{I}_\rho})^{\mathcal{I}_{A(\rho)}} = \left\{g \in S_{\mathcal{I}} \mid (g\boldsymbol{i})_{V(\rho)} \text{ depends only on } \boldsymbol{i}_{V(\rho)} \text{ for any } \rho \in \mathcal{P}\right\}.$$

The proof of the following lemma is easy and omitted.

Lemma 7. *Let A and B be two subsets of $[m]$. Let $g \in S_{\mathcal{I}}$. Assume that $(g\boldsymbol{i})_A$ depends only on $\boldsymbol{i}_A$ and that $(g\boldsymbol{i})_B$ depends only on $\boldsymbol{i}_B$. Then $(g\boldsymbol{i})_{A\cap B}$ depends only on $\boldsymbol{i}_{A\cap B}$, and $(g\boldsymbol{i})_{A\cup B}$ depends only on $\boldsymbol{i}_{A\cup B}$.*

Now we establish the following theorem.

Theorem 2. *The group of invariance coincides with the wreath product of* $(S_{\mathcal{I}_\rho})_{\rho\in\mathcal{P}}$*, that is,*

$$\bigcap_{D\in\operatorname{red}\Delta}(S_{\mathcal{I}_{D^C}}\operatorname{wr}S_{\mathcal{I}_D}) \;=\; \prod_{\rho\in\mathcal{P}}(S_{\mathcal{I}_\rho})^{\mathcal{I}_{A(\rho)}}. \tag{15}$$

Proof. By Lemma 6, the left-hand side in (15) is equal to

$$\{g\in S_{\mathcal{I}} \mid (g\boldsymbol{i})_D \text{ depends only on } \boldsymbol{i}_D \text{ for any } D\in\operatorname{red}\Delta\}.$$

On the other hand, also by Lemma 6, the right-hand side in (15) is equal to

$$\{g\in S_{\mathcal{I}} \mid (g\boldsymbol{i})_{V(\rho)} \text{ depends only on } \boldsymbol{i}_{V(\rho)} \text{ for any } \rho\in\mathcal{P}\}.$$

Now the equality (15) is clear if one uses Lemma 7 with two relations

$$D \;=\; \bigcup_{\rho\in\mathcal{P},\rho\subset D} V(\rho) \quad\text{and}\quad V(\rho) \;=\; \bigcap_{D\in(\operatorname{red}\Delta)(\rho)} D.$$

The former one is from the construction of $\mathcal{P}$. The latter one is Lemma 3. □

Corollary 1. *The group of invariance is equal to the direct product of the symmetric groups* $(S_{\mathcal{I}_\rho})_{\rho\in\mathcal{P}}$ *if and only if the poset* $\mathcal{P}$ *has the trivial order, i.e. no two distinct elements of* $\mathcal{P}$ *are comparable.*

Let us present some examples. Here we abbreviate $(S_{\mathcal{I}_\rho})^{\mathcal{I}_{A(\rho)}}$ to $S^*_{\rho|A(\rho)}$, and $S_{\mathcal{I}_\rho}$ to S^*_ρ, respectively.

Example 1. *Let* $\operatorname{red}\Delta = \{\{1\},\dots,\{m\}\}$*. Then* $\mathcal{P} = \{\{1\},\dots,\{m\}\}$ *with the trivial order. The wreath product is the direct product* $W = \prod_{j=1}^m S^*_{\{j\}}$.

Example 2. *Let* $m = 3$ *and* $\operatorname{red}\Delta = \{\{1\},\{2,3\}\}$*. In this case,* $\{2,3\}$ *is a pseudofactor but not a single factor. Then the sieve poset is* $\{\{1\},\{2,3\}\}$ *with the trivial order. The wreath product is* $W = S^*_{\{1\}}\times S^*_{\{2,3\}}$.

Example 3. *Let* $m = 3$ *and* $\operatorname{red}\Delta = \{\{1,2\},\{2,3\}\}$*. Then the sieve poset is* $\{\{1\},\{2\},\{3\}\}$ *with the order relations* $\{1\}<\{2\}$ *and* $\{3\}<\{2\}$ *(no other relations). The wreath product is* $W = S^*_{\{1\}|\{2\}}\times S^*_{\{2\}}\times S^*_{\{3\}|\{2\}}$.

Example 4. *Let* $m = 3$ *and* $\operatorname{red}\Delta = \{\{1\},\{2\}\}$*. Note that the factor* $\{3\}$ *does not appear explicitly. Then the sieve poset is* $\{\{1\},\{2\},\{3\}\}$ *with the order relations* $\{3\}<\{1\}$ *and* $\{3\}<\{2\}$*. The wreath product is* $W = S^*_{\{1\}}\times S^*_{\{2\}}\times S^*_{\{3\}|\{1,2\}}$.

