MATHEMATICAL ENGINEERING TECHNICAL REPORTS

Improving on the Maximum Likelihood Estimators of the means in Poisson Decomposable Graphical Models

Hisayuki Hara and Akimichi Takemura

METR 2005–08 April 2005

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

WWW page: http://www.i.u-tokyo.ac.jp/mi/mi-e.htm

Improving on the Maximum Likelihood Estimators of the means in Poisson Decomposable Graphical Models

Hisayuki Hara*and Akimichi Takemura†
Department of Geosystem Engineering, University of Tokyo,
Department of Mathematical Informatics, University of Tokyo

April, 2005

Abstract

In this article we study the simultaneous estimation of the means in Poisson decomposable graphical models. We derive some classes of estimators which improve on the maximum likelihood estimator under the normalized squared losses. Our estimators are based on the argument in Chou[3] and shrink the maximum likelihood estimator depending on the marginal frequencies of variables forming a complete subgraph of the conditional independence graph.

Keywords and phrases : chordal graph, contingency table, decomposable graph, decomposable Poisson model, inadmissibility, perfect sequence, simultaneous estimation, shrinkage estimation, unbiased estimation of risk difference.

1 Introduction

Suppose we have I independent Poisson observations $x_1, \ldots, x_I$ with unknown mean parameters $\lambda_1, \ldots, \lambda_I$. Consider the problem of estimating $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_I)$ under the loss

$$\mathrm{L}(\boldsymbol{\lambda}, \hat{\boldsymbol{\lambda}}) = \sum_{i=1}^{I} \frac{1}{\lambda_i}(\lambda_i - \hat{\lambda}_i)^2.$$

The usual estimator is $\boldsymbol{x} = (x_1, \ldots, x_I)'$ which is both the maximum likelihood and the uniformly minimum variance unbiased estimator. For $I = 1$ $\boldsymbol{x}$ is known to be admissible(See Brown and Hwang[2] for example). Since Clevenson and Zidek[4] proved the inadmissibility of $\boldsymbol{x}$ for $I \geq 2$ and derived the class of estimators which improve on $\boldsymbol{x}$ under the above loss, the estimation of $\boldsymbol{\lambda}$ in higher dimensions has received considerable attention and a lot of research have been devoted to this problem.

*hara@geosys.t.u-tokyo.ac.jp
†takemura@stat.t.u-tokyo.ac.jp

Tsui and Press[19] derived improved estimators under k-normalized squared error loss, $\mathrm{L}_k(\boldsymbol{\lambda}, \hat{\boldsymbol{\lambda}}) = \sum_{i=1}^p (\hat{\lambda}_i - \lambda_i)^2/\lambda_i^k$. Hwang[12] generalized the identity of Hudson[11] and derived an unbiased estimator of the difference of two risk functions between the MLE and the improved estimators. Hwang[12] also extended the problem to the estimation of the mean parameters of a subclass of discrete exponential families which includes Poisson distribution and derived some estimators which dominate the MLE by using the identity. Chou[3] gave classes of improved estimators for a wider class of discrete exponential families. In the setting of simultaneous prediction of Poisson random variables, Komaki[14] derived fundamental results on admissibility under the Kullback-Leibler loss. Other important results in this field may be found in Ghosh and Parsian[8], Ghosh, Hwang and Tsui[7], Ghosh and Yang[9], Johnstone[13], Tsui[18] etc.

The present paper considers the problem of estimating the means in Poisson decomposable graphical models. Consider a J-way layout contingency table. Let $\Delta = \{1, \ldots, J\}$ be the set of variables which corresponds to the set of vertices in the conditional independence graph. Denote the number of levels for $\delta \in \Delta$ by I_δ. We assume that $I_\delta \geq 2$ for all δ. We express the set of levels of δ by $\mathcal{I}_\delta = \{1, \ldots, I_\delta\}$. Each cell of the table is the element $i = (i_\delta)_{\delta \in \Delta}$ of the whole cells $\mathcal{I}$,

$$i \in \mathcal{I}, \quad \mathcal{I} = \prod_{\delta \in \Delta} \mathcal{I}_\delta.$$

Let the marginal cell and the set of the marginal cells for $V \subset \Delta$ be expressed by i_V and $\mathcal{I}_V$, respectively. For the vector of the cell frequencies $\boldsymbol{x} = \{x(i)\}_{i \in \mathcal{I}} \in \mathbb{Z}^{|\mathcal{I}|}$ the marginal frequency for i_V is denoted by $x(i_V)$. Define $x^+ = x(i_\emptyset) = \sum_{i \in \mathcal{I}} x(i)$.

The Poisson decomposable graphical model is expressed as follows,

$$x(i) \sim \mathrm{Po}(\lambda(i)), \quad \lambda(i) = \lambda \frac{\prod_{C \in \mathcal{C}} \alpha(i_C)}{\prod_{S \in \mathcal{S}} \alpha(i_S)^{\nu(S)}}, \tag{1}$$

$$\sum_{i_C \in \mathcal{I}_C} \alpha(i_C) = 1, \quad \sum_{i_S \in \mathcal{I}_S} \alpha(i_S) = 1,$$

where $\mathcal{C}$ is the set of cliques of the corresponding decomposable conditional independence graph $\mathcal{G}$ and $\mathcal{S}$ is the set of minimal vertex separators S with multiplicities $\nu(S)$ in any perfect sequence. $x(i)$ are supposed to be independent with respect to $i \in \mathcal{I}$. We note that the above definition includes the case where $\mathcal{G}$ is disconnected. For the disconnected $\mathcal{G}$, we suppose that $\emptyset \in \mathcal{S}$ and that $\nu(\emptyset) = \nu_{\mathcal{G}} - 1$, where $\nu_{\mathcal{G}}$ is the number of connected components of $\mathcal{G}$. In this article we address the problem of the simultaneous estimation of $\boldsymbol{\lambda} = \{\lambda(i)\}_{i \in \mathcal{I}}$ under the following normalized squared loss function

$$\mathrm{L}(\boldsymbol{\lambda}, \hat{\boldsymbol{\lambda}}) = \sum_{i \in \mathcal{I}} \frac{1}{\lambda(i)} (\lambda(i) - \hat{\lambda}(i))^2 \tag{2}$$

from the decision theoretic viewpoint.

When $\mathcal{G}$ is complete, (1) corresponds to a saturated model. A series of results on the shrinkage estimation of multivariate Poisson means which were inspired by Clevenson and Zidek[4] correspond to this setting.

The model (1) with $\mathcal{S} = \{\emptyset\}$ and $\mathcal{C} = \{\{1\}, \ldots, \{J\}\}$ is called Poisson multiplicative model. Recently Hara and Takemura[10] studied the estimation of the means in the Poisson multiplicative models and derived some classes of estimators improving on the MLE by using the argument in Clevenson and Zidek[4] and Chou[3]. Hara and Takemura[10] also showed the inadmissibility of the MLE in the three way decomposable graphical model which corresponds to the decomposable graph in Figure 1.

