Hierarchical subspace models for contingency tables

Hisayuki Hara, Tomonari Sei and Akimichi Takemura

University of Tokyo

August 16, 2010
Universidade de São Paulo
Introduction
Notations for \( m \)-way contingency tables

- \( I := I_1 \times \cdots \times I_m \) : number of cells of an \( m \)-way table
  - \( I_k \) : number of levels for \( k \)-th variable
- \( \mathcal{I} := [I_1] \times \cdots \times [I_m] \) : set of cells
  - \( [I_k] = \{1, \ldots, I_k\} \)
- \( i = (i_1 i_2 \cdots i_m) \in \mathcal{I} \) : each cell
- For \( D \in [m] \),
  - \( i_D \) : marginal cell
  - \( \mathcal{I}_D \) : the set of marginal cells for \( D \)
  - \( I_D \) : the number of marginal cells for \( D \)

- \( x(\cdot) \) : frequencies
- \( p(\cdot) \) : cell probabilities
Space of $m$-way tables

- $V = \mathbb{R}^I = \mathbb{R}^{I_1 \times \cdots \times I_m}$: the set of $m$-way tables with real entries
  
  $V$: $I$-dimensional real vector space of functions $\psi: \mathcal{I} \mapsto \mathbb{R}$

- $L_D$ for $D \subseteq [m]$:
  
  the set of functions depending only on $D$ marginal cells

  $$L_D = \{ \psi \in V \mid \psi(i_1, \ldots, i_m) = \psi(i'_1, \ldots, i'_m) \text{ if } i_v = i'_v, \forall v \in D \}$$

- $L_D$ is considered as $\mathbb{R}^{I_D}$, where $I_D = \prod_{v \in D} I_v$

- If $D = [m]$, $L_D = D_{[m]} = V$
Space of $m$-way tables

- $V = \mathbb{R}^I = \mathbb{R}^{I_1 \times \cdots \times I_m}$: the set of $m$-way tables with real entries
  - $V$: $I$-dimensional real vector space of functions $\psi: \mathcal{I} \rightarrow \mathbb{R}$

- $L_D$ for $D \subset [m]$:
  - the set of functions depending only on $D$ marginal cells
  
  \[ L_D = \{ \psi \in V \mid \psi(i_1, \ldots, i_m) = \psi(i'_1, \ldots, i'_m) \text{ if } i_v = i'_v, \forall v \in D \} \]

- $L_D$ is considered as $\mathbb{R}^{I_D}$, where $I_D = \prod_{v \in D} I_v$
- If $D = [m]$, $L_D = D_{[m]} = V$
Space of $m$-way tables

- $V = \mathbb{R}^I = \mathbb{R}^{I_1 \times \cdots \times I_m}$: the set of $m$-way tables with real entries
  
  $V$: $I$-dimensional real vector space of functions $\psi: \mathcal{I} \mapsto \mathbb{R}$

- $L_D$ for $D \subset [m]$:
  the set of functions depending only on $D$ marginal cells

  $$L_D = \{ \psi \in V \mid \psi(i_1, \ldots, i_m) = \psi(i'_1, \ldots, i'_m) \text{ if } i_v = i'_v, \forall v \in D \}$$

- $L_D$ is considered as $\mathbb{R}^{I_D}$, where $I_D = \prod_{v \in D} I_v$

- If $D = [m]$, $L_D = D_{[m]} = V$
Space of $m$-way tables

- $V = \mathbb{R}^I = \mathbb{R}^{I_1 \times \cdots \times I_m}$: the set of $m$-way tables with real entries
  - $V$: $I$-dimensional real vector space of functions $\psi: \mathcal{I} \mapsto \mathbb{R}$

- $L_D$ for $D \subset [m]$:
  - the set of functions depending only on $D$ marginal cells
  \[ L_D = \{ \psi \in V \mid \psi(i_1, \ldots, i_m) = \psi(i'_1, \ldots, i'_m) \text{ if } i_v = i'_v, \forall v \in D \} \]

- $L_D$ is considered as $\mathbb{R}^{I_D}$, where $I_D = \prod_{v \in D} I_v$
- If $D = [m]$, $L_D = D_{[m]} = V$
$L$: a linear subspace of $V$ s.t. $1 \in L$

Log-affine model:

\[
\log p(\cdot) := \{\log p(i), i \in I\} \in L
\]

$V \leftrightarrow m$-way saturated (full) model
Log affine model

- $L$ : a linear subspace of $V$ s.t. $1 \in L$
- **Log-affine model** :

$$\log p(\cdot) := \{\log p(i), i \in I\} \in L$$

- $V \iff m$-way saturated (full) model
\( L \) : a linear subspace of \( V \) s.t. \( 1 \in L \)

**Log-affine model**:

\[
\log p(\cdot) := \{\log p(i), i \in \mathcal{I}\} \in L
\]

\( V \Leftrightarrow m \)-way saturated (full) model
Hierarchical model

- $\Delta$: a simplicial complex
- $\text{red}\Delta$: the set of maximal elements (facets) of $\Delta$
  - $\text{red}\Delta$ is considered as a hypergraph
- **Hierarchical model $L_\Delta$:**

\[
\log p(\cdot) \in L_\Delta := \sum_{D \in \text{red}\Delta} L_D,
\]

- **Graphical model**
  - $\Leftrightarrow \text{red}\Delta$ is the set of maximal cliques of a graph
Example 1. modeling for $I \times J$ tables

- The saturated model: $\text{red}\Delta := \{1, 2\}$

$$V = L_{\{1,2\}} : \log p_{ij} = \alpha_i + \beta_j + \gamma_{ij}$$

- A log-affine model $L$:

$$L : \log p_{ij} = \alpha_i + \beta_j + \gamma \phi_{ij} \subset V$$

- $\phi_{ij}$: known functions
  - ex 1. uniform association model: $\phi_{ij} = ij$
  - ex 2. two-way change point model (Hirotsu(1997)):

$$\phi_{ij} = \begin{cases} 
1, & \text{if } i \leq I_1 < I \text{ and } j \leq J_1 < J, \\
0, & \text{otherwise,}
\end{cases}$$

