Collapsibility of Marginal Models for Categorical Data

S. GHOSH AND P. VELLAISAMY

Abstract. We consider marginal log-linear models for parameterizing distributions on multidimensional contingency tables. These models generalize ordinary log-linear and multivariate logistic models, besides several others. First, we obtain some characteristic properties of marginal log-linear parameters. Then we define collapsibility and strict collapsibility of these parameters in a general sense. Several necessary and sufficient conditions for collapsibility and strict collapsibility are derived using the technique of Möbius inversion. These include results for an arbitrary set of marginal log-linear parameters having some common effects. The connections of collapsibility and strict collapsibility to various forms of independence of the variables are discussed. Finally, we establish a result on the relationship between parameters with the same effect but different margins, and use it to demonstrate smoothness of marginal log-linear models under collapsibility conditions thereby obtaining a curved exponential family.

1. Introduction

Various models for multidimensional contingency tables have been proposed by imposing restrictions on marginal or conditional distributions, especially in the context of longitudinal and causal models. Some references include McCullagh and Nelder (1989), Liang, Zeger and Qaqish (1992), Becker (1994), Lang and Agresti (1994), Glonek and McCullagh (1995) and Bergsma (1997). In this paper, we consider a general class of marginal log-linear (MLL) models introduced by Bergsma and Rudas (2002). The MLL parameters are computed from marginals of the joint distribution and are characterized by two subsets of the variables - the relevant marginal and the effect (a subset of the marginal).

The MLL models can often be parameterized by combining log-linear parameters within different marginal tables. The resulting MLL parameterizations provide an elegant and flexible way to parameterize a multivariate discrete probability distribution. These models generalize several well-known models in the literature. Useful submodels can be induced by setting some of the parameters to 0, or more generally by restricting attention to a linear or affine subset of the parameter space. If these zero parameters can be embedded into a larger smooth parameterization of the joint distribution, then the model defined by the conditional independence constraints is a curved exponential family, and therefore possesses good statistical properties. This approach was applied by Rudas et al. (2010) and Forcina et al. (2010) for conditional independence models, and Evans and Richardson (2013) to some classes of graphical models. More recently, Evans (2015) demonstrated smoothness of certain MLL parameterizations.

In this paper, we use marginal models for the analysis of a multidimensional contingency table, which may be quite involved for large dimensional tables. So, it is often useful and convenient to reduce the dimension of the table and examine the condensed (summed over certain variables) table. In a condensed table, some extraneous association between the remaining

2010 Mathematics Subject Classification. Primary 62H17; Secondary 62E99.

Key words and phrases. Marginal log-linear parameters; Contingency table; Collapsibility; Conditional independence; Smooth parameterization.
variables may be introduced. Also, any original relationship between certain variables may be lost and/or the monotonicity of dependence among some variables may be reversed. This paradox, due to Simpson (1951), is known as the Simpson’s paradox. For more examples, see Lindley and Novick (1981), Shapiro (1982), Cox and Wermuth (2003) and Vellaisamy (2012). However, there are certain tables which do not exhibit Simpson’s paradox. In such cases, it is advantageous to condense (more technically, collapse) the original table, especially if the table is sparse or the observed cell counts are small (Ducharne and Lepage (1986)). Hence, it is of practical importance to identify various conditions for collapsibility of a table. With the recent explosion in the amount of available data due to rapid advances in information technology, the problem of data mining has become significant and statistically challenging. Hence, collapsibility may also be considered as a “dimension reduction” technique for data condensation.

Wermuth (1987) studied the parametric collapsibility with respect to odds ratio and relative risk, and Guo and Geng (1995) discussed the collapsibility conditions for logistic regression coefficients. Whittemore (1978) obtained some necessary and sufficient conditions for collapsibility and strict collapsibility for a \( n \)-dimensional table. However, due to arbitrary functional representations and the algebraic approach, the results and their proofs are quite involved and non-intuitive. Vellaisamy and Vijay (2007) studied collapsibility for a multidimensional contingency table using ordinary log-linear parameters, which have simple closed-form expressions. Also, Vellaisamy and Vijay (2009) considered log-linear modelling using conditional structures, while Vellaisamy and Vijay (2010) studied collapsibility of contingency tables based on such conditional log-linear models. Our collapsibility results are based on MLL interaction parameters, which can be easily computed from the tables. We use the Möbius inversion technique, which is well-known in combinatorics (Charalambides (2002)). It also plays an important role for proving the Hammersley-Clifford theorem (see Lauritzen (1996), p. 36 or Hammersley and Clifford (1971)) in graphical models. In this paper, we obtain several results related to collapsibility and strict collapsibility of the more general MLL parameters for modelling categorical data.

The remaining paper is organized as follows. In Section 2, we introduce various terms and notations that are used throughout the paper. We discuss the concept of a marginal model introduced by Bergsma and Rudas (2002) and introduce various definitions of MLL parameters available in the literature. We use a simple and intuitive expression for such parameters subsequently in the paper. Several important properties of these parameters are derived using Möbius inversion. In Section 3, we give a general definition of collapsibility by considering two arbitrary margins of a contingency table and obtain a set of equivalent conditions for collapsibility of MLL parameters. These conditions generalize those obtained by Vellaisamy and Vijay (2007) for ordinary log-linear parameters. We obtain necessary and sufficient conditions for collapsibility with respect to a set of MLL interaction parameters having some common effects, which generalize the results of Vellaisamy and Vijay (2007). Also, a general definition of strict collapsibility is provided and some results for strict collapsibility are derived. In Section 4, we explore the relationship of strict collapsibility with various forms of independence of variables for a multidimensional table. New necessary and sufficient conditions are obtained in each case. In Section 5, we obtain a result relating parameters having a common effect but defined within different margins. This result is then used to show that a smooth MLL parameterization or curved exponential family can be obtained.
under collapsibility conditions. Moreover, some sufficient conditions for collapsibility of MLL parameters in a multidimensional table are also provided using the above result. Section 6 mentions some concluding remarks.

2. Marginal log-linear parameters

First, we introduce some notations and terminology that will be used throughout the paper (see Evans and Richardson (2013)). Let \( V \) be a finite index set. Consider a collection of categorical variables \((X_v)_{v \in V}\) assuming values in finite discrete probability spaces \((\mathcal{X}_v)_{v \in V}\) with respect to a strictly positive probability measure. Without loss of generality, let \( \mathcal{X}_v = \{0, 1, \ldots, |\mathcal{X}_v| - 1\} \). For \( \phi \neq A \subseteq V \), we denote \( \mathcal{X}_A = \times_{v \in A} (X_v) \) and \( \mathcal{X} = \mathcal{X}_V \) and similarly \( X_A = (X_v)_{v \in A} \), \( X \equiv X_V \), \( x_A = (x_v)_{v \in A} \) and \( x \equiv x_V \). Also, let \( \tilde{\mathcal{X}}_v = \{0, 1, \ldots, |\mathcal{X}_v| - 2\} \) and \( \tilde{\mathcal{X}} = \times_{v \in V} (\tilde{\mathcal{X}}_v) \). The marginal distribution of \( X_A \) is denoted by \( p_A = P(X_A = x_A) \) and for disjoint \( A, B \subseteq V \), denote the conditional distribution of \( X_A | X_B \) by \( p_{A|B} = p(x_A | x_B) = P(X_A = x_A | X_B = x_B) \). Note that by assumption, \( p_V > 0 \).

The Cartesian product \( \mathcal{X} = \times_{v \in V} \mathcal{X}_v \) is a \(|V|\)-dimensional contingency table with \( x = (x_1, \ldots, x_{|V|}) \) for \( x_v \in \mathcal{X}_v \) being a cell of the table. A cell frequency for cell \( x \) is given by a non-negative real number \( n(x) \). A marginal cell frequency corresponding to the marginal table \( \mathcal{X}_A = \times_{v \in A} \mathcal{X}_v \) is denoted by \( n_A(x_A) = \sum_{j \in \mathcal{X}_j:j_A=x_A} n(j) \). If \( \mathcal{F} \) is the class of strictly positive frequency distributions \( n = (n(x))_{x \in \mathcal{X}} \) on \( \mathcal{X} \), then a function \( \theta : \mathcal{F} \rightarrow \mathbb{R}^k, k \geq 1 \) is called a parameter of \( \mathcal{F} \). For an open set \( D \subseteq \mathbb{R}^k, \theta : \mathcal{F} \rightarrow D \) is a \( p \)-dimensional \((1 \leq p \leq k)\) smooth parameterization of \( \mathcal{F} \) if it is a homeomorphism onto \( D \), is twice continuously differentiable and its Jacobian has full rank \( p \) everywhere. If \( \mathcal{F} \) is an exponential family, then a model \( \mathcal{G} \subseteq \mathcal{F} \) is called curved exponential if it has a smooth parameterization. The vector \( \theta = (\theta_1, \ldots, \theta_k) \) is said to be variation independent if \( \Theta = \Theta_1 \times \cdots \times \Theta_k \), where \( \theta \in \Theta \) and \( \theta_i \in \Theta_i \) for \( 1 \leq i \leq k \).

Next, we discuss the concept of marginal models introduced by Bergsma and Rudas (2002).

**Definition 2.1.** Let \( \mathcal{P} = \{(P, Q)\} \) be a collection of ordered pairs of subsets of \( V \) such that \( P \subseteq Q \subseteq V \). Define

\[
(Q = \{Q|(P, Q) \in \mathcal{P} \text{ for some } P \subseteq Q\})
\]

(2.1)

to be the collection of margins in \( \mathcal{P} \). If \( \mathcal{Q} = \{Q_1, \ldots, Q_t\} \), then define \( \mathbb{P}_i = \{P|(P, Q_i) \in \mathcal{P}\} \) to be the collection of effects defined within the margin \( Q_i \). We call \( \mathcal{P} \) hierarchical if there is an ordering on \( \mathcal{Q} \) such that \( i < j \Rightarrow Q_j \not\subset Q_i \) (the sequence of margins is ascending) and \( P \in \mathbb{P}_j \Rightarrow P \not\in \mathbb{P}_i \). The second condition means that every effect is contained only within the first margin of which it is a subset. We call \( \mathcal{P} \) complete if for every non-empty \( P \subseteq V \), there is exactly one \( Q_i \in \mathcal{Q} \) such that \((P, Q_i) \in \mathcal{P}\).

In other words, \( \mathcal{P} \) is said to be hierarchical because each effect is a subset of the first possible margin in an ascending class of sets. Similarly, \( \mathcal{P} \) is said to be complete because all effects are considered, each defined only once (in a specific margin). Note that a pair \((P, Q_i)\) represents a log-linear interaction over \( P \) within \( Q_i \). Next, we provide various definitions of a MLL parameter.

The following definition is due to Bergsma and Rudas (2002).
Definition 2.2. For \( x \in \mathcal{X} \) and \( \phi \neq L \subseteq M \subseteq V \), the marginal log-linear parameters for a marginal \( M \) are defined recursively as follows.

\[
\lambda^M_{\phi}(x_\phi) = \frac{1}{|\mathcal{X}_M|} \sum_{j_M \in \mathcal{X}_M} \log p_M(j_M);
\]

\[
\lambda^M_L(x_L) = \frac{1}{|\mathcal{X}_{M \setminus L}|} \sum_{j_M \in \mathcal{X}_{M \setminus L} : j_L = x_L} \log p_M(j_M) - \sum_{L' \subset L} \lambda^M_{L'}(x_{L'}).
\]

Definition 2.2 represents effect coding while Glonek and McCullagh (1995) used dummy coding, which yields different but equivalent parameters. The next definition provides an equivalent expression of a MLL parameter (see Evans (2011)).

Definition 2.3. For \( L \subseteq M \subseteq V \) and \( x_L \in \mathcal{X}_L \), let

\[
\nu^M_L(x_L) = \frac{1}{|\mathcal{X}_{M \setminus L}|} \sum_{j_M \in \mathcal{X}_{M \setminus L} : j_L = x_L} \log p_M(j_M)
\]

and

\[
\lambda^M_L(x_L) = \sum_{L' \subseteq L} (-1)^{|L \setminus L'|} \nu^M_{L'}(x_{L'})
\]

\[
= \sum_{L' \subseteq L} (-1)^{|L \setminus L'|} \frac{1}{|\mathcal{X}_{M \setminus L'}|} \sum_{j_M \in \mathcal{X}_{M \setminus L'} : j_{L'} = x_{L'}} \log p_M(j_M).
\]

Then \( \lambda^M_L(x_L) \) is called a marginal log-linear parameter.

Note that for \( L = \phi \), (2.5) reduces to (2.2). Evans and Richardson (2013) (see their Lemma 1) provided another expression of a MLL parameter as shown below.

Definition 2.4. For \( L \subseteq M \subseteq V \) and \( x_L \in \mathcal{X}_L \), a marginal log-linear parameter is given by

\[
\lambda^M_L(x_L) = 2^{-|M|} \sum_{x_M \in \{0,1\}^{|M|}} \log p_M(x_M) \prod_{v \in L} (I_{\{x_v = 1\}} - 1),
\]

where \( I_A \) is the indicator function of \( A \).

It can be shown that (2.5) and (2.6) are equivalent. However, we will be using (2.5) subsequently in our paper as it is more explicit and more intuitive than (2.6). For identifiability purpose, the sum over any argument (support of a variable) of a MLL parameter is 0 but other constraints may also be used.

If we have a finite collection of binary random variables \( X_v \in \{0,1\} \) for \( v \in V \), then Evans (2015) defined a MLL parameter as follows.