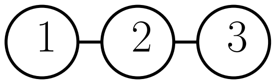

Figure 1: 3-way decomposable model

In this paper we extend the results of Hara and Takemura[10] to the general Poisson decomposable models (1) and give the classes of estimators improving on the MLE under the loss function (2). The paper is organized as follows. In Section 2 we summarize basic facts on the Poisson decomposable graphical models and decomposable graphs. In Section 3 we present some classes of estimators which dominate the MLE in the Poisson decomposable model (1). In Section 4 we give examples of improved estimators for some decomposable models. Section 5 gives Monte Carlo studies which confirm the theoretical results of the dominance relationship.

2 Basic facts on the Poisson decomposable graphical models and the decomposable graphs

2.1 Notation

In this section we define some notations which we use in the following argument. Mostly we follow the notation of Lauritzen[15]. In what follows we assume that the graph $\mathcal{G}$ is decomposable (chordal) and not complete.

For a subset of vertices $V \subset \Delta$ let $\mathcal{G}(V)$ be the subgraph induced by V. $\mathcal{C}(V)$, $\mathcal{S}(V)$ and $\nu(S, V)$ for $S \in \mathcal{S}(V)$ represent the set of the cliques, the set of the minimal vertex separators and the multiplicity of S in $\mathcal{G}(V)$, respectively.

Define I_V for $V \subset \Delta$ as

$$I_V = \prod_{\delta \in V} I_\delta$$

and $I_\emptyset \equiv 1$. Let $\mathrm{adj}(\delta, \mathcal{G})$ be the set of vertices which are adjacent to $\delta \in \Delta$ in $\mathcal{G}$.

For a set of cliques $\mathcal{C}^* \subset \mathcal{C}$, define $\Delta(\mathcal{C}^*)$ by $\Delta(\mathcal{C}^*) = \bigcup_{C \in \mathcal{C}^*} C$. We note that $\mathcal{C}^* \subset \mathcal{C}(\Delta(\mathcal{C}^*))$ but in general $\mathcal{C}^* \neq \mathcal{C}(\Delta(\mathcal{C}^*))$. For example Δ and $\mathcal{C}$ of the graph in Figure 2 is

$$\Delta = \{1, 2, 3, 4, 5\}, \quad \mathcal{C} = \{\{1, 2, 3\}, \{1, 3, 4\}, \{1, 4, 5\}\},$$

respectively. If we set $\mathcal{C}^*$ as $\mathcal{C}^* = \{\{1, 2, 3\}, \{1, 4, 5\}\}$, $\Delta(\mathcal{C}^*) = \Delta$. Thus $\mathcal{C}(\Delta(\mathcal{C}^*)) = \mathcal{C} \supsetneq \mathcal{C}^*$.

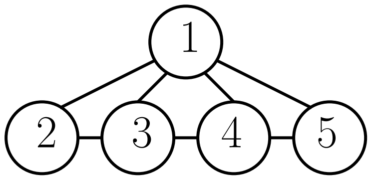

Figure 2: A decomposable graph with five vertices

2.2 Basic facts on the Poisson decomposable models

In this section we summarize some basic facts on the Poisson decomposable model (1). The joint probability function of $\boldsymbol{x}$ is

$$\begin{aligned}\Pr(\boldsymbol{x}) &= \prod_{i\in\mathcal{I}} \frac{\lambda(i)^{x(i)}}{x(i)!} e^{-\lambda(i)} \\ &= e^{-\lambda}\lambda^{x^+} \prod_{i\in\mathcal{I}} \frac{1}{x(i)!} \times \frac{\prod_{C\in\mathcal{C}} \alpha(i_C)^{x(i_C)}}{\prod_{S\in\mathcal{S}} \alpha(i_S)^{\nu(S)x(i_S)}}.\end{aligned}$$

Thus $\boldsymbol{x}_{\mathcal{C}} = \{x(i_C), i_C \in \mathcal{I}_C, C \in \mathcal{C}\}$ is the complete sufficient statistic for this model. The dimension of $\boldsymbol{x}_{\mathcal{C}}$ is $\sum_{C\in\mathcal{C}} I_C$. $\boldsymbol{x}_{\mathcal{C}}$ contains some obvious redundant elements to be minimal sufficient, but it is notationally convenient to use $\boldsymbol{x}_{\mathcal{C}}$. Following Sundberg[17] and Lauritzen[15], the marginal probability function of $\boldsymbol{x}_{\mathcal{C}}$ is

$$\Pr(\boldsymbol{x}_{\mathcal{C}}) = e^{-\lambda}\lambda^{x^+} \cdot \frac{\prod_{C\in\mathcal{C}}\prod_{i_C\in\mathcal{I}_C} \alpha(i_C)^{x(i_C)}}{\prod_{S\in\mathcal{S}}\prod_{i_S\in\mathcal{I}_S} \alpha(i_S)^{\nu(S)x(i_S)}} \cdot t(\boldsymbol{x}_{\mathcal{C}}),$$

where

$$t(\boldsymbol{x}_{\mathcal{C}}) = \frac{\prod_{S\in\mathcal{S}}\{\prod_{i_S\in\mathcal{I}_S} x(i_S)!\}^{\nu(S)}}{\prod_{C\in\mathcal{C}}\prod_{i_C\in\mathcal{I}_C} x(i_C)!}$$

and the MLE of λ and $\alpha(i_C)$ for $C \in \mathcal{C}$ are x^+ and $x(i_C)/x^+$, respectively. Therefore the MLE of $\boldsymbol{\lambda}$ is given by

$$\hat{\boldsymbol{\lambda}}^{ML}(\boldsymbol{x}_{\mathcal{C}}) = \{\hat{\lambda}^{ML}(i, \boldsymbol{x}_{\mathcal{C}})\}_{i\in\mathcal{I}},$$

$$\hat{\lambda}^{ML}(i, \boldsymbol{x}_{\mathcal{C}}) = \begin{cases} \dfrac{\prod_{C\in\mathcal{C}} x(i_C)}{\prod_{S\in\mathcal{S}} x(i_S)^{\nu(S)}}, & \text{if } \ x(i_S) \neq 0 \ \text{ for } \ \forall S \in \mathcal{S}, \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

The following lemma corresponds to the identity of Hudson[11] and Hwang[12].

Lemma 2.1. *For a real valued function g if $E|g(\boldsymbol{x}_{\mathcal{C}})| < \infty$ and $g(\boldsymbol{x}_{\mathcal{C}}) = 0$ whenever there exists $C \in \mathcal{C}$ and $i_C \in \mathcal{I}_C$ such that $x(i_C) < m$, then*

$$E\left[\frac{1}{\lambda(i)^m} g(\boldsymbol{x}_{\mathcal{C}})\right] = E\left[\frac{t(\boldsymbol{x}_{\mathcal{C}} + m\boldsymbol{e}^i_{\mathcal{C}})}{t(\boldsymbol{x}_{\mathcal{C}})} g(\boldsymbol{x}_{\mathcal{C}} + m\boldsymbol{e}^i_{\mathcal{C}})\right], \tag{4}$$

where $\boldsymbol{e}^i_{\mathcal{C}} = \{e^i(j_C), j_C \in \mathcal{I}_C, C \in \mathcal{C}\}$ *is the* $\sum_{C\in\mathcal{C}} I_C$*-dimensional vector such that*

$$e^i(j_C) = \begin{cases} 1, & \textit{for } \; j_C = i_C, \\ 0, & \textit{otherwise} \end{cases}$$

for all $j_C \in \mathcal{I}_C$, $C \in \mathcal{C}$.