- Modeling strategy for higher dimensional tables has not been fully discussed
Example 1. modeling for $I \times J$ tables

- The saturated model: $\text{red} \Delta := \{1, 2\}$

$$V = L_{\{1,2\}} : \log p_{ij} = \alpha_i + \beta_j + \gamma_{ij}$$

- A log-affine model $L$:

$$L : \log p_{ij} = \alpha_i + \beta_j + \gamma \phi_{ij} \subset V$$

  - $\phi_{ij}$: known functions

ex 1. uniform association model: $\phi_{ij} = ij$

ex 2. two-way change point model (Hirotsu(1997)):

$$\phi_{ij} = \begin{cases} 1, & \text{if } i \leq I_1 < I \text{ and } j \leq J_1 < J, \\ 0, & \text{otherwise}, \end{cases}$$

↓

- Modeling strategy for higher dimensional tables has not been fully discussed
Example 1. modeling for $I \times J$ tables

- The saturated model: $\text{red} \Delta := \{1, 2\}$

$$V = L_{\{1,2\}} : \log p_{ij} = \alpha_i + \beta_j + \gamma_{ij}$$

- A log-affine model $L$:

$$L : \log p_{ij} = \alpha_i + \beta_j + \gamma \phi_{ij} \subset V$$

  - $\phi_{ij}$: known functions
  - ex. 1. uniform association model: $\phi_{ij} = ij$
  - ex. 2. two-way change point model (Hirotsu(1997)):

$$\phi_{ij} = \begin{cases} 
1, & \text{if } i \leq I_1 < I \text{ and } j \leq J_1 < J, \\
0, & \text{otherwise,}
\end{cases}$$

$\downarrow$

- Modeling strategy for higher dimensional tables has not been fully discussed
Example 2. 3-way Split model

Split model (Højsgaard (2003)):

The conditional independence structures are different for specific values of the conditioning variables

\[ L = L_{i_2=1}^1 + L_{i_2=1}^3 + L_{i_2=2}^{1,3} \]

- \( i_2 = 1 \) slice \( L_{i_2=1}^1 = L_{1} + L_{3} \)
- \( i_2 = 2 \) slice \( L_{i_2=2}^{1,3} = L_{1,3} \)
- \( L \subset L_{1,2,3} \)

Sophisticated modeling of interaction terms is required for the analysis of contingency tables
Example 2. 3-way Split model

- Split model (Højsgaard (2003)):
  The conditional independence structures are different for specific values of the conditioning variables

  \[ L = L_{i_2=1}^{1} + L_{i_2=1}^{2} + L_{i_2=2}^{2} \]

- \( i_2 = 1 \) slice \( L_{i_2=1}^{1} = L_{\{1\}} + L_{\{3\}} \)
- \( i_2 = 2 \) slice \( L_{i_2=2}^{2} = L_{\{1,3\}} \)
- \( L \subset L_{\{1,2,3\}} \)

\[ \Downarrow \]

- Sophisticated modeling of interaction terms is required for the analysis of contingency tables
We propose "hierarchical subspace model (HSM)" as a generalization of the hierarchical model.

The notion of an HSM gives a modeling strategy of multiway tables and unifies various models of interaction effects.

In this talk we discuss HSM from a viewpoint of localization of the computation of MLE and Markov bases.

We also illustrate practical advantage of our modeling strategy with some data sets.
Hierarchical subspace model
Conformality of Log-affine model

**Definition (conformality)**

\[ W_1, \ldots, W_K : \text{linear subspaces of } V \]

\[ W := W_1 + \cdots + W_K \]

\( L \) is conformal to \( \{ W_j \}_{j=1}^K \) if

\[ L = L \cap W = (L \cap W_1) + \cdots + (L \cap W_K) \]

- \( L \supset (L \cap W_1) + \cdots + (L \cap W_K) \) always holds
- The inclusion is strict in general
Ex. 3-way conditional independence model

- $W_1 := L_{\{1,2\}}$, $W_2 := L_{\{2,3\}}$
- $W = W_1 + W_2 = L_{\{1,2\}} + L_{\{2,3\}}$
- $W : \log p_{i_1i_2i_3} = \alpha_{i_1} + \beta_{i_2} + \gamma_{i_3} + \delta_{i_1i_2} + \delta'_{i_2i_3}$

\[1\] \[2\] \[3\]

- $L : \log p_{i_1i_2i_3} = \alpha_{i_1} + \beta_{i_2} + \gamma_{i_3} + \delta \phi_{i_1i_2} + \delta' \psi_{i_2i_3}$
  - $\phi_{i_1i_2}, \psi_{i_2i_3} :$ known functions
  - $\delta, \delta'$: free parameters
  - $L \cap W_1 = \{\alpha_{i_1} + \beta_{i_2} + \delta \phi_{i_1i_2}\}$
  - $L \cap W_2 = \{\beta_{i_2} + \gamma_{i_3} + \delta' \psi_{i_2i_3}\}$

\[\rightarrow\] $L = (L \cap L_{\{1,2\}}) + (L \cap L_{\{2,3\}})$

- Intuitively conformality represents decomposability of $L$
Ex. 3-way conditional independence model

- $W_1 := L\{1,2\}, \ W_2 := L\{2,3\}$
- $W = W_1 + W_2 = L\{1,2\} + L\{2,3\}$
- $W : \log p_{i_1i_2i_3} = \alpha_{i_1} + \beta_{i_2} + \gamma_{i_3} + \delta_{i_1i_2} + \delta'_{i_2i_3}$