Definition 2.5. For \( L \subseteq M \subseteq V \), a marginal log-linear parameter is given by

\[
\lambda^M_L(x_L) = 2^{-|M|} \sum_{x_M \in \{0,1\}^{|M|}} (-1)^{|x_L|} \log p_M(x_M),
\]

where \( |x_L| = \sum_{v \in L} x_v \) is the number of 1’s in \( x_L \in \{0,1\}^{|L|} \).
However, in this paper, we are concerned with the general case in which \( X_v \) for \( v \in V \) is arbitrary, that is, a categorical variable is not necessarily binary. Next, we consider a set of MLL parameters, which is specified by a set of ordered pairs. Let
\[
(2.8)
\lambda_P = \{ \lambda^M(x_L) | (L, M) \in P, x_L \in X_L \}.
\]
be the collection of MLL parameters corresponding to \( P \) (see Definition 2.1). To avoid redundancy due to the identifiability constraints, consider only \( x_L \in \tilde{X}_L \) so that the MLL parameters \( \tilde{\lambda}^M_L(x_L) : \mathcal{F} \to \mathbb{R}^{[X]} \) are elements of \( \tilde{\lambda}_P \). Bergsma and Rudas (2002) (Theorem 2) showed that if \( P \) is hierarchical and complete, then \( \tilde{\lambda}_P \) is a smooth parameterization of \( \mathcal{F} \) (the saturated model). Theorem 4 of Bergsma and Rudas (2002) shows that for a complete and hierarchical \( P \), \( \tilde{\lambda}_P \) is variation independent if and only if \( P \) is ordered decomposable. Let \( \lambda^M_L \) denote the collection \( \{ \lambda^M_L(x_L) | x_L \in X_L \} \). Important special cases of \( \lambda_P \) are the classical or ordinary log-linear parameters and the multivariate logistic parameters (McCullagh and Nelder (1989) and Glonek and McCullagh (1995)) denoted by \( \{ \lambda^V_L | L \subseteq V \} \) and \( \{ \lambda^E_L | L \subseteq V \} \) respectively. Glonek (1996) considered a mixture of the above parameters, but the MLL parameters are more general than them.

2.1. Properties of marginal log-linear parameters. In this section, we provide some results related to MLL parameters. Henceforth, all subsets of \( V \) are assumed to be non-empty unless stated otherwise. The next result extends a result of Vellaisamy and Vijay (2007) for ordinary log-linear parameters to MLL parameters. It is also stated without proof in Evans and Richardson (2013).

Lemma 2.1. For every \( L \subseteq M \), the marginal log-linear parameters \( \lambda^M_L \) satisfy
\[
(2.9)
\sum_{x_v} \lambda^M_L(x_L) = \sum_{x_v} \lambda^M_L(x_v, x_{L \setminus \{v\}}) = 0 \quad \forall x_L \in X_L, \ v \in L.
\]

Proof. Let \( L_v = L \setminus \{v\} \) and \( L'_v \subseteq L_v \) for any \( v \in L \). To prove the result, we need the following facts.

1. For \( L' \subseteq L \) and \( v \notin L' \), we have \( L' = L'_v \), \( x_v \in X_{L'v} = X_{L'_v} \) and
\[
(2.10)
\frac{1}{|X_v|} \sum_{x_v} \nu^M_{L'_v}(x_{L'_v}) = \frac{1}{|X_v|} \sum_{x_v} \left[ \frac{1}{|X_{M \setminus L'_v}|} \sum_{j_M \in X_{M \setminus L'_v}, x_{L'_v} = x_{L'_v}} \log p_M(j_M) \right]
\]

2. For \( L' \subseteq L \) and \( v \in L' \), we have \( L' = L'_v \), \( x_v \in X_L = X_{L'_v} \) and
\[
(2.11)
\frac{1}{|X_v|} \sum_{x_v} \nu^M_{L'_v}(x_{L'_v}) = \frac{1}{|X_v|} \sum_{x_v} \left[ \frac{1}{|X_{M \setminus L'_v}|} \sum_{j_M \in X_{M \setminus L'_v}, x_{L'_v} = x_{L'_v}} \log p_M(j_M) \right]
\]
For Lemma 2.2, which completes the proof.

Lemma 2.1

Substituting (2.14) in (2.13) and repeating the steps for each

The second last line of (2.11) follows from the fact that

which implies

\[ |\mathcal{X}_{M \setminus L_v'}| = |\mathcal{X}_{(M \setminus L_v') \cup \{v\}}| = |\mathcal{X}_{M \setminus L_v'}| \]

Similarly,

\[ \sum_{x_v} \sum_{j_M \in \mathcal{X}_{M \setminus L_v'}} \log p_M(j_M) = \sum_{j_M \in \mathcal{X}_{M \setminus L_v'}} \log p_M(j_M) = \sum_{j_M \in \mathcal{X}_{M \setminus L_v'}} \log p_M(j_M). \]

Note that\[ \log 0 = -\infty. \]

Hence, using (2.10) and (2.11), we have from (2.5),

\[ \sum_{x_v} \lambda_L^M(x_L) = \sum_{x_v} \sum_{L' \subseteq L} (-1)^{|L \setminus L'|} \nu_L^M(x_L') \]

\[ = \left| \mathcal{X}_v \right| \sum_{L' \subseteq L, v \in L'} (-1)^{|L \setminus L'|} \left( \frac{1}{|\mathcal{X}_v|} \sum_{x_v} \nu_L^M(x_L') \right) + \sum_{L' \subseteq L, v \notin L'} (-1)^{|L \setminus L'|} \left( \frac{1}{|\mathcal{X}_v|} \sum_{x_v} \nu_L^M(x_L') \right) \]

\[ = |\mathcal{X}_v| \left[ \sum_{L' \subseteq L} (-1)^{|L_v \setminus L_v'|} \nu_L^M(x_L) + \sum_{L' \subseteq L} (-1)^{|L_v \setminus L_v'|} \nu_L^M(x_L) \right] \]

\[ = |\mathcal{X}_v| \left[ \sum_{L' \subseteq L} \left\{ (-1)^{|L_v \setminus L_v'|} (1 - 1) \right\} \nu_L^M(x_L) \right] \]

\[ = 0, \]

which completes the proof. \(\square\)

**Lemma 2.2.** For \( L \subseteq N \subseteq M \), we have

\[ \sum_{L' \subseteq L} \lambda_L^M(x_L') = \sum_{L' \subseteq L} \lambda_L^N(x_L') \Leftrightarrow \lambda_L^M(x_L') = \lambda_L^N(x_L') \quad \forall \, \phi \neq L' \subseteq L. \]

**Proof.** The reverse implication is obvious. For the forward implication, we assume

\[ \sum_{L' \subseteq L} \lambda_L^M(x_L') = \sum_{L' \subseteq L} \lambda_L^N(x_L'). \]

First fix \( L' \subseteq L \). Then summing both sides of (2.13) over \( L'' \subseteq L \) with \( L'' \neq L' \), we get from Lemma 2.1

\[ \lambda_L^M(x_L') = \lambda_L^N(x_L'). \]

Substituting (2.14) in (2.13) and repeating the steps for each \( L' \), we obtain

\[ \lambda_L^M(x_L') = \lambda_L^N(x_L') \quad \forall \, \phi \neq L' \subseteq L, \]

which completes the proof. \(\square\)
The next lemma generalizes a result of Vellaisamy and Vijay (2007) for ordinary log-linear parameters.

**Lemma 2.3.** The log-linear marginal model is given by

\[
(2.16) \quad \log p_M(x_M) = \sum_{L \subseteq M} \lambda^M_L(x_L)
\]

if and only if

\[
(2.17) \quad \nu^M_A(x_A) = \sum_{L \subseteq A} \lambda^M_L(x_L) \quad \forall A \subseteq M,
\]

where \( \nu^M_A \) is given by \((2.4)\).

**Proof.** Let \( \log p_M(x_M) = \sum_{L \subseteq M} \lambda^M_L(x_L) \), where \( \lambda^M_L \) satisfies Lemma 2.1. Then

\[
\nu^M_A(x_A) = \frac{1}{|X_{M \setminus A}|} \sum_{j_M \in X_{M \setminus A}, j_A = x_A} \log p_M(j_M)
= \frac{1}{|X_{M \setminus A}|} \sum_{j_M \in X_{M \setminus A}, j_A = x_A} \sum_{L \subseteq M} \lambda^M_L(j_L)
= \frac{1}{|X_{M \setminus A}|} \left[ \sum_{j_M \in X_{M \setminus A}, j_A = x_A} \left( \sum_{L : L \cap A^c = \phi} \lambda^M_L(j_L) + \sum_{L : L \cap A^c \neq \phi} \lambda^M_L(j_L) \right) \right]
= \frac{1}{|X_{M \setminus A}|} \times |X_{M \setminus A}| \sum_{L \subseteq A} \lambda^M_L(x_L) + \frac{1}{|X_{M \setminus A}|} \sum_{L : L \cap A^c = \phi \neq \phi} \sum_{j_M \in X_M, j_L = x_L, j_L \cap A = x_L \cap A} \lambda^M_L(j_L)
= \sum_{L \subseteq A} \lambda^M_L(x_L)
\]

since \( \sum_{j_M \in X_{M \setminus A} : L \cap A^c = x_L \cap A^c} \lambda^M_L(j_L) = 0 \) by Lemma 2.1. For the sufficiency part, observe that by substituting \( A = M \) in \((2.17)\), we have \( \nu^M_M = \log p_M(x_M) \) (LHS of \((2.16)\)) using \((2.4)\). Also, the RHS of \((2.16)\) and \((2.17)\) become identical. \(\square\)

Note that from \((2.4)\), using Möbius inversion for \( L \subseteq M \subseteq V \) and \( x_L \in X_L \), we have

\[
(2.18) \quad \lambda^M_L(x_L) = \sum_{L' \subseteq L} (-1)^{|L \setminus L'|} \nu^M_{L'}(x_{L'}) \iff \nu^M_L(x_L) = \sum_{L' \subseteq L} \lambda^M_{L'}(x_{L'}),
\]

where \( \nu^M_L \) is given by \((2.4)\). The next result states the restrictions on MLL parameters for conditional independence models (see Lemma 1 of Rudas et al. (2010) or Eq. (6) of Forcina et al. (2010)). Henceforth in this paper, we denote \( S \cup T \) as \( ST \).

**Lemma 2.4.** Consider three disjoint subsets \( A, B \) and \( C \) of \( V \), where \( C \) may be empty. Then \( X_A \independent X_B | X_C \) if and only if

\[
(2.19) \quad \lambda^{ABC}_{A'B'C'} = 0 \quad \text{for every} \ A' \subseteq A, B' \subseteq B, C' \subseteq C.
\]
Note that if $C = \phi$ in Lemma 2.4 then we have marginal independence between $X_A$ and $X_B$. The result in this case was proved for multivariate logistic parameters by Kauermann (1997).

We next state another result linking MLL parameters to conditional independence, which appears as Lemma 3 of Evans and Richardson (2013). However, we provide an alternative proof using (2.5).

**Lemma 2.5.** Consider three disjoint subsets $A$, $B$ and $C$ of $V$, where $A$ is non-empty. If $X_A \perp X_B|X_C$, then for any $D \subseteq C$,

$$(2.20) \quad \lambda_{AD}^{ABC}(x_{AD}) = \lambda_{AD}^{AC}(x_{AD}), \quad \forall x_{AD} \in \mathcal{X}_{AD}.$$  

**Proof.** Since $X_A \perp X_B|X_C$, we have

$$p_{ABC}(x_{ABC}) = p_{AB|C}(x_{AB|C}) p_C(x_C)$$

$$= p_{AC}(x_AC) p_{BC|C}(x_B|C) p_C(x_C),$$

which implies

$$\lambda_{AD}^{ABC}(x_{AD}) = \sum_{L' \subseteq AD} (-1)^{|AD\setminus L'|} \frac{1}{|\mathcal{X}_{ABC\setminus L'}|} \sum_{j_{ABC} \in \mathcal{X}_{ABC}} \sum_{j_{L'=x_{L'}}} \log p_{ABC}(j_{ABC})$$

$$= \sum_{L' \subseteq AD} (-1)^{|AD\setminus L'|} \frac{1}{|\mathcal{X}_{ABC\setminus L'}|} \sum_{j_{ABC} \in \mathcal{X}_{ABC}} \sum_{j_{AC} = x_{AC}} \sum_{j_{B|C}} \log p_{AC}(j_{AC})$$

$$+ \sum_{L' \subseteq AD} (-1)^{|AD\setminus L'|} \frac{1}{|\mathcal{X}_{ABC\setminus L'}|} \sum_{j_{ABC} \in \mathcal{X}_{ABC}} \sum_{j_{BC} = x_{BC}} \sum_{j_{B|C}} \log p_{BC|C}(j_{B|C}|j_C)$$

$$(2.21) \quad = A_1 + A_2 \quad \text{(say)}.$$  

Now

$$A_1 = \sum_{L' \subseteq AD} (-1)^{|AD\setminus L'|} \frac{1}{|\mathcal{X}_{AC}\setminus L'|} \sum_{j_{B|C} = x_{B|C}} \sum_{j_{AC} = x_{AC}} \log p_{AC}(j_{AC})$$

$$= \sum_{L' \subseteq AD} (-1)^{|AD\setminus L'|} \frac{1}{|\mathcal{X}_{AC\setminus L'}|} \sum_{j_{AC} = x_{AC}} \sum_{j_{B|C}} \log p_{AC}(j_{AC})$$

$$(2.22) \quad = \lambda_{AD}^{AC}(x_{AD}).$$

Similarly,

$$A_2 = \sum_{L' \subseteq AD} (-1)^{|AD\setminus L'|} \frac{1}{|\mathcal{X}_{BC}\setminus L'|} \sum_{j_{BC} = x_{BC}} \sum_{j_{B|C}} \log p_{BC|C}(j_{B|C}|j_C)$$

$$= \sum_{L' \subseteq AD} (-1)^{|AD\setminus L'|} \frac{1}{|\mathcal{X}_{BC}\setminus L'|} \sum_{j_{BC} = x_{BC}} \sum_{j_{B|C}} \log p_{BC|C}(j_{B|C}|j_C)$$
Remark 2.1. Note that Lemma 2.5 holds for ordinary log-linear parameters only if $V = ABC$ and the parametrizations in RHS and LHS of (2.20) are different.