PROOF.

$$\begin{aligned}
\mathrm{E}\left[\frac{1}{\lambda^m(i)} g(\boldsymbol{x}_{\mathcal{C}})\right] &= \sum_{C\in\mathcal{C}} \sum_{j_C \in \mathcal{I}_C} \frac{1}{\lambda^m(i)} g(\boldsymbol{x}_{\mathcal{C}}) \mathrm{Pr}(\boldsymbol{x}_{\mathcal{C}}) \\
&= \sum_{C\in\mathcal{C}} \sum_{j_C \in \mathcal{I}_C} \left\{ g(\boldsymbol{x}_{\mathcal{C}}) e^{-\lambda} \lambda^{x^+ - m} \cdot \frac{\prod_{C\in\mathcal{C}} \alpha(i_C)^{x(i_C)-m}}{\prod_{S\in\mathcal{S}} \alpha(i_S)^{\nu(S)(x(i_S)-m)}} \right. \\
&\qquad \left. \times \frac{\prod_{C\in\mathcal{C}} \prod_{j_C \neq i_C} \alpha(j_C)^{x(j_C)}}{\prod_{S\in\mathcal{S}} \prod_{j_S \neq i_S} \alpha(j_S)^{\nu(S) x(j_S)}} \cdot t(\boldsymbol{x}_{\mathcal{C}}) \right\} \\
&= \sum_{C\in\mathcal{C}} \sum_{j_C \in \mathcal{I}_C} \frac{t(\boldsymbol{x}_{\mathcal{C}} + m\boldsymbol{e}^i_{\mathcal{C}})}{t(\boldsymbol{x}_{\mathcal{C}})} g(\boldsymbol{x}_{\mathcal{C}} + m\boldsymbol{e}^i_{\mathcal{C}}) \mathrm{Pr}(\boldsymbol{x}_{\mathcal{C}}) \\
&= \mathrm{E}\left[\frac{t(\boldsymbol{x}_{\mathcal{C}} + m\boldsymbol{e}^i_{\mathcal{C}})}{t(\boldsymbol{x}_{\mathcal{C}})} g(\boldsymbol{x}_{\mathcal{C}} + m\boldsymbol{e}^i_{\mathcal{C}})\right].
\end{aligned}$$

□

From this lemma with $m = -1$ and $g(\boldsymbol{x}_{\mathcal{C}}) = 1$, $\hat{\boldsymbol{\lambda}}^{ML}(\boldsymbol{x})$ is found to be the uniformly minimum variance unbiased estimator of $\boldsymbol{\lambda}$. We note that (4) with $m = 1$ is expressed by

$$\begin{aligned}
\mathrm{E}\left[\frac{1}{\lambda(i)} g(\boldsymbol{x}_{\mathcal{C}})\right] &= \mathrm{E}\left[\frac{t(\boldsymbol{x}_{\mathcal{C}} + \boldsymbol{e}^i_{\mathcal{C}})}{t(\boldsymbol{x}_{\mathcal{C}})} g(\boldsymbol{x}_{\mathcal{C}} + \boldsymbol{e}^i_{\mathcal{C}})\right] \\
&= \mathrm{E}\left[\frac{1}{\hat{\lambda}^{ML}(i, \boldsymbol{x}_{\mathcal{C}} + \boldsymbol{e}^i_{\mathcal{C}})} g(\boldsymbol{x}_{\mathcal{C}} + \boldsymbol{e}^i_{\mathcal{C}})\right]. \qquad (5)
\end{aligned}$$

Let the bijection $\sigma : \{1, 2, \ldots, J\} \mapsto \Delta$ be a perfect elimination scheme (e.g. Blair and Peyton[1]) of vertices in $\mathcal{G}$. For $1 \leq j \leq J$ define $\Delta(\sigma, j) \subset \Delta$ as

$$\Delta(\sigma, j) = \{\sigma(j), \sigma(j+1), \ldots, \sigma(J)\}.$$

The following lemmas are required to derive the class of improved estimators.

Lemma 2.2. *Suppose* $x(i_S) \neq 0$ *for all* $S \in \mathcal{S}$. *Then*

$$\begin{aligned}
\hat{\lambda}^{ML}(i_{\Delta(\sigma,j)}, \boldsymbol{x}_{\mathcal{C}}) &= \sum_{i' : i'_{\Delta(\sigma,j)} = i_{\Delta(\sigma,j)}} \hat{\lambda}^{ML}(i', \boldsymbol{x}_{\mathcal{C}}) \\
&= \frac{\prod_{C\in\mathcal{C}(\Delta(\sigma,j))} x(i_C)}{\prod_{S\in\mathcal{S}(\Delta(\sigma,j))} x(i_S)^{\nu(S,\Delta(\sigma,j))}}.
\end{aligned}$$

Lemma 2.3. *Suppose $x(i_S) \neq 0$ for all $S \in \mathcal{S}$. Then*

$$\sum_{i':i'_{\Delta(\sigma,j)}=i_{\Delta(\sigma,j)}} \hat{\lambda}^{ML}(i', \boldsymbol{x}_{\mathcal{C}} + \boldsymbol{e}_{\mathcal{C}}^{i'}) \geq \frac{\prod_{C\in\mathcal{C}(\Delta(\sigma,j))}(x(i_C)+1)}{\prod_{S\in\mathcal{S}(\Delta(\sigma,j))}(x(i_S)+1)^{\nu(S,\Delta(\sigma,j))}}$$
$$= \hat{\lambda}^{ML}(i_{\Delta(\sigma,j)}, \boldsymbol{x}_{\mathcal{C}} + \boldsymbol{e}_{\mathcal{C}}^{i}).$$

The proofs of the Lemma 2.2 and 2.3 are given in the Appendix.

2.3 Some preparations on the decomposable graphs

In this section we prepare some lemmas on the decomposable graphs required for the argument in the following section.

Suppose $|\mathcal{C}| = K$. Let $C_1, C_2, \ldots, C_K$ be a perfect sequence of the cliques in $\mathcal{G}$. We write $H_k = C_1 \cup \cdots \cup C_k$, $k = 1, \ldots, K$ and $S_k = H_{k-1} \cap C_k$, $k = 2, \ldots, K$. Then $\mathcal{S} = \{S_2, \ldots, S_K\}$, where each $S \in \mathcal{S}$ is repeated $\nu(S)$ times. For any S_k there exists $k' < k$ such that $S_k \subset C_{k'}$. This condition is known as the running intersection property of the perfect sequence.

Let $\mathcal{C}_S = \{C_i \in \mathcal{C} \mid C_i \supset S\} = \{C_{k_1}, \ldots, C_{k_q}\}$, $1 \leq k_1 < \cdots < k_q \leq K$, be the set of cliques which includes S. We note that $C_{k_1}, \ldots, C_{k_q}$ is a subsequence of the perfect sequence. Then we obtain the following lemma.

Proposition 2.1. *S decomposes $\mathcal{G}(\Delta(\mathcal{C}_S))$ into $\nu(S)+1$ connected components.*

Darroch, Lauritzen and Speed[5] and Letac and Massam[16] state this proposition but the proof is not given. Since this proposition is essential for the present paper, we give the proof of Proposition 2.1 in a series of lemmas.