- $L : \log p_{i_1i_2i_3} = \alpha_{i_1} + \beta_{i_2} + \gamma_{i_3} + \delta\phi_{i_1i_2} + \delta'\psi_{i_2i_3}$
  - $\phi_{i_1i_2}, \psi_{i_2i_3} : \text{known functions}$
  - $\delta, \delta' : \text{free parameters}$
  - $L \cap W_1 = \{\alpha_{i_1} + \beta_{i_2} + \delta\phi_{i_1i_2}\}$
  - $L \cap W_2 = \{\beta_{i_2} + \gamma_{i_3} + \delta'\psi_{i_2i_3}\}$
  - $\rightarrow L = (L \cap L\{1,2\}) + (L \cap L\{2,3\})$

- Intuitively conformality represents decomposability of $L$
Ex. 3-way conditional independence model

- \( W_1 := L_{\{1,2\}}, \ W_2 := L_{\{2,3\}} \)
- \( W = W_1 + W_2 = L_{\{1,2\}} + L_{\{2,3\}} \)
- \( W : \log p_{i_1 i_2 i_3} = \alpha_{i_1} + \beta_{i_2} + \gamma_{i_3} + \delta_{i_1 i_2} + \delta'_{i_2 i_3} \)

- \( L : \log p_{i_1 i_2 i_3} = \alpha_{i_1} + \beta_{i_2} + \gamma_{i_3} + \delta \phi_{i_1 i_2} + \delta' \psi_{i_2 i_3} \)
  - \( \phi_{i_1 i_2}, \psi_{i_2 i_3} : \text{known functions} \)
  - \( \delta, \delta' : \text{free parameters} \)
  - \( L \cap W_1 = \{ \alpha_{i_1} + \beta_{i_2} + \delta \phi_{i_1 i_2} \} \)
  - \( L \cap W_2 = \{ \beta_{i_2} + \gamma_{i_3} + \delta' \psi_{i_2 i_3} \} \)

\[
\rightarrow L = (L \cap L_{\{1,2\}}) + (L \cap L_{\{2,3\}})
\]

- Intuitively conformality represents decomposability of \( L \)
Ex. 3-way conditional independence model

- $W_1 := L_{1,2}$, $W_2 := L_{2,3}$
- $W = W_1 + W_2 = L_{1,2} + L_{2,3}$
- $W : \log p_{i_1 i_2 i_3} = \alpha_{i_1} + \beta_{i_2} + \gamma_{i_3} + \delta_{i_1 i_2} + \delta'_{i_2 i_3}$

\[ \begin{array}{ccc} 1 & 2 & 3 \end{array} \]

- $L : \log p_{i_1 i_2 i_3} = \alpha_{i_1} + \beta_{i_2} + \gamma_{i_3} + \delta \phi_{i_1 i_2} + \delta' \psi_{i_2 i_3}$
  - $\phi_{i_1 i_2}, \psi_{i_2 i_3}$ : known functions
  - $\delta, \delta'$: free parameters
  - $L \cap W_1 = \{\alpha_{i_1} + \beta_{i_2} + \delta \phi_{i_1 i_2}\}$
  - $L \cap W_2 = \{\beta_{i_2} + \gamma_{i_3} + \delta' \psi_{i_2 i_3}\}$

\[ \rightarrow \quad L = (L \cap L_{1,2}) + (L \cap L_{2,3}) \] comformal!

- Intuitively conformality represents decomposability of $L$
Ex. 3-way conditional independence model

- \( W_1 := L_{\{1,2\}} \), \( W_2 := L_{\{2,3\}} \)
- \( W = W_1 + W_2 = L_{\{1,2\}} + L_{\{2,3\}} \)
- \( L : \log p_{i_1 i_2 i_3} = \alpha_{i_1} + \beta_{i_2} + \gamma_{i_3} + \delta_{i_1 i_2} + \delta'_{i_2 i_3} \)
- \( L' : \log p_{i_1 i_2 i_3} = \alpha_{i_1} + \beta_{i_2} + \gamma_{i_3} + \delta(\phi_{i_1 i_2} + \psi_{i_2 i_3}) \)
- \( L \cap L_{\{1,2\}} = \{\alpha_{i_1} + \beta_{i_2}\} \)
- \( L \cap L_{\{2,3\}} = \{\beta_{i_2} + \gamma_{i_3}\} \)
- \( (L \cap L_{\{1,2\}}) + (L \cap L_{\{2,3\}}) = \alpha_{i_1} + \beta_{i_2} + \gamma_{i_3} \)
- \( L \supset (L \cap L_{\{1,2\}}) + (L \cap L_{\{2,3\}}) \)
Ex. 3-way conditional independence model

- $W_1 := L_{\{1,2\}}, \ W_2 := L_{\{2,3\}}$
- $W = W_1 + W_2 = L_{\{1,2\}} + L_{\{2,3\}}$
- $L : \log p_{i_1 i_2 i_3} = \alpha_{i_1} + \beta_{i_2} + \gamma_{i_3} + \delta_{i_1 i_2} + \delta'_{i_2 i_3}$

\[
\begin{array}{ccc}
1 & 2 & 3 \\
\end{array}
\]

- $L' : \log p_{i_1 i_2 i_3} = \alpha_{i_1} + \beta_{i_2} + \gamma_{i_3} + \delta(\phi_{i_1 i_2} + \psi_{i_2 i_3})$
- $L \cap L_{\{1,2\}} = \{\alpha_{i_1} + \beta_{i_2}\}$
- $L \cap L_{\{2,3\}} = \{\beta_{i_2} + \gamma_{i_3}\}$

\[
\rightarrow (L \cap L_{\{1,2\}}) + (L \cap L_{\{2,3\}}) = \alpha_{i_1} + \beta_{i_2} + \gamma_{i_3}
\]

\[
\rightarrow L \supset (L \cap L_{\{1,2\}}) + (L \cap L_{\{2,3\}})
\]
Ex. 3-way conditional independence model

- $W_1 := L_{\{1,2\}}$, $W_2 := L_{\{2,3\}}$
- $W = W_1 + W_2 = L_{\{1,2\}} + L_{\{2,3\}}$
- $L : \log p_{i_1i_2i_3} = \alpha_{i_1} + \beta_{i_2} + \gamma_{i_3} + \delta_{i_1i_2} + \delta'_{i_2i_3}$
- $L' : \log p_{i_1i_2i_3} = \alpha_{i_1} + \beta_{i_2} + \gamma_{i_3} + \delta(\phi_{i_1i_2} + \psi_{i_2i_3})$