Example 5. *Let* $m\geq 3$ *and* $\operatorname{red}\Delta = \{\{1,2\},\{2,3\},\dots,\{m-1,m\},\{m,1\}\}$*. Then* $\mathcal{P} = \{\{1\},\dots,\{m\}\}$ *with the trivial order. The wreath product is* $W = \prod_{j=1}^m S^*_{\{j\}}$.

Example 6. *Let* $m = 6$ *and* $\operatorname{red}\Delta = \{\{1,4,5\},\{2,5,6\},\{3,4,6\}\}$. *Then the sieve poset is* $\mathcal{P} = \{\{1\},\{2\},\{3\},\{4\},\{5\},\{6\}\}$ *with the order relations* $\{1\} < \{5\}$, $\{1\} < \{4\}$, $\{2\} < \{5\}$, $\{2\} < \{6\}$, $\{3\} < \{4\}$ *and* $\{3\} < \{6\}$ *(no other relations). The wreath product is* $W = S^*_{\{1\}|\{4,5\}} \times S^*_{\{2\}|\{5,6\}} \times S^*_{\{3\}|\{4,6\}} \times S^*_{\{4\}} \times S^*_{\{5\}} \times S^*_{\{6\}}$.

The last example is a counter-example to the conjecture in the discussion of [Aoki and Takemura, 2008b, Section 5]. In our terminology, the conjecture is stated as "If all pseudofactors are single, i.e. $\mathcal{P} = \{\{1\},\dots,\{m\}\}$, and the intersection of $\operatorname{red}\Delta$ is empty, then the group of invariance is the direct product of the symmetric groups on each factor". The conjecture is justified if we impose an additional condition that $\mathcal{P}$ has the trivial order (see Corollary 1).

We show an example in that the inclusion (4) is strict.

Example 7 (Sudoku)**.** *The solution of sudoku is a* 9×9 *table whose each row, each column and each* 3×3 *block contains the digits from 1 to 9 exactly once. Following the terminology of Russel and Jarvis [2006], we call a "row" of* 3 *blocks a band and a "column" of* 3 *blocks a stack. The solution is considered as a* $3 \times 3 \times 3 \times 3 \times 9$ *contingency table* $x(i,j,k,l,c)$ *where we define* $x(i,j,k,l,c) = 1$ *if the number* $c \in [9]$ *is put on the* j*-th row of the* i*-th band and the* l*-th column of the* k*-th stack and* $x(i,j,k,l,c) = 0$ *otherwise. Then the restriction is given by four equations*

$$x(i,j,+,+,c) = 1, \quad x(+,+,k,l,c) = 1, \quad x(i,+,k,+,c) = 1, \quad x(i,j,k,l,+) = 1,$$

where "+" denotes taking marginal (sum) over the index. The maximal simplices of this model is given by

$$\operatorname{red}\Delta = \{\{1,2,5\},\{3,4,5\},\{1,3,5\},\{1,2,3,4\}\}.$$

The sieve poset is $\mathcal{P} = \{\{1\},\{2\},\{3\},\{4\},\{5\}\}$ *with the order* $\{1\} > \{2\}$ *and* $\{3\} > \{4\}$ *(no other relations). The wreath product is given by*

$$W = S^*_{\{1\}} \times S^*_{\{2\}|\{1\}} \times S^*_{\{3\}} \times S^*_{\{4\}|\{3\}} \times S^*_{\{5\}},$$

which consists of permutation of bands, permutation of rows in each band, permutation of stacks, permutation of columns in each stack and permutation of numbers. However, the group of invariance $G_{\ker A_\Delta}$ *has an additional permutation* f *defined by* $f(i,j,k,l,c) = (k,l,i,j,c)$. *The permutation* f *does not belong to the wreath product* W. *Note that the model does not satisfy the assumption of Theorem 1 because* $|\mathcal{I}_{\{1,2,5\}}| = |\mathcal{I}_{\{3,4,5\}}| = |\mathcal{I}_{\{1,3,5\}}| = |\mathcal{I}_{\{1,2,3,4\}}| = 81$. *The group generated by* W *and* f *is used to count the number of essentially different solutions of sudoku in Russel and Jarvis [2006].*

7 Discussions

We derived an explicit formula of the group of invariance provided that the number of levels I_j, $j \in [m]$, are generic. In our future work we intend to generalize this result by weakening the restriction on the number of levels. We conjecture that under mild regularity conditions the group of invariance is generated by the wreath product of this paper and the permutation of factors with a common number of levels. However, it seems to be difficult to solve this problem. For example, as described in Example 3 of Aoki and Takemura [2008b], the group of invariance for the $2 \times 2 \times 2$ contingency tables with fixed two-dimensional marginals is different from the new conjectured candidate group. In the example, as was pointed out by a referee to Aoki and Takemura [2008b], the group of invariance is not faithful. Here an action G to L is called faithful if the kernel $\{g \in G \mid gx = x, \ \forall x \in L\}$ of the action consists only of the unit element. On the other hand, we can prove that the group of invariance is faithful under the assumption of Theorem 1. Indeed, in a similar way to the proof of Lemma 5, we can show that there exists a table $\phi \in N_{[m]} \subset \ker A_\Delta$ such that $\{\phi(\boldsymbol{i})\}_{\boldsymbol{i} \in \mathcal{I}}$ are all distinct. Therefore if $g\phi = \phi$, then g has to be the identity map.