Lemma 2.4. *For a perfect sequence $C_1, C_2, \ldots, C_K$ and $k \leq K$, define the set of cliques $\mathcal{C}_k$ by $\mathcal{C}_k = \{C_1, C_2, \ldots, C_k\}$. Then $C_1, C_2, \ldots, C_k$ is a perfect sequence of $\mathcal{G}(\Delta(\mathcal{C}_k))$.*

Proof. When $k = K$, the lemma is trivial. Assume $k < K$. Since $C_1, C_2, \ldots, C_k$ satisfies the running intersection property, it suffices to show that $\mathcal{C}(\Delta(\mathcal{C}_k)) = \mathcal{C}_k$. Suppose $\mathcal{C}(\Delta(\mathcal{C}_k)) \neq \mathcal{C}_k$. Since $\mathcal{C}(\Delta(\mathcal{C}_k)) \supset \mathcal{C}_k$, there exists a clique A such that $A \in \mathcal{C}(\Delta(\mathcal{C}_k)) \setminus \mathcal{C}_k$. From the fact that $\Delta(\mathcal{C}(\Delta(\mathcal{C}_k))) = \Delta(\mathcal{C}_k)$ and the maximality of cliques, A satisfies $A \subset \Delta(\mathcal{C}_k)$ and $A \not\subseteq C_j$ for $j = 1, \ldots, k$ and there exists $S_{k'}$, $k' > k$ such that $S_{k'} \supset A$ and $S_j \not\supseteq A$ for $j < k'$. Thus $S_{k'}$ satisfies $S_{k'} \not\subseteq C_j$ for $j < k'$. This contradicts the running intersection property of $C_1, C_2, \ldots, C_K$. □

Lemma 2.5. *$C_{k_1}, \ldots, C_{k_q}$ is a perfect sequence of $\mathcal{G}(\Delta(\mathcal{C}_S))$.*

Proof. First we show that $\mathcal{C}_S = \mathcal{C}(\Delta(\mathcal{C}_S))$. Suppose $\mathcal{C}_S \neq \mathcal{C}(\Delta(\mathcal{C}_S))$. Then there exists a clique $C \in \mathcal{C}(\Delta(\mathcal{C}_S)) \setminus \mathcal{C}_S$ such that $C \subset \Delta(\mathcal{C}_S) = \Delta(\mathcal{C}(\Delta(\mathcal{C}_S)))$ and $C \not\supseteq S$. However any vertex in $\Delta(\mathcal{C}_S) \setminus S$ is adjacent to all vertices in S from the completeness of cliques. This implies $C \supset S$ for all $C \in \mathcal{C}(\Delta(\mathcal{C}_S))$, which contradicts the assumption that $\mathcal{C}_S \neq \mathcal{C}(\Delta(\mathcal{C}_S))$. Thus $\mathcal{C}_S = \mathcal{C}(\Delta(\mathcal{C}_S))$.

Next we show that $C_{k_1}, \ldots, C_{k_q}$ is perfect. Since $C_1, \ldots, C_K$ is perfect, for each $l \geq 2$ there exists a $k < k_l$ such that $S_{k_l} \subset C_k \in \mathcal{C}$. Since $C_{k_l} \supset S$, we have

$$C_k \supset S_{k_l} = \bigcup_{j=1}^{k_l-1} (C_j \cap C_{k_l}) \supseteq (C_{k_1} \cap C_{k_l}) \supseteq S,$$

which implies $C_k \in \mathcal{C}_S$. Thus the subsequence $C_{k_1}, \ldots, C_{k_q}$ also satisfies the running intersection property. This proves the lemma. $\square$

Lemma 2.6. $\mathcal{S}(\Delta(\mathcal{C}_S)) = \{S_{k_l} : l \geq 2\}$ *and the multiplicity of* S *in* $\mathcal{G}(\Delta(\mathcal{C}_S))$ *is* $\nu(S)$.

Proof. From Lemma 2.5

$$S_{k_l} = \bigcup_{j=1}^{k_l-1} (C_j \cap C_{k_l}) = \bigcup_{m=1}^{l-1} (C_{k_m} \cap C_{k_l})$$

for $1 \leq l \leq q$. Thus the proof is completed. $\square$

From Lemma 2.5 and 2.6 there exists the sequence of indices $\tilde{k}_0, \tilde{k}_1, \ldots, \tilde{k}_{\nu(S)}$ such that $\tilde{k}_0 < \cdots < \tilde{k}_{\nu(S)}$, $\tilde{k}_0 = k_1$ and $S_{\tilde{k}_m} = S$ for $m = 1, \ldots, \nu(S)$. Define $\mathcal{C}_S^l$ and ν_l, $l = 1, \ldots, q$, by

$$\mathcal{C}_S^l = \{C_{k_0}, C_{k_1}, \ldots, C_{k_l}\}, \quad \nu_l = \max\{m | \tilde{k}_m \leq k_l\},$$

respectively. From Lemma 2.4 and 2.6 $C_{k_1}, \ldots, C_{k_l}$ is a perfect sequence of $\mathcal{G}(\Delta(\mathcal{C}_S^l))$ and ν_l is the multiplicity of S in $\mathcal{G}(\Delta(\mathcal{C}_S^l))$. Proposition 2.1 is obtained from the following lemma with $l = q$.

Lemma 2.7. S *decomposes* $\mathcal{G}(\Delta(\mathcal{C}_S^l))$ *into* $\nu_l + 1$ *connected components for* $l = 1, \ldots, q$.

Proof. We prove this lemma by induction on l. For $l = 1$ the lemma is trivial. Suppose that $p \geq 2$ and that the lemma holds for $l \leq p$. Then there exists $\nu_p + 1$ connected components of $\Delta(\mathcal{C}_S^p) \setminus S$. Denote them by $\Gamma_1^S, \ldots, \Gamma_{\nu_p+1}^S$.

If $S_{k_{p+1}} = S$, then $\nu_{p+1} = \nu_p + 1$ and $C_{k_{l'}} \cap C_{k_{p+1}} = S$ for all l', $1 \leq l' \leq p$. This implies $\Gamma_m^S \cap (C_{k_{p+1}} \setminus S) = \emptyset$ for all $1 \leq m \leq \nu_p + 1$. Thus S decomposes $\Delta(\mathcal{C}_S^{p+1})$ into the following $\nu_{p+1} + 1$ connected components, $\Gamma_1^S, \Gamma_2^S, \ldots, \Gamma_{\nu_p+1}^S$ and $\Gamma_{\nu_{p+1}+1}^S = C_{k_{p+1}} \setminus S$.