- $L \cap L_{\{1,2\}} = \{\alpha_{i_1} + \beta_{i_2}\}$
- $L \cap L_{\{2,3\}} = \{\beta_{i_2} + \gamma_{i_3}\}$

\[ \rightarrow (L \cap L_{\{1,2\}}) + (L \cap L_{\{2,3\}}) = \alpha_{i_1} + \beta_{i_2} + \gamma_{i_3} \]

\[ \rightarrow \boxed{L \supset (L \cap L_{\{1,2\}}) + (L \cap L_{\{2,3\}})} \quad \text{not comformal} \]
Terminologies on hypergraphs

- A divider $S$ of $\text{red}\Delta$:
  - $\exists u, v \text{ s.t. } S$ is a minimal clique separator separating $u$ and $v$
    - $S$ : the set of dividers of $\text{red}\Delta$
- $u$ and $v$ are tightly connected:
  - there is no divider which separates $u$ and $v$
- compact component $C$:
  - any two vertices in $C$ are tightly connected
    - $C$ : the set of maximal compact components of $\text{red}\Delta$

**ex.** A divider : $\{3, 4\}$,

Compact components : $\{1, 2, 3, 4\}$, $\{3, 4, 5, 6\}$
Terminologies on hypergraphs

- A divider $S$ of $\text{red}\Delta$:
  $\exists u, v$ s.t. $S$ is a minimal clique separator separating $u$ and $v$
  - $S$ : the set of dividers of $\text{red}\Delta$

- $u$ and $v$ are tightly connected:
  there is no divider which separates $u$ and $v$

- compact component $C$:
  any two vertices in $C$ are tightly connected
  - $C$ : the set of maximal compact components of $\text{red}\Delta$

ex. A divider : $\{3, 4\}$,
Compact components : $\{1, 2, 3, 4\}$, $\{3, 4, 5, 6\}$

![Diagram of hypergraph]
Terminologies on hypergraphs

- A divider $S$ of $\text{red}\Delta$:
  $\exists u, v \text{ s.t. } S$ is a minimal clique separator separating $u$ and $v$
  - $S$: the set of dividers of $\text{red}\Delta$

- $u$ and $v$ are tightly connected:
  there is no divider which separates $u$ and $v$

- compact component $C$:
  any two vertices in $C$ are tightly connected
  - $C$: the set of maximal compact components of $\text{red}\Delta$

ex. A divider: $\{3, 4\}$,
Compact components: $\{1, 2, 3, 4\}$, $\{3, 4, 5, 6\}$
A divider $S$ of $\text{red}\Delta$:
$\exists u, v \text{ s.t. } S$ is a minimal clique separator separating $u$ and $v$
- $S$: the set of dividers of $\text{red}\Delta$

$u$ and $v$ are tightly connected:
there is no divider which separates $u$ and $v$

compact component $C$:
any two vertices in $C$ are tightly connected
- $C$: the set of maximal compact components of $\text{red}\Delta$

ex. A divider: $\{3, 4\}$,
Compact components: $\{1, 2, 3, 4\}$, $\{3, 4, 5, 6\}$
A divider $S$ of $\text{red}\Delta$:

$\exists u, v$ s.t. $S$ is a minimal clique separator separating $u$ and $v$

- $S$: the set of dividers of $\text{red}\Delta$

$u$ and $v$ are tightly connected:

there is no divider which separates $u$ and $v$

compact component $C$:

any two vertices in $C$ are tightly connected

- $C$: the set of maximal compact components of $\text{red}\Delta$

ex. A divider: $\{3, 4\}$,
Compact components: $\{1, 2, 3, 4\}$, $\{3, 4, 5, 6\}$
Hierarchical subspace model (HSM)

**Definition (hierarchical subspace model)**

$L$ is a hierarchical subspace model of a hierarchical model $L_\Delta$ if:

1. $L_S \subset L$ for each $S \in \mathcal{S}$
2. $L$ is conformal to $\{L_C, C \in \mathcal{C}\}$

- $L_S \subset L \Rightarrow \hat{p}(i_S) = x(i_S)/n$
  - $\hat{p}(i_S)$: MLE for $L \cap L_S$, $x(i_S)$: a marginal frequency for $i_S$
- **Conformality**: Decomposability of the model
  - $L_\Delta = L_{\{1,2\}} + L_{\{2,3\}}$
  - $L$: $\log p_{ijk} = \alpha_i + \beta_j + \gamma_k + \delta \phi_{ij} + \delta' \psi_{jk}$
    - $\Rightarrow$ HSM of $L_\Delta$
  - $L'$: $\log p_{ijk} = \alpha_i + \beta_j + \gamma_k + \delta (\phi_{ij} + \psi_{jk})$
    - $\Rightarrow L'$ is not conformal to $\{L_{\{1,2\}}, L_{\{2,3\}}\}$
    - $\Rightarrow L'$ is not an HSM of $L_\Delta$
  - $L'$ is an HSM of the saturated model
Hierarchical subspace model (HSM)

Definition (hierarchical subspace model)

$L$ is a hierarchical subspace model of a hierarchical model $L_{\Delta}$ if

1. $L_S \subset L$ for each $S \in S$
2. $L$ is conformal to $\{L_C, C \in C\}$

$L_S \subset L \Rightarrow \hat{p}(i_S) = x(i_S)/n$

- $\hat{p}(i_S)$: MLE for $L \cap L_S$, $x(i_S)$: a marginal frequency for $i_S$

Conformality: Decomposability of the model

$L_{\Delta} = L_{\{1,2\}} + L_{\{2,3\}}$
$L$: $\log p_{ijk} = \alpha_i + \beta_j + \gamma_k + \delta \phi_{ij} + \delta' \psi_{jk}$
$\Rightarrow$ HSM of $L_{\Delta}$