The next result provides the relationship between certain conditional distribution parameters and MLL parameters.

Lemma 2.6. For a fixed $L \subseteq M \subseteq V$ with $N = M \setminus L$, define $\kappa_{L|N}(x_L|x_N) = \sum_{L \subseteq A \subseteq M} \lambda^M_A(x_A).$ Then $\kappa_{L|N}(x_L|x_N) = (-1)^{|L|} \sum_{B \subseteq N} \lambda^M_B(x_B).$

Proof. Note that

\[
\kappa_{L|N}(x_L|x_N) = \sum_{L \subseteq A \subseteq M} \lambda^M_A(x_A) = \sum_{L \subseteq A \subseteq M} \sum_{L' \subseteq A} (-1)^{|A'|} \frac{1}{|\mathbf{x}_{B'C}|} \sum_{j_{BC} \in \mathbf{x}_{BC}} \log p_B(j_{B|C})
\]

Now

\[
B_1 = \sum_{L' \subseteq A} (-1)^{|A\setminus L'|} \frac{1}{|\mathbf{x}_{BC}|} \sum_{j_{BC} \in \mathbf{x}_{BC}} \log p_B(j_{B|C}) (\because A \cap D = L' \cap B = L' \cap C = \phi)
\]

\[
B_2 = \sum_{L' \subseteq A} (-1)^{|A\setminus L'|} \frac{1}{|\mathbf{x}_{BC}|} \sum_{j_{BC} \in \mathbf{x}_{BC}} \log p_B(j_{B|C}) (\because L' \cap B = \phi; D \subseteq C)
\]

\[
= \sum_{A' \subseteq A} \sum_{\phi \neq D' \subseteq D} (-1)^{|A\setminus A'|} \frac{1}{|\mathbf{x}_{BC}\setminus A'\setminus D'|} \sum_{j_{BC} \in \mathbf{x}_{BC}} \log p_B(j_{B|C})
\]

(2.23) \[= B_1 + B_2 \quad \text{(say)}.
\]

Hence, the result follows from (2.21)-(2.25). \[\square\]

Remark 2.1. Note that Lemma 2.5 holds for ordinary log-linear parameters only if $V = ABC$ and the parametrizations in RHS and LHS of (2.20) are different.
Similarly,
\[
\begin{align*}
C_2 &= \sum\sum_{L \subseteq A \subseteq M \text{ s.t. } L' \subseteq A \setminus L} (-1)^{|A\setminus L'|} \frac{1}{|\mathcal{X}_{M \setminus L'}|} \sum_{j_M \in \mathcal{X}_{M \setminus j_L = x_L}} \log p_M(j_M) \\
&= \sum\sum_{B \subseteq N \text{ s.t. } L' \subseteq B} (-1)^{|L| + |B \setminus L'|} \frac{1}{|\mathcal{X}_{M \setminus L'}|} \sum_{j_M \in \mathcal{X}_{M \setminus j_L = x_L}} \log p_M(j_M) \quad (\because B = A \setminus L) \\
&= (-1)^{|L|} \sum\sum_{B \subseteq N \text{ s.t. } L' \subseteq B} (-1)^{|B \setminus L'|} \frac{1}{|\mathcal{X}_{M \setminus L'}|} \sum_{j_M \in \mathcal{X}_{M \setminus j_L = x_L}} \log p_M(j_M) \\
&= (-1)^{|L|} \sum_{B \subseteq N} \lambda_B^M(x_B).
\end{align*}
\]

Hence, the result follows from (2.26) – (2.28).

By Lemma 5 of Evans and Richardson (2013), for fixed \( M \) and \( L \subseteq M \), the set of MLL parameters \( \{\lambda_A^M(x_M)\mid L \subseteq A \subseteq M, x_M \in \mathcal{X}_M\} \) together with the \((|L| - 1)\)-dimensional marginal distributions of \( X_L \) conditional on \( X_{M \setminus L} \) smoothly parameterizes the conditional distribution of \( X_L \) given \( X_{M \setminus L} \).
3. Collapsibility

In this section, we establish some necessary and sufficient conditions for collapsibility of a multidimensional contingency table with respect to MLL parameters. Recall that \( \mathcal{F} \) is the class of positive frequency distributions \( n \) on \( \mathcal{X} \) and \( \tilde{\lambda}_P \) is the collection of non-redundant MLL parameters corresponding to \( P \). Bergsma and Rudas (2002) defined a log-affine marginal model \( M_P(q, H) \subseteq \mathcal{F} \) by the constraint

\[
n \in M_P(q, H) \iff \tilde{\lambda}_P \in q + H,
\]

where \( P \) is fixed, \( H \) is a nonempty linear subspace of \( \mathbb{R}^{\dim \tilde{\lambda}_P} \) and \( q \in \mathbb{R}^{\dim \tilde{\lambda}_P} \). Note that log-affine marginal models generalize ordinary log-linear models, log-affine models (Habermann (1974), Rudas and Leimer (1992) and Lauritzen (1996)), multivariate logistic models of McCullagh and Nelder (1989) and Glonek and McCullagh (1995), and the mixture of these models (Glonek (1996)). If \( q = 0 \), then \( M_P(0, H) \) is called a log-linear marginal model.

Whittemore (1978) and Vellaisamy and Vijay (2007), besides others, studied collapsibility of ordinary log-linear parameters for a multidimensional contingency table. Bergsma and Rudas (2002) mentioned that one can specify collapsibility conditions using log-linear marginal models \( M_P(0, H) \) with \( P \) nonhierarchal. They defined a complete table \( \mathcal{X} \) to be collapsible into the marginal table \( \mathcal{X}_M \) with respect to \( L \subseteq M \) if

\[
\lambda^V_L(x_L) = \lambda^M_L(x_L), \quad \forall x_L \in \mathcal{X}_L.
\]

This condition implies that the amount of information about the interaction among the variables in \( L \) is the same in \( \mathcal{X} \) and \( \mathcal{X}_M \). Suppose now one is interested in studying the association among variables in \( L \) only within some strict subsets (margins) of \( V \). An application would be to check collapsibility of the corresponding marginal tables or conditional independence of variables in those margins only. For this purpose, we define collapsibility by considering two arbitrary marginal tables of \( \mathcal{X} \). Specifically, we consider the marginal tables \( \mathcal{X}_M \) and \( \mathcal{X}_N \), where \( M \subseteq N \subseteq V \) with \( |V| \geq 2 \). If \( N = V \), then our definition reduces to Bergsma and Rudas’ (2002) definition.

**Definition 3.1.** A \( |N| \)-dimensional table \( \mathcal{X}_N \) is collapsible (over \( N \setminus M \)) into a \( |M| \)-dimensional table \( \mathcal{X}_M \) with respect to \( \lambda^M_L \), where \( L \subseteq M \subseteq N \) if

\[
\lambda^M_L(x_L) = \lambda^N_L(x_L), \quad \forall x_L \in \mathcal{X}_L.
\]

Define now

\[
d^{(M)}(x_M) = \log p_M(x_M) - \nu^N_M(x_M)
\]

and for any \( Z \subseteq M \)

\[
d^{(M)}_Z(x_Z) = \frac{1}{|\mathcal{X}_M \setminus Z|} \sum_{j_M \in \mathcal{X}_M : j_Z = x_Z} d^{(M)}(j_M).
\]
Then the following result generalizes Theorem 3.1 of Vellaisamy and Vijay (2007) for classical log-linear parameters. It characterizes the conditions for which collapsibility holds with respect to $\lambda_L^M$.

**Theorem 3.1.** Let $L \subseteq M \subseteq N$ and and $\delta_Z = (\lambda_L^M - \lambda_N^N)$ for $Z \subseteq M$. Also, let $\log p_M(x_M) = \sum_{Z \subseteq M} \lambda_Z^M(x_Z)$ and $\log p_N(x_N) = \sum_{Z \subseteq N} \lambda_Z^N(x_Z)$ be the log-linear marginal models corresponding to the marginal tables $\mathcal{X}_M$ and $\mathcal{X}_N$ respectively. Then the following conditions are equivalent.

1. $\delta_L(x_L) = 0$;
2. $\tilde{d}_L^M(x_L) = \sum_{Z \subseteq L} \delta_Z(x_Z)$;
3. $\sum_{Z \subseteq L} (-1)^{|L \setminus Z|} \tilde{d}_Z^M(x_Z) = 0$.

**Proof.** From (3.3), we have

$$\log p_M(x_M) = \nu_M^N(x_M) + d^M(x_M).$$

Also, from (3.4) and (3.5), we have for any $Z \subseteq M$

$$\nu_Z^M(x_Z) = \frac{1}{|\mathcal{X}_M|} \sum_{j_M \in \mathcal{X}_M: j_Z = x_Z} \log p_M(j_M)$$

$$= \frac{1}{|\mathcal{X}_M|} \sum_{j_M \in \mathcal{X}_M: j_Z = x_Z} [\nu_M^N(j_M) + d^M(j_M)]$$

$$= \frac{1}{|\mathcal{X}_M|} \sum_{j_M \in \mathcal{X}_M} \sum_{i_M \in \mathcal{X}_M} \log p_N(i_M) + \frac{1}{|\mathcal{X}_M|} \sum_{j_M \in \mathcal{X}_M} \tilde{d}_Z^M(x_Z)$$

$$= \nu_Z^N(x_Z) + \tilde{d}_Z^M(x_Z),$$

(3.6)

Now using Lemma 2.3 we get

$$\tilde{d}_Z^M(x_Z) = \nu_Z^M(x_Z) - \nu_Z^N(x_Z)$$

$$= \sum_{A \subseteq Z} (\lambda_A^M(x_A) - \lambda_A^N(x_A))$$

$$= \sum_{A \subseteq Z} \delta_A(x_A).$$

(3.7)

Applying Möbius inversion to (3.7) with $Z = L$ leads to

$$\delta_L(x_L) = \sum_{A \subseteq L} (-1)^{|L \setminus A|} \tilde{d}_A^M(x_A).$$

(3.8)

Thus (i) $\iff$ (ii) follows from (3.7). Similarly, (i) $\iff$ (iii) follows from (3.8).  

Note that a log-linear marginal model $\log p_M(x_M) = \sum_{Z \subseteq M} \lambda_Z^M(x_Z)$ is said to be hierarchical if $\lambda_L^M \neq 0 \Rightarrow \lambda_L^M \neq 0$ or equivalently $\lambda_L^M = 0 \Rightarrow \lambda_L^M = 0$ for $L' \subset L$. As shown in the example on pp. 564-565 of Vellaisamy and Vijay (2007), if one is interested in studying the
association (say conditional independence) between a particular variable with other variables, then the full table can be collapsed into a marginal table with respect to MLL parameters involving that variable. This motivates the case for studying collapsibility with respect to more than one MLL parameter. Specifically, Example 3.2 of Vellaisamy and Vijay (2007) motivates collapsibility with respect to MLL parameters having two common effects. The next result considers collapsibility with respect to MLL parameters having a set of common effects, and generalizes Theorems 3.2 and 3.3 of Vellaisamy and Vijay (2007) for ordinary log-linear parameters.

**Theorem 3.2.** Let \( L \subseteq M \subset N \subseteq V \). A \(|N|\)-dimensional table \( \mathbf{x}_N \) is collapsible (over \( N \setminus M \)) into a \(|M|\)-dimensional table \( \mathbf{x}_M \) with respect to the set \( C_S = \{ \lambda_A^N | S \subseteq A \subseteq L \} \) of MLL parameters if and only if

\[
\sum_{R \subseteq S} (-1)^{|S \setminus R|} \lambda^M_R(x_{LR}) = 0,
\]

where \( L_R = L \setminus R \) and \( 1 \leq |S| \leq |M| \).

**Proof.** Using (3.6), note that the condition (3.9) is equivalent to

\[
\sum_{R \subseteq S} (-1)^{|S \setminus R|} [\nu^M_L(x_{LR}) - \nu^N_L(x_{LR})] = 0.
\]

Using Lemma 2.3, (3.10) reduces to

\[
\sum_{R \subseteq S} (-1)^{|S \setminus R|} \left[ \sum_{Z \subseteq L} \lambda^M_Z(x_Z) - \sum_{Z \subseteq L_R} \lambda^N_Z(x_Z) \right] = 0
\]

or equivalently

\[
\sum_{Z \subseteq L_S} \lambda^M_{Z \cup S}(x_Z, x_S) = \sum_{Z \subseteq L_S} \lambda^N_{Z \cup S}(x_Z, x_S).
\]

We next show that (3.12) holds if and only if the \(|N|\)-dimensional table \( \mathbf{x}_N \) is collapsible with respect to \( \lambda_A^N \) for any \( A \subseteq L \) and \( S \subseteq A \).