In the case where $S_{k_{p+1}} \neq S$, ν_{p+1} is equal to ν_p. Since $S_{k_{p+1}} \supset S$, there exists l', $1 \leq l' \leq p$ such that

$$S_{k_{p+1}} \subset C_{k_{l'}} \in \mathcal{C}_S, \quad (C_{k_{p+1}} \setminus S) \cap (C_{k_{l'}} \setminus S) \neq \emptyset. \tag{6}$$

Then $C_{k_{p+1}} \setminus S$ and $C_{k_{l'}} \setminus S$ are connected. If there are two cliques $C_{k_{p_1}}$ and $C_{k_{p_2}}$ satisfying (6), it is necessary for them to satisfy that

$$(C_{k_{p_1}} \setminus S) \cap (C_{k_{p_2}} \setminus S) \supset (S_{k_{l+1}} \setminus S) \neq \emptyset.$$

Thus $C_{k_{p_1}} \setminus S$ and $C_{k_{p_2}} \setminus S$ are connected and they belong to the same connected component. Let Γ_1^S be the connected component. Then $C_{k_{p+1}}$ satisfies

$$C_{k_{p+1}} \cap \Gamma_1^S \neq \emptyset \quad \text{and} \quad C_{k_{p+1}} \cap \Gamma_m^S = \emptyset$$

for $2 \leq m \leq \nu_p + 1$. Hence S decomposes $\Delta(\mathcal{C}_S^p)$ into $\Gamma_1^S \cup (C_{k_{p+1}} \setminus S)$, $\Gamma_2^S, \ldots, \Gamma_{\nu_p+1}^S = \Gamma_{\nu_{p+1}+1}^S$. This completes the proof. $\square$

We have completed the proof of Proposition 2.1. We present two additional lemmas needed later.

Lemma 2.8. *For any $C \in \mathcal{C}$, there exists a perfect elimination scheme σ and j, $1 \le j \le J$, such that*

$$\Delta(\sigma, j) = \{\sigma(j), \ldots, \sigma(J)\} = C. \tag{7}$$

Proof. The proof is by induction on the number of vertices J. The lemma is trivial if $J \le 3$. Suppose $J > 3$ and assume that the lemma holds for all decomposable graphs with fewer than J vertices.

Since $\mathcal{G}$ is decomposable and is not complete, $\mathcal{G}$ has at least two non-adjacent simplicial vertices(see Dirac[6]). Thus there exists a simplicial vertex δ such that $\delta \notin C$. $\mathcal{C}(\Delta \setminus \{\delta\})$ includes C and then from the inductive assumption $\mathcal{G}(\Delta \setminus \{\delta\})$ has a perfect elimination scheme $\sigma_{\Delta \setminus \{\delta\}} : \{1, 2, \ldots, J-1\} \mapsto \Delta \setminus \{\delta\}$ and for some j, $2 \le j \le J$,

$$\{\sigma_{\Delta \setminus \{\delta\}}(j-1), \ldots, \sigma_{\Delta \setminus \{\delta\}}(J-1)\} = C.$$

If we set $\sigma(1) = \delta$ and $\sigma(l) = \sigma_{\Delta \setminus \{\delta\}}(l-1)$, $l = 2, \ldots, J$, then σ satisfies (7). The proof is completed. □

From Proposition 2.1 $S \in \mathcal{S}$ decomposes $\Delta(\mathcal{C}_S)$ into $\nu(S)+1$ connected components. Denote them by $\Gamma_1^S, \Gamma_2^S, \ldots, \Gamma_{\nu(S)+1}^S$. Let $\mathbb{C}(S, \mathcal{G})$ be the class of $\nu(S)+1$ cliques in $\mathcal{C}_S$ such that

$$\mathbb{C}(S, \mathcal{G}) = \Big\{ \{C_1, \ldots, C_{\nu(S)+1}\} \mid C_1 \in \mathcal{C}(\Gamma_1^S \cup S), \ldots, C_{\nu(S)+1} \in \mathcal{C}(\Gamma_{\nu(S)+1}^S \cup S) \Big\}. \tag{8}$$

We note that the cliques in $\{C_1, \ldots, C_{\nu(S)+1}\} \in \mathbb{C}(S, \mathcal{G})$ satisfy $C_j \in \mathcal{C}$ for $j = 1, \ldots, \nu(S)+1$ and $C_j \cap C_k = S$ for $j \ne k$. Then we obtain the following lemma.

Lemma 2.9. *For any $S \in \mathcal{S}$ and $\tilde{\mathcal{C}} \in \mathbb{C}(S, \mathcal{G})$, there exists a perfect elimination scheme σ and j, $1 \le j \le J$, such that*

$$\mathcal{G}(\Delta(\tilde{C})) = \mathcal{G}(\Delta(\sigma, j)). \tag{9}$$

We give the proof of this lemma in the Appendix.

3 Improved estimation of the means in the decomposable models for connected graphs

In this section we give some classes of estimators improving on $\boldsymbol{\lambda}^{ML}$ under the loss function (2). We introduce the following class of Chou[3]-type estimators,

$$\hat{\boldsymbol{\lambda}}^{\phi_V, \gamma} = \{\hat{\lambda}^{\phi_V, \gamma}(i)\}_{i \in \mathcal{I}},$$

$$\hat{\lambda}^{\phi_V, \gamma}(i) = \hat{\lambda}^{ML}(i, \boldsymbol{x}_{\mathcal{C}}) \left(1 - \frac{\phi_V(x(i_V))}{(x(i_V)+\beta)^{\gamma+1}}\right), \tag{10}$$

where V is a set of vertices such that $\mathcal{G}(V)$ is complete, i.e., $V \subset C$ for some $C \in \mathcal{C}$. $\beta > 0$ and $\gamma \ge 0$ are the constants. Suppose that $\phi_V(x)$ is nondecreasing and satisfies

$$0 \le \frac{\phi_V(x)}{(x+\beta)^{\gamma+1}} \le 1 \tag{11}$$

for all nonnegative integer x.

By using (5), the difference between two risk function of $\hat{\boldsymbol{\lambda}}^{ML}$ and $\hat{\boldsymbol{\lambda}}^{\phi_V,\gamma}$ is expressed by

$$
\begin{aligned}
&\mathrm{R}(\boldsymbol{\lambda}, \hat{\boldsymbol{\lambda}}^{ML}) - \mathrm{R}(\boldsymbol{\lambda}, \hat{\boldsymbol{\lambda}}^{\phi_V,\gamma}) \\
&\quad = \mathrm{E}[\mathrm{L}(\boldsymbol{\lambda}, \hat{\boldsymbol{\lambda}}^{ML}) - \mathrm{L}(\boldsymbol{\lambda}, \hat{\boldsymbol{\lambda}}^{\phi_V,\gamma})] \\
&\quad = \sum_{i \in \mathcal{I}} \mathrm{E}\left[\frac{1}{\lambda(i)} \left\{ (\hat{\lambda}^{ML}(i, \boldsymbol{x}_{\mathcal{C}}) - \lambda(i))^2 - (\hat{\lambda}^{\phi_V,\gamma}(i) - \lambda(i))^2 \right\} \right] \\
&\quad = \sum_{i \in \mathcal{I}} \mathrm{E}\left[\frac{2}{\lambda(i)} \hat{\lambda}^{ML}(i, \boldsymbol{x}_{\mathcal{C}}) \frac{\phi_V(x(i_V))}{(x(i_V)+\beta)^{\gamma+1}} (\hat{\lambda}^{ML}(i, \boldsymbol{x}_{\mathcal{C}}) - \lambda(i)) \right. \\
&\qquad\qquad\qquad \left. - \frac{1}{\lambda(i)} \left\{ \hat{\lambda}^{ML}(i, \boldsymbol{x}_{\mathcal{C}}) \frac{\phi_V(x(i_V))}{(x(i_V)+\beta)^{\gamma+1}} \right\}^2 \right] \\
&\quad = \sum_{i \in \mathcal{I}} \mathrm{E}\left[\frac{2\phi_V(x(i_V)+1)}{(x(i_V)+\beta+1)^{\gamma+1}} \hat{\lambda}^{ML}(i, \boldsymbol{x}_{\mathcal{C}} + \boldsymbol{e}^i_{\mathcal{C}}) - \frac{2\phi_V(x(i_V))}{(x(i_V)+\beta)^{\gamma+1}} \hat{\lambda}^{ML}(i, \boldsymbol{x}_{\mathcal{C}}) \right. \\
&\qquad\qquad\qquad \left. - \left\{ \frac{\phi_V(x(i_V)+1)}{(x(i_V)+\beta+1)^{\gamma+1}} \right\}^2 \hat{\lambda}^{ML}(i, \boldsymbol{x}_{\mathcal{C}} + \boldsymbol{e}^i_{\mathcal{C}}) \right] \\
&\quad = \sum_{i_V \in \mathcal{I}_V} \mathrm{E}\left[\frac{2\phi_V(x(i_V)+1)}{(x(i_V)+\beta+1)^{\gamma+1}} \hat{\lambda}^{ML}(i_V, \boldsymbol{x}_{\mathcal{C}} + \boldsymbol{e}^i_{\mathcal{C}}) - \frac{2\phi_V(x(i_V))}{(x(i_V)+\beta)^{\gamma+1}} \hat{\lambda}^{ML}(i_V, \boldsymbol{x}_{\mathcal{C}}) \right. \\
&\qquad\qquad\qquad \left. - \left\{ \frac{\phi_V(x(i_V)+1)}{(x(i_V)+\beta+1)^{\gamma+1}} \right\}^2 \hat{\lambda}^{ML}(i_V, \boldsymbol{x}_{\mathcal{C}} + \boldsymbol{e}^i_{\mathcal{C}}) \right], \qquad (12)
\end{aligned}
$$