$L'$: $\log p_{ijk} = \alpha_i + \beta_j + \gamma_k + \delta(\phi_{ij} + \psi_{jk})$
$\Rightarrow L'$ is not conformal to $\{L_{\{1,2\}}, L_{\{2,3\}}\}$
$\Rightarrow L'$ is not an HSM of $L_{\Delta}$

$L'$ is an HSM of the saturated model
Hierarchical subspace model (HSM)

Definition (hierarchical subspace model)

$L$ is a hierarchical subspace model of a hierarchical model $L_\Delta$ if

1. $L_S \subset L$ for each $S \in S$
2. $L$ is conformal to $\{L_C, C \in C\}$

- $L_S \subset L \Rightarrow \hat{p}(i_S) = x(i_S)/n$
- $\hat{p}(i_S)$: MLE for $L \cap L_S$, $x(i_S)$: a marginal frequency for $i_S$

Conformality: Decomposability of the model

- $L_\Delta = L_{\{1,2\}} + L_{\{2,3\}}$
- $L$: $\log p_{ijk} = \alpha_i + \beta_j + \gamma_k + \delta \phi_{ij} + \delta' \psi_{jk}$
  $\Rightarrow$ HSM of $L_\Delta$
- $L'$: $\log p_{ijk} = \alpha_i + \beta_j + \gamma_k + \delta (\phi_{ij} + \psi_{jk})$
  $\Rightarrow L'$ is not conformal to $\{L_{\{1,2\}}, L_{\{2,3\}}\}$
  $\Rightarrow L'$ is not an HSM of $L_\Delta$
- $L'$ is an HSM of the saturated model
MLE of HSM

- $L$ : HSM of $L_{\Delta}$
- $C$ : the set of compact components of $\text{red}\Delta$
- $S$ : the set of divider of $\text{red}\Delta$
- $\hat{p}(i)$ : MLE of $p(i)$
- $\hat{p}(i_C)$ : MLE for $L \cap L_{C}$

\[
\hat{p}(i) = \frac{\prod_{C \in C} \hat{p}(i_C)}{\prod_{S \in S} \hat{p}(i_S)} = \frac{\prod_{C \in C} \hat{p}(i_C)}{\prod_{S \in S} x(i_S)/n}
\]
So far we have discussed the definition of HSM

For a given $L_\Delta$, we can obtain an HSM with the same decomposability as $L_\Delta$

Next we discuss how we can localize the inference of a given log-affine model $L$

Every log-affine model $L$ is an HSM of saturated model

Every log affine model has a hierarchical model for which $L$ is HSM

From a viewpoint of localization of the inference, a natural question is to look for a small $L_\Delta$ for which $L$ is an HSM

We derive the smallest decomposable model for which $L$ is an HSM

Ambient decomposable model
So far we have discussed the definition of HSM

For a given $L_\Delta$, we can obtain an HSM with the same decomposability as $L_\Delta$

Next we discuss how we can localize the inference of a given log-affine model $L$

Every log-affine model $L$ is an HSM of saturated model

Every log affine model has a hierarchical model for which $L$ is HSM

From a viewpoint of localization of the inference, a natural question is to look for a small $L_\Delta$ for which $L$ is an HSM

We derive the smallest decomposable model for which $L$ is an HSM

\[\Rightarrow\]

Ambient decomposable model
So far we have discussed the definition of HSM.

For a given $L_\Delta$, we can obtain an HSM with the same decomposability as $L_\Delta$.

Next we discuss how we can localize the inference of a given log-affine model $L$.

Every log-affine model $L$ is an HSM of saturated model.

Every log-affine model has a hierarchical model for which $L$ is HSM.

From a viewpoint of localization of the inference, a natural question is to look for a small $L_\Delta$ for which $L$ is an HSM.

We derive the smallest decomposable model for which $L$ is an HSM.

Ambient decomposable model
So far we have discussed the definition of HSM. For a given $L_\Delta$, we can obtain an HSM with the same decomposability as $L_\Delta$.

Next we discuss how we can localize the inference of a given log-affine model $L$.

Every log-affine model $L$ is an HSM of saturated model.

Every log affine model has a hierarchical model for which $L$ is HSM.

From a viewpoint of localization of the inference, a natural question is to look for a small $L_\Delta$ for which $L$ is an HSM.

We derive the smallest decomposable model for which $L$ is an HSM.
Decomposability of log-affine model

- So far we have discussed the definition of HSM
- For a given $L_\Delta$, we can obtain an HSM with the same decomposability as $L_\Delta$
- Next we discuss how we can localize the inference of a given log-affine model $L$
- Every log-affine model $L$ is an HSM of saturated model
- Every log affine model has a hierarchical model for which $L$ is HSM
- From a viewpoint of localization of the inference, a natural question is to look for a small $L_\Delta$ for which $L$ is an HSM
- We derive the smallest decomposable model for which $L$ is an HSM

Ambient decomposable model
So far we have discussed the definition of HSM

For a given $L_\Delta$, we can obtain an HSM with the same decomposability as $L_\Delta$

Next we discuss how we can localize the inference of a given log-affine model $L$

Every log-affine model $L$ is an HSM of saturated model

Every log affine model has a hierarchical model for which $L$ is HSM

From a viewpoint of localization of the inference, a natural question is to look for a small $L_\Delta$ for which $L$ is an HSM

We derive the smallest decomposable model for which $L$ is an HSM

$\downarrow$

Ambient decomposable model
Ambient decomposable model
Decomposition of log-affine model

- In the case of the hierarchical model
  interaction terms $\Leftrightarrow \Delta$
  decomposition of $L_\Delta \Leftrightarrow$ decomposition of red$\Delta$ (hypergraph)