First suppose (3.12) holds. Summing both sides over all \( x_j (j \in Z) \) except \( x_S \), we obtain using Lemma 2.1

\[
\lambda^M_S(x_S) = \lambda^N_S(x_S).
\]

This is because \( \sum_{Z \subseteq L_S} \lambda^M_{Z \cup S}(x_Z, x_S) = 0 \) for every \( j \in Z \) and every nonempty set \( Z \subseteq L_S \) and similarly for \( \sum_{Z \subseteq L_S} \lambda^N_{Z \cup S}(x_Z, x_S) \). Substituting (3.13) in (3.12), we get

\[
\sum_{Z \subseteq L_S, Z \neq \phi} \lambda^M_{Z \cup S}(x_Z, x_S) = \sum_{Z \subseteq L_S, Z \neq \phi} \lambda^N_{Z \cup S}(x_Z, x_S).
\]

Now summing over \( x_m \) for all \( m \in L \setminus \{ j, S \} \) in (3.14), we have

\[
\lambda^M_{jS}(x_j, x_S) = \lambda^N_{jS}(x_j, x_S).
\]

Repeating this process leads to

\[
\lambda^M_A(x_A) = \lambda^N_A(x_A) \quad \forall \ A \subseteq L, S \subseteq A.
\]

Hence, the \(|N|\)-dimensional table \( \mathbf{x}_N \) is collapsible into the \(|M|\)-dimensional table \( \mathbf{x}_M \) with respect to the set \( C_S = \{ \lambda_A^N | S \subseteq A \subseteq L \} \).
Conversely, assume now $\mathfrak{X}_N$ is collapsible with respect to the set $C_S$ so that
\begin{equation}
\lambda^M_A(x_A) = \lambda^N_A(x_A) \quad \forall x_A \in \mathfrak{X}_A,
\end{equation}
where $A \subseteq L$ and $S \subseteq A$. Then
\begin{equation}
\sum_{A \subseteq L; S \subseteq L} \lambda^M_A(x_A) = \sum_{A \subseteq L; S \subseteq L} \lambda^N_A(x_A) \quad \forall x_A \in \mathfrak{X}_A,
\end{equation}
which is equivalent to
\begin{equation}
\sum_{Z \subseteq L_S} \lambda^M_{Z \cup S}(x_Z, x_S) = \sum_{Z \subseteq L_S} \lambda^N_{Z \cup S}(x_Z, x_S),
\end{equation}
that is, \((3.12)\) holds. This completes the proof. \hfill \square

Remark 3.1. (i) Consider the collapsibility of a 3-dimensional table into a 2-dimensional one with respect to $\lambda_L^{123}$. There is only one possibility of $A = L$. Hence, \((3.9)\) is necessary and sufficient for collapsibility with respect to $\lambda_L^{123}$. Also, in this case, \((3.9)\) reduces to Condition (iii) of Theorem 3.1. For a multidimensional table, if we consider larger subsets of $A$ until $A = L$, then the necessary and sufficient condition for collapsibility with respect to $\lambda_L^N$ is given by Condition (iii) of Theorem 3.1.

(ii) Parameters of a model are said to be upwards compatible (see Roverato and Lupparelli (2013)) if they are invariant with respect to marginalization and their interpretations remain same when a submodel is chosen. Note that the multivariate logistic parameters satisfy the upward compatibility property because each such parameter can be compared within different marginal distributions. For example, if $X_i \in \{0, 1\}$ for $1 \leq i \leq 3$, then the value of the multivariate logistic parameter $\lambda_{12}^{12}(1, 1) = \frac{1}{4} \log \left( \frac{p_{012} p_{102}}{p_{002} p_{112}} \right)$ does not change whether it is calculated in the $X_1X_2$-marginal or the full $X_1X_2X_3$ table. This implies that these parameters are always collapsible by definition. Hence, any higher dimensional marginal table can always be collapsed into a lower dimensional one with respect to a multivariate logistic parameter.

3.1. Strict Collapsibility. We next consider a stronger version of collapsibility, namely, strict collapsibility. Whittemore (1978) and Vellaisamy and Vijay (2007), among others, studied strict collapsibility of ordinary log-linear parameters for a multidimensional contingency table. For marginal models, Bergsma and Rudas (2002) mentioned that in addition to $\lambda^M_L(x_L) = \lambda^M_L(x_L)$ for all $x_L \in \mathfrak{X}_L$, if $\lambda^N_K(x_K) = 0$ for all $L \subset K \not\subset M$, then the full table $\mathfrak{X}_V$ is said to be strictly collapsible into the marginal table $\mathfrak{X}_M$ with respect to $L$. This implies that the association between the variables in $L$ is the same in both the tables $\mathfrak{X}_V$ and $\mathfrak{X}_M$ conditionally on any subset of variables not in $M$. Theorem 7 of Bergsma and Rudas (2002) shows that such strict collapsibility conditions can always be imposed upon a log-linear marginal model. Similar to Definition 3.1, we provide a definition of strict collapsibility as follows.

Definition 3.2. A $|N|$-dimensional table $\mathfrak{X}_N$ is strictly collapsible over $N \setminus M$ into a $|M|$-dimensional table $\mathfrak{X}_M$ with respect to $\lambda^N_L$, where $\phi \neq L \subseteq M \subset N$ if

(i) $\lambda^M_L(x_L) = \lambda^N_L(x_L) \quad \forall x_L \in \mathfrak{X}_L,$

(ii) $\lambda^N_L(x_Z) = 0 \quad \forall L \subset Z \not\subset M, Z \subset N.$

The result below gives an equivalent expression for Condition (ii) of Definition 3.2 and generalizes Lemma 4.1 of Vellaisamy and Vijay (2007).
Lemma 3.1. For a \(|N|\)-dimensional table \(\mathbf{X}_N\), we have
\[
\lambda_N^Z(x_Z) = 0 \quad \forall \ L \subset Z \subseteq N, Z \cap (N\setminus M) \neq \phi
\]
if and only if
\[
(3.20) \quad \sum_{\substack{L \subset Z \subseteq N \\mid \\text{Z \cap (N\setminus M) \neq \phi}}} \lambda_N^Z(x_Z) = 0.
\]

Proof. The necessary part is obvious. Now if (3.20) holds, then
\[
(3.21) \quad \sum_{\substack{Z \subseteq L \subseteq N \\mid \\text{Z \cap (N\setminus M) \neq \phi}}} \lambda_N^{L \cup Z}(x_L, x_Z) = 0.
\]
Summing over \(x_m\), where \(m \in L^c \setminus \{k\} \) for \(k \in N\setminus M\), we have
\[
(3.22) \quad \lambda_N^{L \cup \{k\}}(x_L, x_k) = 0 \quad \forall \ k \in N\setminus M.
\]
Substituting (3.22) in (3.21) leads to
\[
(3.23) \quad \sum_{\substack{Z \subseteq L \subseteq N \\mid \\text{Z \cap (N\setminus M) \neq \phi, \ Z \neq \{k\}}} \lambda_N^{L \cup Z}(x_L, x_Z) = 0 \quad \forall \ k \in N\setminus M.
\]
Performing the above steps repeatedly, we obtain
\[
\lambda_N^Z(x_Z) = 0 \quad \forall \ L \subset Z \subseteq N, Z \cap (N\setminus M) \neq \phi,
\]
which completes the proof. \(\square\)

The following result establishes a necessary and sufficient condition for strict collapsibility with respect to a MLL parameter, generalizing Theorem 4.1 of Vellaisamy and Vijay (2007).

Theorem 3.3. Suppose a \(|N|\)-dimensional table \(\mathbf{X}_N\) is collapsible over \(N\setminus M\) into a \(|M|\)-dimensional table \(\mathbf{X}_M\) with respect to \(\lambda_M^L\), where \(L \subseteq M \subset N\). Then it is strictly collapsible if and only if
\[
(3.24) \quad \sum_{Z \subseteq L} (-1)^{|L \setminus Z|} \nu_{Z \cup (N\setminus L)}^N(x_Z, x_{N\setminus L}) = \sum_{Z \subseteq L} (-1)^{|L \setminus Z|} \nu_{Z \cup (M\setminus L)}^N(x_Z, x_{M\setminus L}).
\]

Proof. Suppose (3.24) holds, which is equivalent to
\[
(3.25) \quad \log p_N(x_N) - \nu_{L \cup (M\setminus L)}^N(x_L, x_{M\setminus L}) + \sum_{Z \subseteq L} (-1)^{|L \setminus Z|} \left[ \nu_{Z \cup (N\setminus L)}^N(x_Z, x_{N\setminus L}) - \nu_{Z \cup (M\setminus L)}^N(x_Z, x_{M\setminus L}) \right] = 0.
\]
Note that
\[
(3.26) \quad \log p_N(x_N) - \nu_{L \cup (M\setminus L)}^N(x_L, x_{M\setminus L}) = \sum_{Z \subseteq N} \lambda_N^Z(x_Z) - \sum_{Z \subseteq L \cup (M\setminus L)} \lambda_N^Z(x_Z)
\]
\[
= \sum_{Z \subseteq N \\mid \\text{Z \cap (N\setminus M) \neq \phi}} \lambda_N^Z(x_Z)
\]
\[
= \sum_{L \subseteq Z \subseteq N \\mid \\text{Z \cap (N\setminus M) \neq \phi}} \lambda_N^Z(x_Z) + \sum_{L \subseteq Z \subseteq N \\mid \\text{Z \cap (N\setminus M) \neq \phi}} \lambda_N^Z(x_Z).
\]
Also, observe

\[
\sum_{\substack{L \subseteq Z \\
\text{Z} \cap (N \setminus M) \neq \emptyset}} \lambda^N_Z(x_Z) = \sum_{j=1}^{\vert L \vert} \sum_{\substack{Z \subseteq L_j \cup (N \setminus L) \\
\text{Z} \cap (N \setminus M) \neq \emptyset}} \lambda^N_Z(x_Z) - \sum_{j,l=1}^{\vert L \vert} \sum_{\substack{Z \subseteq L_j \cup (N \setminus L) \\
\text{Z} \cap (N \setminus M) \neq \emptyset}} \lambda^N_Z(x_Z) \\
+ \cdots + (-1)^{\lfloor \vert L \rfloor \rfloor} \sum_{\substack{Z \subseteq L^c \cup (N \setminus L) \\
\text{Z} \cap (N \setminus M) \neq \emptyset}} \lambda^N_Z(x_Z)
\]

\[
= \sum_{j=1}^{\vert L \vert} \left[ \sum_{\substack{Z \subseteq L_j \cup (N \setminus L) \\
\text{Z} \cap (N \setminus M) \neq \emptyset}} \lambda^N_Z(x_Z) - \sum_{\substack{Z \subseteq L_j \cup (N \setminus L) \\
\text{Z} \cap (N \setminus M) \neq \emptyset}} \lambda^N_Z(x_Z) \right] \\
- \sum_{j,l=1}^{\vert L \vert} \left[ \sum_{\substack{Z \subseteq L_j \cup (N \setminus L) \\
\text{Z} \cap (N \setminus M) \neq \emptyset}} \lambda^N_Z(x_Z) - \sum_{\substack{Z \subseteq L_j \cup (N \setminus L) \\
\text{Z} \cap (N \setminus M) \neq \emptyset}} \lambda^N_Z(x_Z) \right] \\
+ \cdots + (-1)^{\lfloor \vert L \rfloor \rfloor} \left[ \sum_{\substack{Z \subseteq (N \setminus L) \\
\text{Z} \cap (N \setminus M) \neq \emptyset}} \lambda^N_Z(x_Z) - \sum_{\substack{Z \subseteq (N \setminus L) \\
\text{Z} \cap (N \setminus M) \neq \emptyset}} \lambda^N_Z(x_Z) \right]
\]

(3.27)

Substituting (3.27) in (3.26), we obtain

\[
\log p_N(x_N) - \nu^N_{L \cup (M \setminus L)}(x_L, x_{M \setminus L}) = \sum_{\substack{L \subseteq Z \\
\text{Z} \cap (N \setminus M) \neq \emptyset}} \lambda^N_Z(x_Z) - \sum_{\substack{Z \subseteq L \\
\text{Z} \cap (N \setminus M) \neq \emptyset}} (-1)^{\lfloor \vert L \rfloor \rfloor} \left[ \nu^N_{Z \cup (N \setminus L)}(x_Z, x_N) - \nu^N_{Z \cup (M \setminus L)}(x_Z, x_M) \right]
\]

(3.28)

Again, substituting (3.28) in (3.26) leads to

\[
\sum_{\substack{L \subseteq Z \\
\text{Z} \cap (N \setminus M) \neq \emptyset}} \lambda^N_Z(x_Z) = 0.
\]

(3.29)

Since (3.24)-(3.29) are all equivalent, the result now follows from Lemma 3.1.

Next, we provide necessary and sufficient conditions for strict collapsibility with respect to a collection of MLL parameters having a set of common effects, which generalizes Theorem 4.2 of Vellaisamy and Vijay (2007).
Theorem 3.4. A \(|N|\)-dimensional table \(\mathbf{x}_N\) is strictly collapsible over \(N\setminus M\) into a \(|M|\)-dimensional table \(\mathbf{x}_M\) with respect to the set \(C_S = \{\lambda^N_A | S \subseteq A \subseteq L\}\) of interaction parameters if and only if
\[
\sum_{R \subseteq S} (-1)^{|S\setminus R|} d_{LR}(x_{LR}) = 0
\]
and
\[
\sum_{Z \subseteq S} (-1)^{|S\setminus Z|} \nu^N_{NS\cup Z}(x_{NS}, x_Z) = \sum_{Z \subseteq S} (-1)^{|S\setminus Z|} \nu^N_{MS\cup Z}(x_{MS}, x_Z),
\]
where \(M_S = M\setminus S\), \(N_S = N\setminus S\) and \(1 \leq |S| \leq |M|\).

Proof. The collapsibility follows from (3.30) by Theorem 3.2. Assume now \(C_A = \{\lambda^N_A | A \subseteq Z, Z \cap (N\setminus M) \neq \emptyset\}\). Note that if \(A_1 \subset A_2\), then \(C_{A_1} \subseteq C_{A_2}\). Hence, for \(A = S\), it is sufficient to show the equivalence of (3.31) and the strict collapsibility condition (ii) of Definition 3.2.