where

$$
\hat{\lambda}^{ML}(i_V) = \sum_{i' : i'_V = i_V} \hat{\lambda}^{ML}(i').
$$

(12) implies that

$$
\hat{\mathrm{Rd}}(\hat{\boldsymbol{\lambda}}^{\phi_V,\gamma}) = \sum_{i_V \in \mathcal{I}_V} \hat{\mathrm{Rd}}(i_V, \hat{\boldsymbol{\lambda}}^{\phi_V,\gamma}),
$$

where

$$
\begin{aligned}
\hat{\mathrm{Rd}}(i_V, \hat{\boldsymbol{\lambda}}^{\phi_V,\gamma}) &= 2\left\{ \frac{\phi_V(x(i_V)+1)}{(x(i_V)+\beta+1)^{\gamma+1}} \hat{\lambda}^{ML}(i_V, \boldsymbol{x}_{\mathcal{C}} + \boldsymbol{e}^i_{\mathcal{C}}) - \frac{\phi_V(x(i_V))}{(x(i_V)+\beta)^{\gamma+1}} \hat{\lambda}^{ML}(i_V, \boldsymbol{x}_{\mathcal{C}}) \right\} \\
&\quad - \left(\frac{\phi_V(x(i_V)+1)}{(x(i_V)+\beta+1)^{\gamma+1}} \right)^2 \hat{\lambda}^{ML}(i_V, \boldsymbol{x}_{\mathcal{C}} + \boldsymbol{e}^i_{\mathcal{C}}) \qquad (13)
\end{aligned}
$$

is an unbiased estimator of $\mathrm{R}(\boldsymbol{\lambda}, \hat{\boldsymbol{\lambda}}^{ML}) - \mathrm{R}(\boldsymbol{\lambda}, \hat{\boldsymbol{\lambda}}^{\phi_V,\gamma})$. Thus in order to examine the dominance of $\hat{\boldsymbol{\lambda}}^{\phi_V,\gamma}$ over $\hat{\boldsymbol{\lambda}}^{ML}$, it suffices to show that $\phi_V(\cdot)$ satisfies $\hat{\mathrm{Rd}}(i_V, \hat{\boldsymbol{\lambda}}^{\phi_V,\gamma}) \geq 0$.

Let $\mathcal{C}_V$ be the set of the cliques which include V. Define $I_{\mathcal{C}_V}$ by

$$
I_{\mathcal{C}_V} = \max_{C \in \mathcal{C}_V} I_{C \setminus V}.
$$

Then we obtain the following theorem.

Theorem 3.1. *If $\phi_V(\cdot)$ satisfies*

$$0 \le \phi_V(x) \le \min\Big(\ 2I_{\mathcal{C}_V} - 2\gamma - 2, \ \ 2\beta - 2\gamma, \ \ (x+\beta)^{\gamma+1} \ \Big) \tag{14}$$

for all nonnegative integer x, $\hat{\boldsymbol{\lambda}}^{\phi_V,\gamma}$ dominates $\hat{\boldsymbol{\lambda}}^{ML}$ under the loss function (2).

Proof. By applying Lemma 2.2, 2.3 with $\Delta(\sigma, j) = C \in \mathcal{C}_V$ and Lemma 2.8, we have

$$\begin{aligned}
\hat{\lambda}^{ML}(i_V, \boldsymbol{x}_{\mathcal{C}}) &= x(i_V), \\
\hat{\lambda}^{ML}(i_V, \boldsymbol{x}_{\mathcal{C}} + \boldsymbol{e}^i_{\mathcal{C}}) &\ge \sum_{i'_C : i'_V = i_V} (x(i_C) + 1) \\
&= x(i_V) + I_{C \setminus V}
\end{aligned}$$

for all $C \in \mathcal{C}_V$. Thus from the definition of $I_{\mathcal{C}_V}$

$$\hat{\lambda}^{ML}(i_V, \boldsymbol{x}_{\mathcal{C}} + \boldsymbol{e}^i_{\mathcal{C}}) \ge x(i_V) + I_{\mathcal{C}_V}.$$

Define B as $B = x(i_V) + \beta + 1$. Then from (13)

$$\begin{aligned}
&\hat{\mathrm{Rd}}(i_V, \hat{\boldsymbol{\lambda}}^{\phi_V,\gamma}) \\
&= \frac{\phi_V(x(i_V)+1)}{B^{\gamma+2}} \left(2B - \frac{\phi_V(x(i_V)+1)}{B^\gamma} \right) \hat{\lambda}^{ML}(i_V, \boldsymbol{x}_{\mathcal{C}} + \boldsymbol{e}^i_{\mathcal{C}}) - \frac{\phi_V(x(i_V))}{(B-1)^{\gamma+1}} \hat{\lambda}^{ML}(i_V, \boldsymbol{x}_{\mathcal{C}}) \\
&\ge \frac{\phi_V(x(i_V)+1)}{B^{\gamma+2}(B-1)^{\gamma+1}} \Bigg\{ (B-1)^{\gamma+1} \left(2B - \frac{\phi_V(x(i_V)+1)}{B^\gamma} \right) (x(i_V) + I_{\mathcal{C}_V}) \\
&\qquad\qquad - 2B^{\gamma+2} x(i_V) \Bigg\} \\
&\ge \frac{\phi_V(x(i_V)+1)}{B(B-1)^{\gamma+1}} \Big\{ \Big(2(B - (\gamma+1)) - \phi_V(x(i_V)+1) \Big) (x(i_V) + I_{\mathcal{C}_V}) - 2Bx(i_V) \Big\} \\
&= \frac{\phi_V(x(i_V)+1)}{B(B-1)^{\gamma+1}} \Big\{ \Big(2I_{\mathcal{C}_V} - 2\gamma - 2 - \phi_V(x(i_V)+1) \Big) x(i_V) \\
&\qquad\qquad + \Big(2\beta - 2\gamma - \phi_V(x(i_V)) \Big) I_{\mathcal{C}_V} \Big\}.
\end{aligned} \tag{15}$$