- $L$ is not necessarily an HSM of a hierarchical model which has the same interaction as $L$
  ex) 3-way conditional independence model
  $L_\Delta = L_{\{1,2\}} + L_{\{2,3\}}$
  $\log p_{ijk} = \alpha_i + \beta_j + \gamma_k + \delta_{ij} + \delta'_{jk}$
  $L: \log p_{ijk} = \alpha_i + \beta_j + \gamma_k + \delta \phi_{ij} + \delta' \psi_{jk}$
  $L': \log p_{ijk} = \alpha_i + \beta_j + \gamma_k + \delta (\phi_{ij} + \psi_{jk})$

- Decomposition of $L$ does not always correspond to simplicial complex induced by interaction terms

- We need to define the notion of decomposition of $L$
Decomposition of log-affine model

In the case of the hierarchical model
interaction terms $\Leftrightarrow \Delta$
decomposition of $L_{\Delta} \Leftrightarrow$ decomposition of red$\Delta$ (hypergraph)

$L$ is not necessarily an HSM of a hierarchical model which has the same interaction as $L$

ex) 3-way conditional independence model

$L_{\Delta} = L_{\{1,2\}} + L_{\{2,3\}}$

$log p_{ijk} = \alpha_i + \beta_j + \gamma_k + \delta_{ij} + \delta'_{jk}$

$L$: $log p_{ijk} = \alpha_i + \beta_j + \gamma_k + \delta \phi_{ij} + \delta' \psi_{jk}$

$L'$: $log p_{ijk} = \alpha_i + \beta_j + \gamma_k + \delta (\phi_{ij} + \psi_{jk})$

Decomposition of $L$ does not always correspond to simplicial complex induced by interaction terms

We need to define the notion of decomposition of $L$
Decomposition of log-affine model

- In the case of the hierarchical model interaction terms $\Leftrightarrow \Delta$
  decomposition of $L_\Delta \Leftrightarrow$ decomposition of $\text{red}\Delta$ (hypergraph)

- $L$ is not necessarily an HSM of a hierarchical model which has the same interaction as $L$
  ex) 3-way conditional independence model
    - $L_\Delta = L_{\{1,2\}} + L_{\{2,3\}}$
      $\log p_{ijk} = \alpha_i + \beta_j + \gamma_k + \delta_{ij} + \delta'_{jk}$
    - $L$: $\log p_{ijk} = \alpha_i + \beta_j + \gamma_k + \delta\phi_{ij} + \delta'\psi_{jk}$
    - $L'$: $\log p_{ijk} = \alpha_i + \beta_j + \gamma_k + \delta(\phi_{ij} + \psi_{jk})$

- Decomposition of $L$ does not always correspond to simplicial complex induced by interaction terms

- We need to define the notion of decomposition of $L$
Decomposition of log-affine model

In the case of the hierarchical model
interaction terms $\Leftrightarrow \Delta$
decomposition of $L_\Delta \Leftrightarrow$ decomposition of $\text{red}\Delta$ (hypergraph)

$L$ is not necessarily an HSM of a hierarchical model which has the same interaction as $L$
ex) 3-way conditional independence model

- $L_\Delta = L_{\{1,2\}} + L_{\{2,3\}}$
- $\log p_{ijk} = \alpha_i + \beta_j + \gamma_k + \delta_{ij} + \delta'_{jk}$

$L$: $\log p_{ijk} = \alpha_i + \beta_j + \gamma_k + \delta_{ij} + \delta'_{jk}$

$L'$: $\log p_{ijk} = \alpha_i + \beta_j + \gamma_k + \delta(\phi_{ij} + \psi_{jk})$

Decomposition of $L$ does not always correspond to simplicial complex induced by interaction terms

We need to define the notion of decomposition of $L$
 Hypergraph induced by HSM

- $S \subset [m]$ is a partial edge separator (pes) of $L$ if
  - $L_S \subset L$
  - For disjoint subsets $A_1 \cup A_2 \cup S = [m]$, $L$ is conformal to $\{L_{A_1 \cup S}, L_{A_2 \cup S}\}$

- $(A_1, A_2, S)$ is called a decomposition of $L$

- $u$ and $v$ are tightly connected in $L$ if there is no pes of $L$ s.t. $u \in A_1$ and $v \in A_2$

- Extended compact component (ECC)
  - a set of vertices any two of which are tightly connected in $L$

- Hypergraph $\mathcal{H}$ induced by $L$
  - the set of maximal ECCs
\( S \subset [m] \) is a partial edge separator (pes) of \( L \) if

- \( L_S \subset L \)
- For disjoint subsets \( A_1 \cup A_2 \cup S = [m] \), \( L \) is conformal to \( \{ L_{A_1 \cup S}, L_{A_2 \cup S} \} \)

\((A_1, A_2, S)\) is called a decomposition of \( L \)

- \( u \) and \( v \) are tightly connected in \( L \) if there is no pes of \( L \) s.t. \( u \in A_1 \) and \( v \in A_2 \)

- Extended compact component (ECC)
  - a set of vertices any two of which are tightly connected in \( L \)

- Hypergraph \( \mathcal{H} \) induced by \( L \)
  - the set of maximal ECCs
$S \subset [m]$ is a partial edge separator (pes) of $L$ if

- $L_S \subset L$
- For disjoint subsets $A_1 \cup A_2 \cup S = [m]$, $L$ is conformal to $\{L_{A_1 \cup S}, L_{A_2 \cup S}\}$

$(A_1, A_2, S)$ is called a decomposition of $L$

$u$ and $v$ are tightly connected in $L$ if there is no pes of $L$ s.t. $u \in A_1$ and $v \in A_2$

Extended compact component (ECC)

a set of vertices any two of which are tightly connected in $L$

Hypergraph $\mathcal{H}$ induced by $L$

the set of maximal ECCs
Hypergraph induced by HSM

- $S \subset [m]$ is a **partial edge separator (pes)** of $L$ if
  - $L_S \subset L$
  - For disjoint subsets $A_1 \cup A_2 \cup S = [m]$, $L$ is conformal to $\{L_{A_1 \cup S}, L_{A_2 \cup S}\}$