By Theorem 3.3 with \(L = A\), we have
\[
\lambda^N_Z(x_Z) = 0 \quad \forall A \subset Z, \ Z \cap (N\setminus M) \neq \emptyset
\]
if and only if
\[
\sum_{Z \subseteq A} (-1)^{|A\setminus Z|} \nu^N_{Z\cup(M\setminus A)}(x_Z, x_{M\setminus A}) = \sum_{Z \subseteq A} (-1)^{|A\setminus Z|} \nu^N_{Z\cup(N\setminus A)}(x_Z, x_{N\setminus A}).
\]

When \(A = S\), we have \(M\setminus A = M_S\) and \(N\setminus A = N_S\). Therefore using (3.32), we obtain
\[
\lambda^N_Z(x_Z) = 0 \quad \forall S \subset Z, \ Z \cap (N\setminus M) \neq \emptyset
\]
if and only if
\[
\sum_{Z \subseteq S} (-1)^{|S\setminus Z|} \nu^N_{NS\cup Z}(x_{NS}, x_Z) = \sum_{Z \subseteq S} (-1)^{|S\setminus Z|} \nu^N_{MS\cup Z}(x_{MS}, x_Z),
\]
which completes the proof.

4. Collapsibility and Independence

In this section, we study the relationship between strict collapsibility and various forms of independence - conditional, joint and mutual in a multidimensional table. Suppose \(A, B\) and \(C\) form a partition of \(M\) (a finite index set). Then

(i) \(X_A \perp X_B | X_C\) (conditional independence) if and only if
\[
p_M(x_M) = \frac{p_{AC}(x_{AC})p_{BC}(x_{BC})}{p_C(x_C)}
\]
\(\Leftrightarrow \log p_M(x_M) - \log p_{M\setminus A}(x_{M\setminus A}) = \log p_{M\setminus B}(x_{M\setminus B}) - \log p_{M\setminus (AB)}(x_{M\setminus (AB)})\).

(ii) \(X_A \perp (X_B, X_C)\) (joint independence) if and only if
\[
p_M(x_M) = p_A(x_A)p_{B\cup C}(x_{B\cup C})
\]
\(\Leftrightarrow \log p_M(x_M) = \log p_{M\setminus (BC)}(x_{M\setminus (BC)}) + \log p_{M\setminus A}(x_{M\setminus A})\).
(iii) \( X_A \perp X_B \perp X_C \) (mutual independence) if and only if
\[
p_M(x_M) = p_A(x_A)p_B(x_B)p_C(x_C)
\]
\[
\Leftrightarrow \log p_M(x_M) = \log p_{M|BC}(x_M|BC) + \log p_{M|AC}(x_M|AC) + \log p_{M|AB}(x_M|AB).
\]

**Remark 4.1.** For a 3-dimensional table, Theorem 2.4-1 of Bishop, Fienberg and Holland (1975) states that conditional independence is a necessary and sufficient condition for collapsibility of ordinary log-linear parameters, while Theorem 2.5-1 states the same result for a 2-dimensional table in terms of conditional independence.

The following result gives necessary and sufficient conditions for strict collapsibility of a multidimensional table in terms of conditional independence.

**Theorem 4.1.** A \(|M|-dimensional table \( X_M \) is said to be strictly collapsible over \( X_A \) (\( X_B \)) into \( X_{BC} \) (\( X_{AC} \)) with respect to \( \lambda_{ABC} \) and \( \lambda_{ABC'} \) (\( \lambda_{A'B'C} \) and \( \lambda_{A'B'C'} \)) if and only if \( X_A \perp X_B|X_C \), where \( A' \subseteq A \), \( B' \subseteq B \), \( C' \subseteq C \) and \( A, B, C \) form a partition of \( M \).

**Proof.** (a) Sufficiency:
The log-linear marginal model for the \(|M|-dimensional table is
\[
\log p_M(x_M) = \sum_{L \subseteq ABC} \lambda^L_{ABC}(x_L)
\]
\[
= \lambda^A_{ABC} + \sum_{A' \subseteq A} \lambda^A_{ABC'}(x_{A'}) + \sum_{B' \subseteq B} \lambda^B_{ABC}(x_{B'}) + \sum_{C' \subseteq C} \lambda^C_{ABC}(x_{A'})
\]
\[
+ \sum_{A'B' \subseteq AB} \lambda^A_{ABC'}(x_{A'B'}) + \sum_{A'C' \subseteq AC} \lambda^A_{ABC'}(x_{A'C'}) + \sum_{B'C' \subseteq BC} \lambda^A_{ABC'}(x_{B'C'})
\]
\[
(4.1)
\]
\[
+ \sum_{A'B'C' \subseteq ABC} \lambda^A_{ABC'}(x_{A'B'C'}).
\]
Since \( A' \subseteq A \), \( B' \subseteq B \) and \( C' \subseteq C \), using Lemma 2.4, we have
\[
X_A \perp X_B|X_C \Leftrightarrow \lambda^A_{ABC'} = \lambda^A_{A'B'C'} = 0.
\]
From (4.1) and (4.2), the log-linear marginal model under \( X_A \perp X_B|X_C \) is
\[
\log p_M(x_M) = \lambda^A_{ABC} + \sum_{A' \subseteq A} \lambda^A_{ABC'}(x_{A'}) + \sum_{B' \subseteq B} \lambda^B_{ABC}(x_{B'}) + \sum_{C' \subseteq C} \lambda^C_{ABC}(x_{A'})
\]
\[
+ \sum_{A'C' \subseteq AC} \lambda^A_{ABC'}(x_{A'C'}) + \sum_{B'C' \subseteq BC} \lambda^A_{ABC'}(x_{B'C'})
\]
\[
(4.3)
\]
The logarithms of the marginal probabilities in (4.3) are
\[
\log p_{AB}(x_{AB}) = \lambda^A_{ABC} + \sum_{A' \subseteq A} \lambda^A_{ABC'}(x_{A'}) + \sum_{B' \subseteq B} \lambda^B_{ABC}(x_{B'}) + \lambda(x_A, x_B).
\]

18
(4.5) \( \log p_{AC}(x_{AC}) = \lambda^ABC_{\phi} + \sum_{A' \subseteq A} \lambda^ABC_{A'}(x_{A'}) + \sum_{C' \subseteq C} \lambda^ABC_{C'}(x_{C'}) + \sum_{A'C' \subseteq AC} \lambda^ABC_{A'C'}(x_{A'C'}) + \lambda(x_C), \)

(4.6) \( \log p_{BC}(x_{BC}) = \lambda^ABC_{\phi} + \sum_{B' \subseteq B} \lambda^ABC_{B'}(x_{B'}) + \sum_{C' \subseteq C} \lambda^ABC_{C'}(x_{C'}) + \sum_{B'C' \subseteq BC} \lambda^ABC_{B'C'}(x_{B'C'}) + \lambda'(x_C), \)

where

\[
\lambda(x_A, x_B) = \log \left( \sum_{x_C} \exp \left\{ \sum_{C' \subseteq C} \lambda^ABC_{C'}(x_{C'}) + \sum_{A'C' \subseteq AC} \lambda^ABC_{A'C'}(x_{A'C'}) + \sum_{B'C' \subseteq BC} \lambda^ABC_{B'C'}(x_{B'C'}) \right\} \right),
\]

\[
\lambda(x_C) = \log \left( \sum_{x_B} \exp \left\{ \sum_{B' \subseteq B} \lambda^ABC_{B'}(x_{B'}) + \sum_{B'C' \subseteq BC} \lambda^ABC_{B'C'}(x_{B'C'}) \right\} \right),
\]

\[
\lambda'(x_C) = \log \left( \sum_{x_A} \exp \left\{ \sum_{A' \subseteq A} \lambda^ABC_{A'}(x_{A'}) + \sum_{A'C' \subseteq AC} \lambda^ABC_{A'C'}(x_{A'C'}) \right\} \right).
\]

If we collapse the \(|M|\)-dimensional table over \(x_A\), we get

(4.7) \( \log p_{BC}(x_{BC}) = \lambda^ABC_{\phi} + \sum_{B' \subseteq B} \lambda^ABC_{B'}(x_{B'}) + \sum_{C' \subseteq C} \lambda^ABC_{C'}(x_{C'}) + \sum_{B'C' \subseteq BC} \lambda^ABC_{B'C'}(x_{B'C'}). \)

We now compare (4.6) and (4.7). Summing RHS of both over \(x_B\) and \(x_C\) gives

(4.8) \( \lambda^ABC_{\phi} = \lambda^ABC_{\phi} + \sum_{x_C} \lambda'(x_C). \)

Summing RHS of (4.6) and (4.7) over \(x_C\) only, we have

\[
|\mathcal{X}_C| \left( \lambda^ABC_{\phi} + \sum_{B' \subseteq B} \lambda^ABC_{B'}(x_{B'}) \right) = |\mathcal{X}_C| \left( \lambda^ABC_{\phi} + \sum_{B' \subseteq B} \lambda^ABC_{B'}(x_{B'}) \right) + \sum_{x_C} \lambda'(x_C)
\]

(4.9) \( \Rightarrow \sum_{B' \subseteq B} \lambda^ABC_{B'}(x_{B'}) = \sum_{B' \subseteq B} \lambda^ABC_{B'}(x_{B'}) \) (using (4.8)).

Using Lemma 2.2, we have from (4.9)

(4.10) \( \lambda^ABC_{B'}(x_{B'}) = \lambda^ABC_{B'}(x_{B'}) \) \( \forall B' \subseteq B. \)

Now summing RHS of (4.6) and (4.7) over \(x_B\) only gives

\[
|\mathcal{X}_B| \left( \lambda^ABC_{\phi} + \sum_{C' \subseteq C} \lambda^ABC_{C'}(x_{C'}) \right) = |\mathcal{X}_B| \left( \lambda^ABC_{\phi} + \sum_{C' \subseteq C} \lambda^ABC_{C'}(x_{C'}) + \lambda'(x_C) \right)
\]

(4.11) \( \Rightarrow \sum_{C' \subseteq C} \lambda^ABC_{C'}(x_{C'}) = \sum_{C' \subseteq C} \lambda^ABC_{C'}(x_{C'}) + \lambda'(x_C) - \frac{\sum_{x_C} \lambda'(x_C)}{|\mathcal{X}_C|} \) (using (4.8)).

Summing both sides of (4.8), (4.9) and (4.11), we get

(4.12) \( \lambda^ABC_{\phi} + \sum_{B' \subseteq B} \lambda^ABC_{B'}(x_{B'}) + \sum_{C' \subseteq C} \lambda^ABC_{C'}(x_{C'}) = \lambda^ABC_{\phi} + \sum_{B' \subseteq B} \lambda^ABC_{B'}(x_{B'}) + \sum_{C' \subseteq C} \lambda^ABC_{C'}(x_{C'}) + \lambda'(x_C). \)
From (4.6), (4.7) and (4.12), we have

\[ (4.13) \quad \sum_{B'C' \subseteq BC} \lambda^{BC}_{B'C'}(x_{B'C'}) = \sum_{B'C' \subseteq BC} \lambda^{ABC}_{B'C'}(x_{B'C'}). \]

Using Lemma 2.2 it can be shown from (4.13) that

\[ (4.14) \quad \lambda^{BC}_{B'C'}(x_{B'C'}) = \lambda^{ABC}_{B'C'}(x_{B'C'}) \quad \forall \ B'C' \subseteq BC. \]

Analogous results can be obtained by collapsing \( \mathfrak{X}_M \) over \( \mathfrak{X}_B \) and then comparing \( p_{AC}(x_{AC}) \) in \( \mathfrak{X}_M \) and \( \mathfrak{X}_{AC} \). In this case, we get

\[ (4.15) \quad \lambda^{AC}_{A'}(x_{A'}) = \lambda^{ABC}_{A'B'C'}(x_{A'B'C'}) \quad \forall \ A' \subseteq A; \quad \lambda^{AC}_{A'C'}(x_{A'C'}) = \lambda^{ABC}_{A'B'C'}(x_{A'B'C'}) \quad \forall \ A'C' \subseteq AC. \]

Hence, collapsibility over \( \mathfrak{X}_A (\mathfrak{X}_B) \) follows from (4.10) and (4.14) (4.15).

Since \( B' \subseteq A'B' \not\subseteq BC, B'C' \subseteq A'B'C' \not\subseteq BC \) and \( \lambda^{ABC}_{A'B'} = \lambda^{ABC}_{A'B'C'} = 0 \) (see (4.2)), strict collapsibility over \( \mathfrak{X}_A \) with respect to \( \lambda^{ABC}_{B'} \) and \( \lambda^{ABC}_{B'C'} \) follows from Defintion 3.2.

Also, since \( A' \subseteq A'B' \not\subseteq AC, A'C' \subseteq A'B'C' \not\subseteq AC \) and \( \lambda^{ABC}_{A'B'} = \lambda^{ABC}_{A'B'C'} = 0 \) (see (4.2)), strict collapsibility over \( \mathfrak{X}_B \) with respect to \( \lambda^{ABC}_{A'B'} \) and \( \lambda^{ABC}_{A'B'C'} \) follows.