Random sampling from the group of invariance is important for performing the Markov Chain Monte Carlo (MCMC) method on contingency tables. See Aoki and Takemura [2008a] for details. In Theorem 2 we rewrote the group of invariance from an intersection form to a wreath-product form. The wreath product is useful for random sampling. Let us briefly describe it. The wreath product is given by $W = \prod_{\rho \in \mathcal{P}} (S_{\mathcal{I}_\rho})^{\mathcal{I}_{A(\rho)}}$. We show an algorithm to obtain a uniformly random sample $w = (w_\rho)_{\rho \in \mathcal{P}}$ from W. Let us number $\mathcal{P}$ as $\mathcal{P} = \{\rho_1, \ldots, \rho_l\}$ such that $i < j$ whenever $\rho_i < \rho_j$. Then, from $i = l$ down to 1, we independently generate $w_{\rho_i}(\boldsymbol{i}_{A(\rho_i)})$ from $S_{\mathcal{I}_{\rho_i}}$ for each $\boldsymbol{i}_{A(\rho_i)} \in \mathcal{I}_{A(\rho_i)}$. The resulting element $w = (w_\rho)_{\rho \in \mathcal{P}}$ is a uniformly random sample from W. Remark that the intersection form in Theorem 1 does not give such a procedure.

References

Satoshi Aoki and Akimichi Takemura. Minimal basis for a connected Markov chain over $3 \times 3 \times K$ contingency tables with fixed two-dimensional marginals. *Aust. N. Z. J. Stat.*, 45(2):229–249, 2003. ISSN 1369-1473.

Satoshi Aoki and Akimichi Takemura. Minimal invariant Markov basis for sampling contingency tables with fixed marinals. *Ann. Inst. Statist. Math.*, 60(2):229–256, 2008a.

Satoshi Aoki and Akimichi Takemura. The largest group of invariance for Markov bases and toric ideals. *J. Symbolic Comput.*, 43(5):342–358, 2008b. ISSN 0747-7171.

R.A. Bailey, C.E. Praeger, C.A. Rowley, and T.P. Speed. Generalized wreath products of permutation groups. *Proc. London Math. Soc.*, 47(3):69–82, 1983.

Persi Diaconis and Bernd Sturmfels. Algebraic algorithms for sampling from conditional distributions. *Ann. Statist.*, 26(1):363–397, 1998. ISSN 0090-5364.

Adrian Dobra and Seth Sullivant. A divide-and-conquer algorithm for generating Markov bases of multi-way tables. *Comput. Statist.*, 19(3):347–366, 2004. ISSN 0943-4062.

Hisayuki Hara, Satoshi Aoki, and Akimichi Takemura. Fibers of sample size two of hierarchical models and Markov bases of decomposable models for contingency tables, 2007. Preprint. arXiv:math/0701429v1.

Raymond Hemmecke and Peter Malkin. Computing generating sets of lattice ideals, 2006. Preprint. arXiv:math/0508359v2.

W.C. Holland. The characterization of generalized wreath products. *J. Algebra*, 13: 152–172, 1969.

Dmitry Kozlov. *Combinatorial Algebraic Topology*, volume 21 of *Algorithms and Computation in Mathematics*. Springer, Berlin, 2008. ISBN 978-3-540-71961-8.

Steffen L. Lauritzen. *Graphical Models*. Oxford University Press, Oxford, 1996.

H. Monod and R.A. Bailey. Pseudofactors: normal use to improve design and facilitate analysis. *Appl. Statist.*, 41(2):317–336, 1992.

Hidefumi Ohsugi and Takayuki Hibi. Toric ideals arising from contingency tables. In *Commutative Algebra and Combinatorics*, Ramanujan Society Lecture Note Series, No.4, 2008. In press.

E. Russel and F. Jarvis. Mathematics of sudoku II, 2006. Preprint.

H.L. Silcock. Generalized wreath products and the lattice of normal subgroups of a group. *Algebra Universalis*, 7:361–372, 1977.

Akimichi Takemura and Satoshi Aoki. Some characterizations of minimal Markov basis for sampling from discrete conditional distributions. *Ann. Inst. Statist. Math.*, 56(1): 1–17, 2004. ISSN 0020-3157.

Charles Wells. Some applications of the wreath product construction. *Amer. Math. Monthly*, 83(5):317–338, 1976. ISSN 0002-9890.