- $(A_1, A_2, S)$ is called a **decomposition** of $L$

- $u$ and $v$ are **tightly connected** in $L$ if there is no pes of $L$ s.t. $u \in A_1$ and $v \in A_2$

- **Extended compact component (ECC)**
  - A set of vertices any two of which are tightly connected in $L$

- Hypergraph $\mathcal{H}$ induced by $L$
  - The set of maximal ECCs
Hypergraph induced by HSM

- $S \subset [m]$ is a **partial edge separator (pes)** of $L$ if
  - $L_S \subset L$
  - For disjoint subsets $A_1 \cup A_2 \cup S = [m]$, $L$ is conformal to $\{L_{A_1 \cup S}, L_{A_2 \cup S}\}$
- $(A_1, A_2, S)$ is called a **decomposition** of $L$
- $u$ and $v$ are **tightly connected** in $L$ if there is no pes of $L$ s.t. $u \in A_1$ and $v \in A_2$
- **Extended compact component (ECC)**
  - a set of vertices any two of which are tightly connected in $L$
- **Hypergraph $H$ induced by $L$**
  - the set of maximal ECCs
**Ambient decomposable model**

- $L_H$: the hierarchical model induced by $H$

**Theorem**

$L_H$ is the ambient decomposable model of $L$

**MLE**

$$\hat{p}(i) = \frac{\prod_{C \in H} \hat{p}(i_C)}{\prod_{S \in S} \hat{p}(i_S)} = \frac{\prod_{C \in H} \hat{p}(i_C)}{\prod_{S \in S} x(i_S)/n}.$$  

- $S$: the set of divider of $H$
- $\hat{p}(i_C)$ depends only on the marginal table $x(i_C)$

⚠️ The computation of the MLE is localized to ECCs
**Ambient decomposable model**

- $L_H$ : the hierarchical model induced by $H$

**Theorem**

$L_H$ is the ambient decomposable model of $L$

- **MLE**

  $$\hat{p}(i) = \frac{\prod_{C \in H} \hat{p}(i_C)}{\prod_{S \in S} \hat{p}(i_S)} = \frac{\prod_{C \in H} \hat{p}(i_C)}{\prod_{S \in S} x(i_S)/n}.$$  

- $S$ : the set of divider of $H$
- $\hat{p}(i_C)$ depends only on the marginal table $x(i_C)$

↓

The computation of the MLE is localized to ECCs
Markov basis for HSM
Dobra and Sullivant (2004)
A Markov basis of a hierarchical model $L_\Delta$ is computed recursively from Markov bases of marginal models $L_C$ for all $C \in \mathcal{C}$

A Markov basis of an HSM is also computed from Markov bases of $L \cap L_C$, $C \in \mathcal{H}$
Dobra and Sullivant (2004)
A Markov basis of a hierarchical model $L_\Delta$ is computed recursively from Markov bases of marginal models $L_C$ for all $C \in \mathcal{C}$

A Markov basis of an HSM is also computed from Markov bases of $L \cap L_C$, $C \in \mathcal{H}$
Dobra and Sullivant (2004)
A Markov basis of a hierarchical model $L_\Delta$ is computed recursively from Markov bases of marginal models $L_C$ for all $C \in \mathcal{C}$

A Markov basis of an HSM is also computed from Markov bases of $L \cap L_C$, $C \in \mathcal{H}$
Dobra and Sullivant (2004)
A Markov basis of a hierarchical model $L_\Delta$ is computed recursively from Markov bases of marginal models $L_C$ for all $C \in \mathcal{C}$

A Markov basis of an HSM is also computed from Markov bases of $L \cap L_C$, $C \in \mathcal{H}$
Dobra and Sullivant (2004)
A Markov basis of a hierarchical model $L_\Delta$ is computed recursively from Markov bases of marginal models $L_C$ for all $C \in C$

A Markov basis of an HSM is also computed from Markov bases of $L \cap L_C$, $C \in \mathcal{H}$
Dobra and Sullivant (2004)
A Markov basis of a hierarchical model $L_\Delta$ is computed recursively from Markov bases of marginal models $L_C$ for all $C \in \mathcal{C}$

A Markov basis of an HSM is also computed from Markov bases of $L \cap L_C$, $C \in \mathcal{H}$
Dobra and Sullivant (2004)
A Markov basis of a hierarchical model $L_\Delta$ is computed recursively from Markov bases of marginal models $L_C$ for all $C \in \mathcal{C}$.

A Markov basis of an HSM is also computed from Markov bases of $L \cap L_C$, $C \in \mathcal{H}$. 
Dobra and Sullivant (2004)
A Markov basis of a hierarchical model $L_\Delta$ is computed recursively from Markov bases of marginal models $L_C$ for all $C \in \mathcal{C}$

A Markov basis of an HSM is also computed from Markov bases of $L \cap L_C$, $C \in \mathcal{H}$
Dobra and Sullivant (2004)
A Markov basis of a hierarchical model $L_\Delta$ is computed recursively from Markov bases of marginal models $L_C$ for all $C \in \mathcal{C}$

A Markov basis of an HSM is also computed from Markov bases of $L \cap L_C$, $C \in \mathcal{H}$
Dobra and Sullivant (2004)
A Markov basis of a hierarchical model $L_\Delta$ is computed recursively from Markov bases of marginal models $L_C$ for all $C \in \mathcal{C}$.