(b) Necessity:

Since \( \mathfrak{X}_M \) is strictly collapsible over \( \mathfrak{X}_A (\mathfrak{X}_B) \) with respect to \( \lambda^{ABC}_{B'} \) and \( \lambda^{ABC}_{B'C'} (\lambda^{ABC}_{A'} \) and \( \lambda^{ABC}_{A'C'} \), we have

\[ 1. \quad \lambda^{BC}_{B'} = \lambda^{ABC}_{B'} \quad \text{and} \quad \lambda^{BC}_{B'C'} = \lambda^{ABC}_{B'C'}; \quad (\lambda^{AC}_{A'} = \lambda^{ABC}_{A'B'C'} \quad \text{and} \quad \lambda^{AC}_{A'C'} = \lambda^{ABC}_{A'B'C'}). \]

\[ 2. \quad \lambda^{ABC}_{A'B'} = \lambda^{ABC}_{A'B'C'} = 0. \]

From Point 2 above and using (4.2), \( X_A \perp \perp X_B | X_C. \)

\[ \square \]

**Remark 4.2.** From Theorem 4.1 we observe that conditional independence is always sufficient for strict collapsibility in a \(|M|\)-dimensional table. However, it is not necessary for the particular case when \(|A| = |B| = |C| = 1 \) (a 3-dimensional table) and \( \mathfrak{X}_M \) is strictly collapsible with respect to two-way interaction parameters only (\( \lambda^{ABC}_{A'C'} \) or \( \lambda^{ABC}_{B'C'} \)). Then \( \lambda^{ABC}_{A'B'C'} = 0 \), while \( \lambda^{ABC}_{A'B'C'} \) is non-zero. This observation is consistent with Theorem 1 of Whittemore (1978), which states the existence of arbitrary 3-dimensional tables that are strictly collapsible over each variable such that no two-way log-linear interaction vanishes.

**Corollary 4.1.** Consider a 3-dimensional table \( \mathfrak{X}_{pqr} \), where \( p \neq q \neq r \in \{1, 2, 3\} \). Then the table is strictly collapsible over \( \mathfrak{X}_p (\mathfrak{X}_q) \) into a 2-dimensional table \( \mathfrak{X}_{qr} (\mathfrak{X}_{pr}) \) with respect to \( \lambda^{123}_{q} \) and \( \lambda^{123}_{p} \) (\( \lambda^{123}_{q} \) and \( \lambda^{123}_{p} \)) if and only if \( X_p \perp \perp X_q | X_r \).

The next result states the relationship between strict collapsibility and joint independence in a multidimensional table.

**Theorem 4.2.** Consider a \(|M|\)-dimensional table \( \mathfrak{X}_{ABC} \) where \( A, B \) and \( C \) form a partition of \( M \). Then the table is strictly collapsible with respect to \( \lambda^{ABC}_{B'}, \lambda^{ABC}_{C'} \) and \( \lambda^{ABC}_{B'C'} \) by collapsing over \( \mathfrak{X}_A \) into \( \mathfrak{X}_{BC} \) if and only if \( X_A \perp \perp (X_B, X_C) \), where \( A' \subseteq A, B' \subseteq B \) and \( C' \subseteq C \).

Proof. (a) Sufficiency:

The log-linear marginal model for \( \mathfrak{X}_M \) is given by (4.11). Using Lemma 2.4 we have

\[ (4.16) \quad X_A \perp \perp (X_B, X_C) \iff \lambda^{ABC}_{AB'C'} = \lambda^{ABC}_{A'B'C'} = \lambda^{ABC}_{A'B'C'} = 0. \]
Hence, from (4.16), the log-linear marginal model under $X_A \perp \perp (X_B, X_C)$ is

\begin{equation}
\log p_M(x_M) = \lambda^{ABC}_\phi + \sum_{A' \subseteq A} \lambda^{ABC}_A(x_A') + \sum_{B' \subseteq B} \lambda^{ABC}_B(x_{B'}) + \sum_{C' \subseteq C} \lambda^{ABC}_C(x_{C'}) + \sum_{B'C' \subseteq BC} \lambda^{ABC}_{B'C'}(x_{B'C'}).\end{equation}

The logarithms of the two-dimensional marginal probabilities in (4.17) are

\begin{equation}
\log p_{AB}(x_{AB}) = \lambda^{ABC}_\phi + \sum_{A' \subseteq A} \lambda^{ABC}_A(x_A') + \sum_{B' \subseteq B} \lambda^{ABC}_B(x_{B'}) + \lambda(x_B),\end{equation}

\begin{equation}
\log p_{AC}(x_{AC}) = \lambda^{ABC}_\phi + \sum_{A' \subseteq A} \lambda^{ABC}_A(x_A') + \sum_{C' \subseteq C} \lambda^{ABC}_C(x_{C'}) + \lambda(x_C),\end{equation}

\begin{equation}
\log p_{BC}(x_{BC}) = \lambda^{ABC}_\phi + \sum_{B' \subseteq B} \lambda^{ABC}_{A'B'}(x_{B'}) + \sum_{C' \subseteq C} \lambda^{ABC}_{C'}(x_{C'}) + \sum_{B'C' \subseteq BC} \lambda^{ABC}_{B'C'}(x_{B'C'}) + \lambda(x_A),\end{equation}

where

\begin{align*}
\lambda(x_B) &= \log \left( \sum_{x_C} \exp \left\{ \sum_{C' \subseteq C} \lambda^{ABC}_{C'}(x_{C'}) + \sum_{B'C' \subseteq BC} \lambda^{ABC}_{B'C'}(x_{B'C'}) \right\} \right), \\
\lambda(x_C) &= \log \left( \sum_{x_B} \exp \left\{ \sum_{B' \subseteq B} \lambda^{ABC}_{B'}(x_{B'}) + \sum_{B'C' \subseteq BC} \lambda^{ABC}_{B'C'}(x_{B'C'}) \right\} \right), \\
\lambda(x_A) &= \log \left( \sum_{x_A} \exp \left\{ \sum_{A' \subseteq A} \lambda^{ABC}_A(x_A') \right\} \right).
\end{align*}

If we collapse $\mathfrak{X}_M$ over $\mathfrak{X}_A$, we get

\begin{equation}
\log p_{BC}(x_{BC}) = \lambda^{BC}_\phi + \sum_{B' \subseteq B} \lambda^{BC}_{B'}(x_{B'}) + \sum_{C' \subseteq C} \lambda^{BC}_{C'}(x_{C'}) + \sum_{B'C' \subseteq BC} \lambda^{BC}_{B'C'}(x_{B'C'}).\end{equation}

We now compare (4.20) and (4.21). Summing RHS of both over $x_B$ and $x_C$ gives

\begin{equation}
\lambda^{BC}_\phi = \lambda^{ABC}_\phi + \lambda(x_A).\end{equation}

Summing RHS of (4.20) and (4.21) over $x_C$ only, we have

\begin{equation}
|\mathfrak{X}_C| \left( \lambda^{BC}_\phi + \sum_{B' \subseteq B} \lambda^{BC}_{B'}(x_{B'}) \right) = |\mathfrak{X}_C| \left( \lambda^{ABC}_\phi + \sum_{B' \subseteq B} \lambda^{ABC}_{B'}(x_{B'}) + \lambda(x_A) \right) \Rightarrow \sum_{B' \subseteq B} \lambda^{BC}_{B'}(x_{B'}) = \sum_{B' \subseteq B} \lambda^{ABC}_{B'}(x_{B'}) \ \text{(using (4.22))}.\end{equation}

Using Lemma 2.2, we have from (4.23)

\begin{equation}
\lambda^{BC}_{B'}(x_{B'}) = \lambda^{ABC}_{B'}(x_{B'}) \ \forall \ B' \subseteq B.
\end{equation}

Now summing RHS of (4.20) and (4.21) over $x_B$ only gives

\begin{equation}
|\mathfrak{X}_B| \left( \lambda^{BC}_\phi + \sum_{C' \subseteq C} \lambda^{BC}_{C'}(x_{C'}) \right) = |\mathfrak{X}_B| \left( \lambda^{ABC}_\phi + \sum_{C' \subseteq C} \lambda^{ABC}_{C'}(x_{C'}) + \lambda(x_A) \right).
\end{equation}
Using Lemma 2.2, it can be shown from (4.25) that

\[ \lambda^{BC}_{C'}(x_{C'}) = \sum_{C' \subseteq C} \lambda^{ABC}_{C'}(x_{C'}) \quad \text{(using (1.22)).} \]

Using Lemma 2.2, it can be shown from (4.25) that

\[ \lambda^{BC}_{C'}(x_{C'}) = \lambda^{ABC}_{C'}(x_{C'}) \quad \forall \ C' \subseteq C. \]

From (4.20)-(4.23) and (4.25), we have

\[ \sum_{B'C' \subseteq BC} \lambda^{BC}_{B'C'}(x_{B'C'}) = \sum_{B'C' \subseteq BC} \lambda^{ABC}_{B'C'}(x_{B'C'}), \]

which implies from Lemma 2.2

\[ \lambda^{BC}_{B'C'}(x_{B'C'}) = \lambda^{ABC}_{B'C'}(x_{B'C'}) \quad \forall \ B'C' \subseteq BC. \]

Hence, collapsibility follows from (4.24), (4.26) and (4.28).

Since \( B' \subset A'B' \not\subset BC \), \( C' \subset A'C' \not\subset BC \) and \( B', C', B'C' \subset A'B'C' \not\subset BC \) with \( \lambda^{ABC} = \lambda^{ABC}_{A'B'C'} = 0 \) (see (4.16)), strict collapsibility follows from Definition 3.2.

(b) Necessity:

Strict collapsibility over \( \mathfrak{X}_A \) with respect to \( \lambda^{ABC}_{B'C'} \), \( \lambda^{ABC}_{C'} \) and \( \lambda^{ABC}_{B'C'} \) implies

1a. \( \lambda^{BC}_{B'} = \lambda^{ABC}_{B'C'} \), \( \lambda^{BC}_{C'} = \lambda^{ABC}_{C'} \) and \( \lambda^{BC}_{B'C'} = \lambda^{ABC}_{B'C'} \),

1b. \( \lambda^{ABC}_{A'B'} = \lambda^{ABC}_{A'C'} = \lambda^{ABC}_{A'B'C'} = 0. \)

From 1b above, we have \( X_A \perp (X_B, X_C) \) using (4.16). Hence, the result follows.

\[ \square \]

**Remark 4.3.** (i) Note that the necessary and sufficient condition \( X_A \perp (X_B, X_C) \) in Theorem 4.2 remains same if we consider strict collapsibility of \( \mathfrak{X}_M \) with respect to \( \lambda^{ABC}_{B'C'} \) by collapsing over \( \mathfrak{X}_B \) (\( \mathfrak{X}_C \)) into \( \mathfrak{X}_{AC} \) (\( \mathfrak{X}_{AB} \)).

(ii) From Theorem 4.2, note that joint independence is always sufficient for strict collapsibility in a \(|M|\)-dimensional table. However, it is not necessary for the specific case when \(|A| = |B| = |C| = 1 \) (a 3-dimensional table) and \( \mathfrak{X}_A \) is strictly collapsible over \( \mathfrak{X}_A \) with respect to \( \lambda^{ABC}_{B'C'} \) only. Then we have \( \lambda^{ABC}_{A'B'C'} = 0 \) while \( \lambda^{ABC}_{A'B'} \) and \( \lambda^{ABC}_{A'C'} \) are non-zero.

**Corollary 4.2.** A 3-dimensional table \( \mathfrak{X}_{pqr} \), where \( p \neq q \neq r \in \{1, 2, 3\} \), is strictly collapsible with respect to \( \lambda^{123}_q \), \( \lambda^{123}_r \) and \( \lambda^{123}_{qr} \) by collapsing over \( \mathfrak{X}_p \) into \( \mathfrak{X}_{pq} \) if and only if \( X_p \perp (X_q, X_r) \). Similarly, \( \mathfrak{X}_{pqr} \) is strictly collapsible with respect to \( \lambda^{123}_p \) by collapsing over \( \mathfrak{X}_q \) (\( \mathfrak{X}_r \)) into \( \mathfrak{X}_{qr} \) (\( \mathfrak{X}_{pq} \)) if and only if \( X_p \perp (X_q, X_r) \).

The following result shows the connection between strict collapsibility and mutual independence in a multidimensional table.

**Theorem 4.3.** Consider a \(|M|\)-dimensional table \( \mathfrak{X}_{ABC} \) where \( A, B \) and \( C \) form a partition of \( M \). Then the table is strictly collapsible with respect to any two of the following sets of MLL parameters:

1. \( \lambda^{ABC}_{B'C'} \) and \( \lambda^{ABC}_{C'} \) by collapsing over \( \mathfrak{X}_A \) into \( \mathfrak{X}_{BC} \),

2. \( \lambda^{ABC}_{A'B'} \) and \( \lambda^{ABC}_{ABC} \) by collapsing over \( \mathfrak{X}_B \) into \( \mathfrak{X}_{AC} \),

3. \( \lambda^{ABC}_{A'C'} \) and \( \lambda^{ABC}_{B'C'} \) by collapsing over \( \mathfrak{X}_C \) into \( \mathfrak{X}_{AB} \).

22
3. \( \lambda_{A'}^{ABC} \) and \( \lambda_B^{ABC} \) by collapsing over \( \mathcal{X}_C \) into \( \mathcal{X}_{AB} \) if and only if \( X_A \perp X_B \perp X_C \), where \( A' \subseteq A \), \( B' \subseteq B \) and \( C' \subseteq C \).

**Proof.** Without loss of generality, we consider strict collapsibility with respect to MLL parameters in Parts 1 and 2 of Theorem 4.3. The proof for Parts 1 and 3 or Parts 2 and 3 follows similarly.