The second inequality follows from the fact that $(B-1)^{\gamma+1} \ge B^{\gamma+1} - (\gamma+1)B^\gamma$ for $\gamma \ge 0$ and the assumption (11). The right hand side of (15) is always nonnegative under the condition (14), which completes the proof. □

So far we considered any $V \subset \Delta$ such that V is a proper subset of some clique in $\mathcal{G}$. We can obtain wider conditions on ϕ_V to dominate the MLE, if V is a subset of a minimal vertex separator. Consider any of the following three condition,

(i) V is a minimal vertex separator, i.e. $V \in \mathcal{S}$

(ii) There exists a minimal vertex separator S such that $V \subset S$

(iii) $V = \emptyset$, i.e. $x(i_V) = x^+$.

For each of these cases we derive wider class of estimators, which dominate the MLE.

We begin with the case (i) where $V \in \mathcal{S}$. Define $I_{\mathbb{C}(V,\mathcal{G})}$ by

$$I_{\mathbb{C}(V,\mathcal{G})} = \max_{\tilde{\mathcal{C}} \in \mathbb{C}(V,\mathcal{G})} \sum_{C \in \tilde{\mathcal{C}}} I_{C \setminus V}, \tag{16}$$

where $\mathbb{C}(V,\mathcal{G})$ is given in (8). Then we obtain the following theorem.

Theorem 3.2. *Suppose V is a minimal vertex separator in $\mathcal{G}$. If $\phi_V(\cdot)$ satisfies*

$$0 \le \phi_V(x) \le \min\Big(\; 2(I_{\mathbb{C}(V,\mathcal{G})} - \nu(V)) - 2\gamma - 2, \;\; 2\beta - 2\gamma, \;\; (x+\beta)^{\gamma+1} \; \Big) \tag{17}$$

for all nonnegative integer x, then $\hat{\boldsymbol{\lambda}}^{\phi_V,\gamma}$ dominates $\hat{\boldsymbol{\lambda}}^{ML}$ under the loss function (2).

Proof. By using Lemma 2.2, 2.3 with $\Delta(\sigma, j) = \Delta(\tilde{\mathcal{C}})$ and Lemma 2.9, we have

$$\begin{aligned}
\hat{\lambda}^{ML}(i_V, \boldsymbol{x}_{\mathcal{C}}) &= x(i_V), \\
\hat{\lambda}^{ML}(i_V, \boldsymbol{x}_C + \boldsymbol{e}^i_{\mathcal{C}}) &\ge \sum_{C\in\tilde{\mathcal{C}}} \sum_{i'_C : i'_V = i_V} \frac{\prod_{C\in\tilde{\mathcal{C}}}(x(i'_C)+1)}{(x(i_V)+1)^{\nu(V)}} \\
&= \frac{\prod_{C\in\tilde{\mathcal{C}}}(x(i_V)+I_{C\setminus V})}{(x(i_V)+1)^{\nu(V)}} \\
&\ge x(i_V) + \sum_{C\in\tilde{\mathcal{C}}} I_{C\setminus V} - \nu(V)
\end{aligned}$$

for all $\tilde{C} \in \mathbb{C}(V,\mathcal{G})$. The last inequality follows from Lemma 3.2 in Hara and Takemura[10]. Thus from the definition of $I_{\mathbb{C}(V,\mathcal{G})}$

$$\hat{\lambda}^{ML}(i_V, \boldsymbol{x}_C + \boldsymbol{e}^i_{\mathcal{C}}) \ge x(i_V) + I_{\mathbb{C}(V,\mathcal{G})} - \nu(V). \tag{18}$$

Define B as $B = x(i_V) + \beta + 1$. In the same way as (15)

$$\begin{aligned}
&\hat{\mathrm{Rd}}(i_V, \hat{\boldsymbol{\lambda}}^{\phi_V,\gamma}) \\
&= \frac{\phi_V(x(i_V)+1)}{B^{\gamma+2}} \left(2B - \frac{\phi_V(x(i_V)+1)}{B^\gamma}\right) \hat{\lambda}^{ML}(i_V, \boldsymbol{x}_{\mathcal{C}} + \boldsymbol{e}^i_{\mathcal{C}}) - \frac{\phi_V(x(i_V))}{(B-1)^{\gamma+1}} \hat{\lambda}^{ML}(i_V, \boldsymbol{x}_{\mathcal{C}}) \\
&\ge \frac{\phi_V(x(i_V)+1)}{B^{\gamma+2}(B-1)^{\gamma+1}} \left\{ (B-1)^{\gamma+1} \left(2B - \frac{\phi_V(x(i_V)+1)}{B^\gamma}\right)(x(i_V) + I_{\mathbb{C}(V,\mathcal{G})} - \nu(V)) \right. \\
&\qquad\qquad \left. -2B^{\gamma+2} x(i_V) \right\} \\
&\ge \frac{\phi_V(x(i_V)+1)}{B(B-1)^{\gamma+1}} \Big\{ \Big(2(B-(\gamma+1)) - \phi_V(x(i_V)+1)\Big)(x(i_V) + I_{\mathbb{C}(V,\mathcal{G})} - \nu(V)) - 2Bx(i_V) \Big\} \\
&= \frac{\phi_V(x(i_V)+1)}{B(B-1)^{\gamma+1}} \Big\{ \Big(2(I_{\mathbb{C}(V,\mathcal{G})} - \nu(V)) - 2\gamma - 2 - \phi_V(x(i_V)+1)\Big) x(i_V) \\
&\qquad\qquad + \Big(2\beta - 2\gamma - \phi_V(x(i_V))\Big)(I_{\mathbb{C}(V,\mathcal{G})} - \nu(V)) \Big\}.
\end{aligned} \tag{19}$$

The right hand side of (19) is always nonnegative under the condition (17), which completes the proof. □

Remark 3.1. We note that $|\tilde{\mathcal{C}}| = \nu(V) + 1$ for all $\tilde{\mathcal{C}} \in \mathbb{C}(V, \mathcal{G})$ and $V \in \mathcal{S}$ from (8). Since we assumed that $I_\delta \geq 2$ for all $\delta \in \Delta$, we have

$$\begin{aligned} I_{\mathbb{C}(V,\mathcal{G})} - \nu(V) &\geq \sum_{C \in \tilde{\mathcal{C}}} I_{C\setminus V} - \nu(V) \\ &= I_{\tilde{C}\setminus V} + \sum_{C \neq \tilde{C}} (I_{C\setminus V} - 1) \\ &> I_{\tilde{C}\setminus S} \end{aligned}$$

for all $\tilde{C} \in \tilde{\mathcal{C}}$ and $\tilde{\mathcal{C}} \in \mathbb{C}$. This implies $I_{\mathbb{C}(V,\mathcal{G})} - \nu(V) \geq I_{\mathcal{C}_V}$. Thus the class (17) is wider than (14).

Next we consider the case (ii) where there exists at least one minimal vertex separator $S \in \mathcal{S}$ such that $V \subset S$. We note that such V may itself be a minimal vertex separator.