A Markov basis of an HSM is also computed from Markov bases of $L \cap L_C$, $C \in \mathcal{H}$. 
Local computation of Markov bases

- \((A_1, A_2, S)\) : decomposition of \(L\)
- \(z_1\) : a move \(L \cap L_{A_1 \cup S}\)

\[
z_1 = \left\{ \{(i_1, j_1), \ldots, (i_d, j_d)\} \parallel \{(i'_1, j_1), \ldots, (i'_d, j_d)\} \right\},
\]

\[i_k, i'_k \in \mathcal{I}_{A_1}, \quad j_k \in \mathcal{I}_{S}\]

- \((i_1, j_1), \ldots, (i_d, j_d)\) : cells (with replication) of “+” elements of \(z_1\)
- \((i'_1, j_1), \ldots, (i'_d, j_d)\) : cells (with replication) of “−” elements of \(z_1\)

\(L_S \subset L \Rightarrow z_1(i_S) = 0\)
Extension of moves

**Definition (Ext(\(\mathcal{B}(A_1 \cup S) \rightarrow L\))**

\(\mathcal{B}(A_1 \cup S)\): a Markov basis of \(L \cap L_{A_1 \cup S}\)

\(k := \{k_1, \ldots, k_d\} \in \mathcal{I}_{A_2} \times \cdots \times \mathcal{I}_{A_2}\),

Define \(z_1^k\) by

\[
z_1^k := \{(i_1, j_1, k_1), \ldots, (i_d, j_d, k_d)\} \ll \{(i'_1, j_1, k_1), \ldots, (i'_d, j_d, k_d)\}\.
\]

Then define \(\text{Ext}(\mathcal{B}(V_1) \rightarrow L)\) by

\[
\text{Ext}(\mathcal{B}(V_1) \rightarrow L) := \{z_1^k \mid k \in \mathcal{I}_{A_2} \times \cdots \times \mathcal{I}_{A_2}\}
\]

\[
\begin{array}{cccc}
i_1j_1 & i'_1j_1 & i_2j_2 & i'_2j_2 \\
1 & -1 & 1 & -1 \\
\end{array}
\Rightarrow
\begin{array}{cccc}
k_1 \\
1 & -1 & 0 & 0 \\
k_2 \\
0 & 0 & 1 & -1 \\
k_1 \\
1 & -1 & 1 & -1 \\
\end{array}
\]
Local computation of Markov bases

Theorem

\[ V_1 := A_1 \cup S, \quad V_2 := A_2 \cup S \]
\[ \mathcal{B}(V_1), \quad \mathcal{B}(V_2) : \text{MB of } L \cup L_{V_1}, \quad L \cup L_{V_2} \]
\[ \mathcal{B}_{V_1,V_2} : \text{a MB of decomposable model with two cliques } V_1, \ V_2 \]

Then

\[ \mathcal{B} := \text{Ext}(\mathcal{B}(V_1) \rightarrow L) \cup \text{Ext}(\mathcal{B}(V_2) \rightarrow L) \cup \mathcal{B}_{V_1,V_2} \]

is a Markov basis of \( L \)
Numerical example
Woman and Mathematics (WAM) data: 6-way contingency table
a questionnaire from high school students in NJ
(Fowlkes(1988), Højsgaard(2003))

1. Attendance in math lectures (attended=1, not=2)
2. Sex (female=1, male=2)
3. School type (suburban=1, urban=2)
4. Agree in statement “I’ll need mathematics in my future work”
   (agree=1, disagree=2)
5. Subject preference (math-science=1, liberal arts=2)
6. Future plans (college=1, job=2)
1. **$H_1$: decomposable model**

$$L_1 = L_{\{1,2,3,5\}} + L_{\{2,3,4,5\}} + L_{\{3,4,5,6\}}$$

2. **$H_0$: split model**

$$L_0 = L_{\{1,2,3,5\}} + L^{j_3=1}_{\{2,5\}} + L^{j_3=1}_{\{4,5\}} + L^{j_3=2}_{\{2,4,5\}} + L_{\{3,4,5,6\}}.$$
Numerical example: model

1. \( H_1: \) decomposable model

\[
L_1 = L\{1,2,3,5\} + L\{2,3,4,5\} + L\{3,4,5,6\}
\]

2. \( H_0: \) split model

\[
L_0 = L\{1,2,3,5\} + L_{j_3=1}^{2,5} + L_{j_3=1}^{4,5} + L_{j_3=2}^{2,4,5} + L\{3,4,5,6\}.
\]

\[
\Rightarrow
\]

1. attendance in math
2. sex
3. school type
4. necessity of math in future
5. subject preference
6. future plan
Numerical example: Markov basis

- $\mathcal{B}_1$: Markov basis of $L_1$
  \[
  \mathcal{B}_1 = \mathcal{B}_{\{1,2,3,5\},\{2,3,4,5,6\}} \cup \mathcal{B}_{\{1,2,3,4,5\},\{3,4,5,6\}}
  \]

- $\mathcal{B}_0$: Markov basis of $L_0$
  \[
  \mathcal{B}_0 = \mathcal{B}_{\{1,2,5\},\{4,5,6\}} \cup \mathcal{B}_1
  \]

- $\mathcal{B}_{C_1,C_2}$: a Markov basis of a decomposable model with two cliques $C_1$ and $C_2$

- $\mathcal{B}_{C_1}^{i_\delta}$: $\mathcal{B}_C$ on $i_\delta$-slice
Numerical example: results

- We used LR statistic as a test statistic

<table>
<thead>
<tr>
<th>Deviance</th>
<th>asymptotic $\chi_2^2$</th>
<th>MCMC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.851</td>
<td>0.396</td>
<td>0.399±0.012</td>
</tr>
</tbody>
</table>

- Split model is accepted

- Histogram MCMC
- Dotted line asymptotic $\chi_2^2$
Conclusion
Summary

- We proposed a hierarchical subspace model by defining the notion of conformality of linear subspaces to a given hierarchical model.
- The notion of an HSM gives a modeling strategy of multiway tables and unifies various models of interaction effects.
- We illustrated practical advantage of our modeling strategy with some data sets.

Future work

- Is it possible to treat nonlinear models such as the RC association model in the framework of HSM?
Hierarchical subspace models for contingency tables
in preparation

Algebraic algorithms for sampling from conditional distributions

Decomposition of a hypergraph by partial-edge separators

A divide-and-conquer algorithm for generating Markov bases of multi-way tables

Split models for contingency tables
Computational Statistics & Data Analysis, 42, 621–645.
Thank you for your attention!