(a) Sufficiency:
The log-linear marginal model for \( \mathcal{X}_M \) is given by (4.1). Using Lemma 2.4, we have

\[
X_A \perp X_B \perp X_C \Leftrightarrow \lambda_{A'B'}^{ABC} = \lambda_{A'C'}^{ABC} = \lambda_{B'C'}^{ABC} = 0.
\]

Hence, from (4.29), the log-linear marginal model under \( X_A \perp X_B \perp X_C \) is

\[
\log p_M(x_M) = \lambda_{\phi}^{ABC} + \sum_{A' \subseteq A} \lambda_{A'}^{ABC}(x_{A'}) + \sum_{B' \subseteq B} \lambda_{B'}^{ABC}(x_{B'}) + \sum_{C' \subseteq C} \lambda_{C'}^{ABC}(x_{C'}).
\]

The logarithms of the two-dimensional marginal probabilities in (4.17) are

\[
\log p_{AB}(x_{AB}) = \lambda_{\phi}^{ABC} + \sum_{A' \subseteq A} \lambda_{A'}^{ABC}(x_{A'}) + \sum_{B' \subseteq B} \lambda_{B'}^{ABC}(x_{B'}) + \lambda(x_C),
\]

\[
\log p_{AC}(x_{AC}) = \lambda_{\phi}^{ABC} + \sum_{A' \subseteq A} \lambda_{A'}^{ABC}(x_{A'}) + \sum_{C' \subseteq C} \lambda_{C'}^{ABC}(x_{C'}) + \lambda(x_B),
\]

\[
\log p_{BC}(x_{BC}) = \lambda_{\phi}^{ABC} + \sum_{B' \subseteq B} \lambda_{B'}^{ABC}(x_{B'}) + \sum_{C' \subseteq C} \lambda_{C'}^{ABC}(x_{C'}) + \lambda(x_A),
\]

where

\[
\lambda(x_C) = \log \left( \sum_{x_C} \exp \left\{ \sum_{C' \subseteq C} \lambda_{C'}^{ABC}(x_{C'}) \right\} \right),
\]

\[
\lambda(x_B) = \log \left( \sum_{x_B} \exp \left\{ \sum_{B' \subseteq B} \lambda_{B'}^{ABC}(x_{B'}) \right\} \right),
\]

\[
\lambda(x_A) = \log \left( \sum_{x_A} \exp \left\{ \sum_{A' \subseteq A} \lambda_{A'}^{ABC}(x_{A'}) \right\} \right).
\]

If we collapse \( \mathcal{X}_M \) over \( \mathcal{X}_A \), we get

\[
\log p_{BC}(x_{BC}) = \lambda_{\phi}^{BC} + \sum_{B' \subseteq B} \lambda_{B'}^{BC}(x_{B'}) + \sum_{C' \subseteq C} \lambda_{C'}^{BC}(x_{C'}).
\]

We now compare (4.33) and (4.34). Summing RHS of both over \( x_B \) and \( x_C \) gives

\[
\lambda_{\phi}^{BC} = \lambda_{\phi}^{ABC} + \lambda(x_A).
\]

Summing RHS of (4.33) and (4.34) over \( x_C \) only, we have

\[
|\mathcal{X}_C| \left( \lambda_{\phi}^{BC} + \sum_{B' \subseteq B} \lambda_{B'}^{BC}(x_{B'}) \right) = |\mathcal{X}_C| \left( \lambda_{\phi}^{ABC} + \sum_{B' \subseteq B} \lambda_{B'}^{ABC}(x_{B'}) + \lambda(x_A) \right)
\]

\[
|\mathcal{X}_C| \left( \lambda_{\phi}^{BC} + \sum_{B' \subseteq B} \lambda_{B'}^{BC}(x_{B'}) \right) = |\mathcal{X}_C| \left( \lambda_{\phi}^{ABC} + \sum_{B' \subseteq B} \lambda_{B'}^{ABC}(x_{B'}) + \lambda(x_A) \right)
\]
Using Lemma 2.2, we have from (4.36)
(4.37) \[ \lambda_{B'}^{BC}(x_{B'}) = \lambda_{B'}^{ABC}(x_{B'}) \quad \forall B' \subseteq B. \]
Now summing RHS of (4.33) and (4.34) over \( x_B \) only gives
\[
|\mathbf{x}_B| \left( \lambda_\phi^{BC} + \sum_{C' \subseteq C} \lambda_{C'}^{BC}(x_{C'}) \right) = |\mathbf{x}_B| \left( \lambda_\phi^{ABC} + \sum_{C' \subseteq C} \lambda_{C'}^{ABC}(x_{C'}) + \lambda(x_A) \right)
\]
(4.38) \[ \Rightarrow \sum_{C' \subseteq C} \lambda_{C'}^{BC}(x_{C'}) = \sum_{C' \subseteq C} \lambda_{C'}^{ABC}(x_{C'}) \quad (\text{using (4.35)}) \]
Using Lemma 2.2, it can be shown from (4.38) that
(4.39) \[ \lambda_{C'}^{BC}(x_{C'}) = \lambda_{C'}^{ABC}(x_{C'}) \quad \forall C' \subseteq C. \]
Similarly, by collapsing \( \mathbf{x}_M \) over \( \mathbf{x}_B \), we get
(4.40) \[ \lambda_{A'}^{AC}(x_{A'}) = \lambda_{A'}^{ABC}(x_{A'}) \quad \text{and} \quad \lambda_{C'}^{BC}(x_{C'}) = \lambda_{C'}^{ABC}(x_{C'}) \quad \forall A', C'. \]
Hence, collapsibility follows from (4.37) and (4.39) for Part 1, and from (4.40) for Part 2.
Since \( B' \subseteq A'B' \not\subseteq BC, C' \subseteq A'C' \not\subseteq BC \) and \( B', C' \subseteq A'B'C' \not\subseteq BC \) with \( \lambda_{A'B'}^{ABC} = \lambda_{A'C'}^{ABC} = \lambda_{A'B'C'}^{ABC} = 0 \) (see (4.29)), strict collapsibility follows for Part 1 from Definition 3.2.
For Part 2, note that \( A' \subseteq A'B' \not\subseteq AC, C' \subseteq B'C' \not\subseteq AC \) and \( A', C' \subseteq A'B'C' \not\subseteq BC \) with \( \lambda_{A'B'}^{ABC} = \lambda_{A'C'}^{ABC} = \lambda_{A'B'C'}^{ABC} = 0 \) (see (4.29)) implying strict collapsibility.

(b) Necessity:
For Part 1, strict collapsibility over \( \mathbf{x}_A \) with respect to \( \lambda_{A'}^{ABC} \) and \( \lambda_{C'}^{ABC} \) implies
1a. \( \lambda_{B'}^{BC} = \lambda_{B'}^{ABC} \) and \( \lambda_{C'}^{BC} = \lambda_{C'}^{ABC} \),
1b. \( \lambda_{A'B'}^{ABC} = \lambda_{A'C'}^{ABC} = \lambda_{A'B'C'}^{ABC} = 0. \)
For Part 2, strict collapsibility over \( \mathbf{x}_B \) with respect to \( \lambda_{A'}^{ABC} \) and \( \lambda_{C'}^{ABC} \) implies
2a. \( \lambda_{A'}^{AC} = \lambda_{A'}^{ABC} \) and \( \lambda_{C'}^{AC} = \lambda_{C'}^{ABC} \),
2b. \( \lambda_{A'B'}^{ABC} = \lambda_{A'B'C'}^{ABC} = \lambda_{A'B'C'}^{ABC} = 0. \)
From 1b and 2b above, we have \( \lambda_{A'B'}^{ABC} = \lambda_{A'C'}^{ABC} = \lambda_{B'C'}^{ABC} = \lambda_{A'B'C'}^{ABC} = 0 \iff X_A \perp \perp X_B \perp \perp X_C \)
using (4.29). Hence, the result follows.

Remark 4.4. From Theorem 4.3 we observe that mutual independence is always necessary and sufficient for strict collapsibility in a \(|M|\)-dimensional table.

Corollary 4.3. A 3-dimensional table \( \mathbf{x}_{pqr} \), where \( p \neq q \neq r \in \{1, 2, 3\} \), is strictly collapsible with respect to any two of the following sets of MLL parameters:
1. \( \lambda_{123}^{pqr} \) and \( \lambda_{123}^{pqr} \) by collapsing over \( \mathbf{x}_p \) into \( \mathbf{x}_{qr} \),
2. \( \lambda_{123}^{pqr} \) and \( \lambda_{123}^{pqr} \) by collapsing over \( \mathbf{x}_q \) into \( \mathbf{x}_{pr} \),
3. $\lambda_{p}^{123}$ and $\lambda_{q}^{123}$ by collapsing over $X_r$ into $X_{pq}$ if and only if $X_p \perp X_q \perp X_r$.

5. Smoothness of Log-linear Marginal Models under Collapsibility

In this section, we explore the relationship between smooth parameterization and collapsibility. To this effect, we first prove a result on MLL parameters defined within different margins. Then using this result, a sufficient condition is provided to show the existence of a smooth MLL parameterization under collapsibility conditions.

The MLL parameters $\{\lambda_{L}^{M}|L \subseteq M\}$ parameterize a marginal distribution $p_M$. Similarly, the conditional distribution $X_A|X_B$ for disjoint $A$ and $B$ can be smoothly parameterized (see Evans (2015)) by

$$\lambda_{A|B} \equiv (\lambda_{L}^{M}|L \subseteq AB, L \cap A \neq \phi).$$

That is, $\lambda_{A|B}$ is the collection of all MLL parameters defined within the margin $AB$, whose effects contain some element of $A$. For $L \subseteq M \subseteq N$, Theorem 3.1 of Evans (2015) provides the exact relationship between $\lambda_{L}^{M}$ and $\lambda_{N}^{M}$ for the case of binary variables. We extend their result to the case of general categorical variables, each with arbitrary number of levels as shown below.

**Theorem 5.1.** Let $A$ and $B$ be disjoint subsets of $V$ with $|X_v| \geq 2$ for $v \in V$. Then the MLL parameter $\lambda_{L}^{AB}$ may be decomposed as

$$\lambda_{L}^{AB} = \lambda_{L}^{B} + f(\lambda_{A|B})$$

for a smooth function $f$ which vanishes if $X_A \perp X_v|X_B\setminus\{v\}$ for some $v \in L$.

**Proof.** Note that

$$\lambda_{L}^{AB}(x_L) = \sum_{L' \subseteq L} (-1)^{|L\setminus L'|} \lambda_{L'}^{AB}(x_{L'})$$

$$= \sum_{L' \subseteq L} (-1)^{|L\setminus L'|} \frac{1}{|X_{AB\setminus L'}|} \sum_{j_{AB} \in x_{AB}} \log p_{AB}(j_{AB})$$

$$= \sum_{L' \subseteq L} (-1)^{|L\setminus L'|} \frac{1}{|X_{A\setminus L'}| |X_{B\setminus L'}|} \sum_{j_{A} \in x_{A}} \sum_{j_{B} \in x_{B}} \log p_{B}(j_{B}) + \log p_{A|B}(j_{A}|j_{B})$$

$$+ \sum_{L' \subseteq L} (-1)^{|L\setminus L'|} \frac{1}{|X_{A\setminus L'}| |X_{B\setminus L'}|} \sum_{j_{AB} \in x_{AB}} \log p_{AB}(j_{AB})$$

$$= \sum_{L' \subseteq L} (-1)^{|L\setminus L'|} \frac{1}{|X_{A\setminus L'}| |X_{B\setminus L'}|} \times |X_{A\setminus L'}| \sum_{j_{B} \in x_{B}} \log p_{B}(j_{B})$$

25
\[+ \sum_{L' \subseteq L} (-1)^{|L\setminus L'|} \frac{1}{|X_{AB\setminus L'}|} \sum_{j_{AB} \in \mathcal{X}_{AB} \atop j_{L'} = x_{L'}} \log p_{A|B}(j_A|j_B)\]

(5.3) \[= \lambda^B_L(x_L) + f(\lambda_{A|B}) \text{ (say).}\]

Since the second term on the RHS of (5.3) is a smooth function of the conditional probabilities \(p_{A|B}(j_A|j_B)\), it implies that \(f\) is also a smooth function of the MLL parameters \(\lambda_{A|B}\) defined in (5.1). Suppose now \(X_A \perp \! \! \! \perp X_v | X_{B\setminus\{v\}}\) for some \(v \in L\). Then

\[f(\lambda_{A|B}) = \sum_{L' \subseteq L} (-1)^{|L\setminus L'|} \frac{1}{|X_{AB\setminus L'}|} \sum_{j_{AB} \in \mathcal{X}_{AB} \atop j_{L'} = x_{L'}} \log p_{A|B}(j_A|j_B)\]

\[= \sum_{L' \subseteq L} (-1)^{|L\setminus L'|} \frac{1}{|X_{AB\setminus L'}|} \sum_{j_{AB} \in \mathcal{X}_{AB} \atop j_{L'} = x_{L'}} \log p_{A|B\setminus\{v\}}(j_A|j_B\setminus\{v\})\]

\[= \frac{1}{|X_{AB\setminus L'}|} \sum_{L' \subseteq L} (-1)^{|L\setminus L'|} \frac{1}{|X_{L\setminus L'}|} \sum_{j_{AB} \in \mathcal{X}_{AB} \atop j_{L'} = x_{L'}} \log p_{A|B\setminus\{v\}}(j_A|j_B\setminus\{v\})\]

\[= \frac{1}{|X_{AB\setminus L'}|} \sum_{L' \subseteq L} (-1)^{|L\setminus L'|} \frac{1}{|X_{L\setminus L'}|} \sum_{j_{AB} \in \mathcal{X}_{AB} \atop j_{L'} = x_{L'}} \log p_{A|B\setminus\{v\}}(j_A|j_B\setminus\{v\})\]

\[+ \frac{1}{|X_{AB\setminus L'}|} \sum_{L' \subseteq L} (-1)^{|L\setminus L'|} \frac{1}{|X_{L\setminus L'}|} \sum_{j_{AB} \in \mathcal{X}_{AB} \atop j_{L'} = x_{L'}} \log p_{A|B\setminus\{v\}}(j_A|j_B\setminus\{v\})\]

(5.4) \[= D_1 + D_2 \text{ (say).}\]

Now

\[D_2 = \frac{1}{|X_{AB\setminus L'}|} \sum_{L' \subseteq L \atop v \in L'} (-1)^{|L\setminus L'|} \frac{1}{|X_{L\setminus L'}|} \sum_{j_{AB} \in \mathcal{X}_{AB} \atop j_{L'} = x_{L'}} \log p_{A|B\setminus\{v\}}(j_A|j_B\setminus\{v\})\]

\[= \frac{1}{|X_{AB\setminus L'}|} \sum_{L' \subseteq L \atop v \in L'} (-1)^{|L\setminus L'|} \frac{1}{|X_{L\setminus L'}|} \sum_{j_{AB} \in \mathcal{X}_{AB} \atop j_{L'} = x_{L'}} \log p_{A|B\setminus\{v\}}(j_A|j_B\setminus\{v\})\]

\[= \frac{1}{|X_{AB\setminus L'}|} \sum_{L' \subseteq L \atop v \in L'} (-1)^{|L\setminus L'|} \frac{1}{|X_{L\setminus L'}|} \sum_{j_{AB} \in \mathcal{X}_{AB} \atop j_{L'} = x_{L'}} \log p_{A|B\setminus\{v\}}(j_A|j_B\setminus\{v\})\]

\[= \frac{1}{|X_{AB\setminus L'}|} \sum_{L' \subseteq L \atop v \in L'} (-1)^{|L\setminus L'|} \frac{1}{|X_{L\setminus L'}|} \sum_{j_{AB} \in \mathcal{X}_{AB} \atop j_{L'} = x_{L'}} \log p_{A|B\setminus\{v\}}(j_A|j_B\setminus\{v\})\]

(5.5) \[= -D_1.\]
From (5.4) and (5.3), we get \( f(\lambda_{A|B}) = 0 \). This completes the proof. \( \square \)

Theorem 3 of Bergsma and Rudas shows that for \( L \subseteq M \subseteq N \), the MLL parameters \( \lambda^M_L \) and \( \lambda^N_L \) are linearly dependent at certain points in the parameter space. Hence, no smooth parameterization can include two such parameters. As a result, collapsibility conditions (see (3.2)) generally do not define a curved exponential family. However, we provide a sufficient condition using Theorem 5.1 for which a complete and hierarchical \( \mathcal{P} \) and hence a smooth MLL parameterization or a curved exponential family can be obtained under collapsibility conditions.

**Theorem 5.2.** Let \( \{A, B\} \) be a partition of \( M \). Then there exists a smooth and possibly variation independent MLL parameterization of \( \mathcal{F} \) on \( \mathfrak{X}_M \) under collapsibility conditions with respect to \( L \subseteq M \) if \( X_A \perp \perp X_v | X_{B\{v\}} \) for some \( v \in L \).

**Proof.** By Theorem 5.1 we have

\[
\lambda^M_L = \lambda^B_L + f(\lambda_{A|B})
\]

for a smooth function \( f \). Also, since \( X_A \perp \perp X_v | X_{B\{v\}} \) for some \( v \in L \), we have \( f(\lambda_{A|B}) = 0 \). So

\[
\lambda^M_L(x_L) = \lambda^B_L(x_L) \quad \forall x_L \in \mathfrak{X}_L,
\]

which implies that \( \mathfrak{X}_M \) is collapsible into \( \mathfrak{X}_B \) with respect to \( \lambda^M_B \) for \( L \subseteq M \). Now consider two complete MLL parameterizations of \( \mathcal{F} \) on \( \mathfrak{X}_M \) corresponding to the collections \( \mathcal{S} \) and \( \mathcal{T} \) (say). Let \( \mathcal{S} \) be non-hierarchical, while \( \mathcal{T} \) is hierarchical. Since \( \mathcal{T} \) is both hierarchical and complete, \( \lambda_{\mathcal{T}} \) is a smooth MLL parameterization of \( \mathcal{F} \) on \( \mathfrak{X}_M \) by Theorem 2 of Bergsma and Rudas (2002). The MLL parameters defined by (5.7) are non-smooth by Theorem 3 of Bergsma and Rudas (2002), and are hence embedded not in \( \mathcal{T} \) but in \( \mathcal{S} \). Specifically, the effect \( L \) is defined in \( \mathcal{S} \) not within the first but some subsequent margin of which it is a subset. However, this is not the case with respect to \( L \) in \( \mathcal{T} \). Let \( B \) be the first margin of which \( L \) is a subset in \( \mathcal{S} \) and \( \mathcal{T} \). Then from (5.7), \( L \) is defined within \( M \) instead of \( B \) in \( \mathcal{S} \), while it has to be defined within \( B \) in \( \mathcal{T} \).

In general, if the conditional distribution \( X_A | X_B \) is fixed, that is, \( p_{A|B} \) or \( f(\lambda_{A|B}) \) is known, then the relationship between \( \lambda^B_L \) and \( \lambda^M_L \) is linear from (5.6). Indeed, \( \lambda^B_L \) and \( \lambda^M_L \) become interchangeable as part of a parameterization, preserving smoothness and (when relevant) variation independence. From (5.7), \( f \) is known since \( f = 0 \). This implies \( \lambda^B_L \) and \( \lambda^M_L \) are interchangeable, that is, \( \lambda_{\mathcal{S}} \) is smooth if and only if \( \lambda_{\mathcal{T}} \) is also smooth, which is true. Thus \( \lambda_{\mathcal{S}} \) provides a smooth parameterization of \( \mathcal{F} \) on \( \mathfrak{X}_M \) under collapsibility conditions thereby defining a curved exponential family. In addition, if \( \mathcal{T} \) is ordered decomposable, then by Theorem 4 of Bergsma and Rudas (2002), \( \lambda_{\mathcal{T}} \) and hence \( \lambda_{\mathcal{S}} \) is a variation independent parameterization of \( \mathcal{F} \) on \( \mathfrak{X}_M \). \( \square \)

Using Theorem 5.2 we can provide sufficient conditions for collapsibility of a multidimensional table with respect to MLL parameters as shown below.

**Theorem 5.3.** Let \( A, B \) and \( C \) form a partition of \( M \). Then for \( R \in \{A, B, C\} \), the table \( \mathfrak{X}_M \) is collapsible over \( \mathfrak{X}_R \) into \( \mathfrak{X}_{M\setminus R} \) with respect to \{\( \lambda^M_L | L \subseteq M\setminus R \)\} if \( X_R \perp \perp X_v | X_{(M\setminus R)\{v\}} \) for some \( v \in M\setminus R \).
Proof. In Theorem 5.1, take $A = R$, $B = M \setminus R$ and $L \subseteq B$ (see (5.2)). Also, note that $f(\lambda_{A[B]} = 0$ if $X_R \perp X_v | X_{(M \setminus R) \setminus \{v\}}$ for some $v \in M \setminus R$ so that from (5.2), we have $\lambda^M_L(x_L) = \lambda^M_{L \setminus R}(x_L)$ for all $x_L \in \mathcal{X}_L$. Hence, the result follows from Definition 3.1.

6. Conclusions

In this paper, our main aim has been to investigate collapsibility for categorical data in a multidimensional contingency table. For this purpose, we consider a large class of models called marginal models introduced by Bergsma and Rudas (2002) for studying strictly positive discrete distributions on such tables. The MLL parameters include the ordinary log-linear and multivariate logistic parameters as special cases. Moreover, the marginal models also generalize several other models studied in the literature. For the multidimensional table, it is assumed that each categorical variable has an arbitrary number of levels.

First, we obtain some distinctive properties of MLL parameters using simple expressions for such parameters. Then collapsibility and strict collapsibility of these parameters are defined in a general sense by considering two arbitrary margins of a table. We derive several necessary and sufficient conditions for collapsibility and strict collapsibility using the technique of Möbius inversion. These conditions are easily verifiable from a table. We also provide various results on collapsibility and strict collapsibility with respect to an arbitrary set of MLL parameters containing some common effects. Such results are useful for studying associations among various categorical variables in a table. Further, we explore the relationships of collapsibility and strict collapsibility with various forms of independence of the variables. We establish necessary and sufficient conditions for each case. Finally, we provide a result on the connection between parameters having a common effect but defined within different margins. This result is then used to demonstrate the existence of a smooth MLL parameterization under collapsibility conditions, thereby obtaining a curved exponential family. Moreover, some sufficient conditions for collapsibility in a multidimensional table are also provided using the result.

References

[1] Agresti, A., Lang, J. B., 1993. A proportional odds model with subject-specific effects for repeated ordered categorical responses. Biometrika 80, 527-534.
[2] Becker, M. P., 1994. Analysis of repeated categorical measurements using models for marginal distributions: an application to trends in attitudes on legalized abortion. Sociological Methodology. Blackwell, Oxford.
[3] Bergsma, W. P., Rudas, T., 2002. Marginal models for categorical data. Ann. Statist. 30(1), 140-159.
[4] Bergsma, W. P., Croon, M. A., Hagenaars, J. A., 2009. Marginal Models: For Dependent, Clustered, and Longitudinal Categorical Data. Springer Science & Business Media, 2009.
[5] Bishop, Y. M. M., Fienberg, S. E., Holland, P. W., 1975. Discrete multivariate analysis: theory and practice. Cambridge: MIT Press.
[6] Charalambides, A. C., 2002. Enumerative combinatorics. Florida: Chapman and Hall.
[7] Cox, D. R., Wermuth, N. (2003). A general condition for avoiding effect reversal after marginalization. J. R. Stat. Soc. Ser. B 65, 937-941.
[8] Croon, M. A., Bergsma, W. P., Hagenaars, J. A., 2000. Analyzing change in categorical variables by generalized log-linear interaction models. Sociol. Methods Res. 29, 195-229.
[9] Ducharme, G. R., Lepage, Y., 1986. Testing collapsibility in multidimensional tables. J. R. Stat. Soc. Ser. B 48, 197-205.
[10] Evans, R. J., 2011. Parametrizations of Discrete Graphical Models (Thesis). University of Washington, http://www.stats.ox.ac.uk/~evans/Thesis.pdf
[11] Evans, R. J., Richardson, T. J., 2013. Marginal log-linear parameters for graphical Markov models. J. R. Stat. Soc. Ser. B 75, 743-768.
[12] Evans, R. J., 2015. Smoothness of marginal log-linear parameterizations. Electron. J. Stat. 9, 475-491.
[13] Forcina, A., Lupparelli, M., Marchetti, G. M., 2010. Marginal parameterizations of discrete models defined by a set of conditional independencies. J. Multivariate Anal. 101, 2519-2527.
[14] Glonek, G. F. V., McCullagh, P., 1995. Multivariate logistic models. J. R. Stat. Soc. Ser. B 57, 533-546.
[15] Glonek, G. F. V., 1996. A class of regression models for multivariate categorical responses. Biometrika 83, 15-28.
[16] Guo, J. H., Geng, Z., 1995. Collapsibility of logistic regression coefficients. J. R. Stat. Soc. Ser. B 57, 263-267.
[17] Haberman, S. J., 1974. The Analysis of Frequency Data. Univ. Chicago Press.
[18] Hammersley, J. M., Clifford, P. E., 1971. Markov fields on finite graphs and lattices. Unpublished manuscript.
[19] Kauermann, G., 1997. A note on multivariate logistic models for contingency tables. Austral. J. Statist. 39, 261-276.
[20] Lang, J. B., Agresti, A., 1994. Simultaneously modelling joint and marginal distributions of multivariate categorical responses. J. Amer. Statist. Assoc. 89, 625-632.
[21] Lauritzen, S. L., 1996. Graphical Models. Clarendon Press, Oxford, UK.
[22] Liang, K. Y., Zeger, S. L., Qaqish, B., 1992. Multivariate regression analyses for categorical data (with discussion). J. R. Stat. Soc. Ser. B 54, 3-40.
[23] Lindley, D. V., Novick, M. R., 1981. The role of exchangeability in inference. Ann. Statist. 9, 45-58.
[24] McCullagh, P., Nelder, J. A., 1989. Generalized Linear Models. 2nd ed. Chapman and Hall, London.
[25] Roverato, A., Lupparelli, M., 2013. Log-mean linear models for binary data. Biometrika. 100, 485-494.
[26] Rudas, T., Leimer, H. G., 1992. Analysis of contingency tables with known conditional odds ratios or known log-linear parameters. Stat. Model. Elsevier, New York, 312-322.
[27] Rudas, T., Bergsma, W. P., Nemeth, R., 2010. Marginal log-linear parameterization of conditional independence models. Biometrika 94, 1006-1012.
[28] Shapiro, S. H., 1982. Collapsing contingency tables: a geometrical approach. Amer. Statist. 36, 43-46.
[29] Simpson, E. H., 1951. The interpretation of interaction in contingency tables. J. R. Stat. Soc. Ser. B 13, 238-241.
[30] Vellaisamy, P., 2012. Simpson’s paradox and collapsibility. J. Indian Statist. Assoc., 50, 297-317.
[31] Vellaisamy, P., Vijay, V., 2007. Some collapsibility results for n-dimensional contingency tables. Ann. Inst. Statist. Math. 59, 577-576.
[32] Vellaisamy, P., Vijay, V., 2009. Log-linear modelling using conditional log-linear structures. Ann. Inst. Statist. Math. 61, 309-329.
[33] Vellaisamy, P., Vijay, V., 2010. Collapsibility of contingency tables based on conditional models. J. Statist. Plann. Inference 140, 1243-1255.
[34] Wermuth, N., 1987. Parametric collapsibility and the lack of moderating effects in contingency tables with a dichotomous response variable. J. R. Stat. Soc. Ser. B 49, 353-364.
[35] Whittemore, A. S., 1978. Collapsibility of multidimensional contingency tables. J. R. Stat. Soc. Ser. B 40, 328-340.

Sayan Ghosh, Department of Mathematics, Indian Institute of Technology Bombay, Powai, Mumbai 400076, INDIA.
E-mail address: sayang@math.iitb.ac.in

P. Vellaisamy, Department of Mathematics, Indian Institute of Technology Bombay, Powai, Mumbai 400076, INDIA.
E-mail address: pv@math.iitb.ac.in