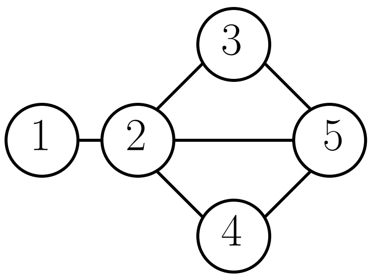

Figure 3: 5-way decomposable model (a)

For example $\mathcal{S}$ of the graph in Fig 3 is $\{\{2\}, \{2, 5\}\}$. $\{2\}$ is itself a minimal vertex separator and is also a subset of $\{2, 5\}$.

Define $\mathcal{S}_V$ as follows,

$$\mathcal{S}_V = \{S \in \mathcal{S} \mid S \supseteq V\}.$$

Then we obtain the following theorem.

Theorem 3.3. *Suppose $V \subset S$ for some $S \in \mathcal{S}$. Define I_V^* by*

$$I_V^* = \max_{S \in \mathcal{S}_V} I_{S\setminus V}(I_{\mathbb{C}(S,\mathcal{G})} - \nu(S)), \tag{20}$$

where $I_\emptyset = I_{V\setminus V} \equiv 1$. If $\phi_V(\cdot)$ satisfies

$$0 \leq \phi_V(x) \leq \min\Big(\ 2I_V^* - 2\gamma - 2,\ \ 2\beta - 2\gamma,\ \ (x+\beta)^{\gamma+1}\ \Big) \tag{21}$$

for all nonnegative integer x, $\hat{\boldsymbol{\lambda}}^{\phi_V,\gamma}$ dominates $\hat{\boldsymbol{\lambda}}^{ML}$ under the loss function (2).

PROOF. Define S_V^* by

$$S_V^* = \operatorname*{argmax}_{S \in \mathcal{S}_V} I_{S\setminus V}(I_{\mathbb{C}(S,\mathcal{G})} - \nu(S)).$$

From (18) we have

$$\begin{aligned}\hat{\lambda}^{ML}(i_V, \boldsymbol{x}_C + \boldsymbol{e}^i_{\mathcal{C}}) &\geq \sum_{i'_{S^*_V} : i'_V = i_V} \left\{ x(i_{S^*_V}) + (I_{\mathbb{C}(S^*_V, \mathcal{G})} - \nu(S^*_V)) \right\} \\ &= x(i_V) + I_{S^*_V \setminus V}(I_{\mathbb{C}(S^*_V, \mathcal{G})} - \nu(S^*_V)).\end{aligned}$$

Let B be $B = x(i_V) + \beta + 1$. In the same way as (19)

$$\begin{aligned}\hat{\mathrm{Rd}}(i_V, \hat{\boldsymbol{\lambda}}^{\phi_V, \gamma}) &\geq \frac{\phi_V(x(i_V)+1)}{B(B-1)^{\gamma+1}} \Big\{ \Big(2 I_{S^*_V \setminus V}(I_{\mathbb{C}(S^*_V)} - \nu(S^*_V)) - 2\gamma - 2 - \phi_V(x(i_V)+1) \Big) x(i_V) \\ &\qquad + \Big(2\beta - 2\gamma - \phi_V(x(i_V)) \Big) I_{S^*_V \setminus V}(I_{\mathbb{C}(S^*_V)} - \nu(S^*_V)) \Big\}.\end{aligned}$$

When ϕ_V satisfies (21), $\hat{\mathrm{Rd}}(i_V, \hat{\boldsymbol{\lambda}}^{\phi_V, \gamma}) \geq 0$, which completes the proof. □

Remark 3.2. We note that

$$\begin{aligned}\max_{C \in \mathcal{C}_V} I_{C \setminus V} &= \max_{S \in \mathcal{S}_V} \max_{C \in \mathcal{C}_V} I_{C \setminus S} I_{S \setminus V} \\ &\leq \max_{S \in \mathcal{S}_V} I_{S \setminus V}(I_{\mathbb{C}(S, \mathcal{G})} - \nu(S)) \\ &= I^*_V.\end{aligned}$$

The inequality follows from the argument in Remark 3.1. Thus the class (21) is wider than (14) in Theorem 3.1. Even in the case where $V \in \mathcal{S}$,

$$\begin{aligned}I^*_V &= \max_{S \in \mathcal{S}_V} I_{S \setminus V}(I_{\mathbb{C}(S, \mathcal{G})} - \nu(S)) \\ &\geq I_{S \setminus V}(I_{\mathbb{C}(S, \mathcal{G})} - \nu(S))\end{aligned}$$

from the fact that $V \in \mathcal{S}_V$, that is, the class (21) is wider than (17). Thus we may as well apply Theorem 3.3 to such V. If $V \in \mathcal{S}$ and there exist no minimal vertex separators $S \in \mathcal{S}$ such that $V \subset S$, Theorem 3.2 should be applied.

Define I^* as

$$I^* = \max_{S \in \mathcal{S}} I_S(I_{\mathbb{C}(S, \mathcal{G})} - \nu(S)). \tag{22}$$

From the result of Theorem 3.3 with $V = \emptyset$, we can obtain the following result.

Theorem 3.4. *If $\phi(\cdot)$ satisfies*

$$0 \leq \phi(x) \leq \min\Big(\; 2I^* - 2\gamma - 2, \;\; 2\beta - 2\gamma, \;\; (x+\beta)^{\gamma+1} \; \Big), \tag{23}$$

then

$$\hat{\boldsymbol{\lambda}}^{\phi, \gamma} = \hat{\boldsymbol{\lambda}}^{ML} \left(1 - \frac{\phi(x^+)}{(x^+ + \beta)^{\gamma+1}} \right)$$

dominates $\hat{\boldsymbol{\lambda}}^{ML}$ under the loss function (2).

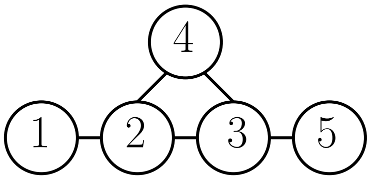

Figure 4: 5-way decomposable model (b)

From Theorem 3.2 to 3.4 we obtained wider classes of improved estimators than the class of Theorem 3.1 corresponding to the case (i) to (iii), respectively. Consider the following other two cases,

(iv) $V \notin \mathcal{S}$ is a union of some minimal vertex separators.

(v) V includes a simplicial vertex.

We give an example of the case (iv). $\mathcal{C}$ and $\mathcal{S}$ of the graph in Figure 4 are {{1,2},{2,3,4},{3,5}} and {{2}, {3}}, respectively. If we set $V = \{2, 3\}$, then V is a subset of the clique $\{2, 3, 4\}$ and is the union of the minimal vertex separators $\{2\}$ and $\{3\}$.

Since V is supposed to be a subset of a clique of $\mathcal{G}$ or $V = \emptyset$, in view of Lemma A.1 in the Appendix, V necessarily belongs to any of the above (i) to (v). Thus Theorem 3.1 should be applied only to the cases (iv) and (v).

4 Examples

4.1 3-way model of Figure 1

As mentioned in the Section 1, Hara and Takemura[10] derived the improved estimators in the 3-way model in Figure 1. In this model $\mathcal{C} = \{\{1, 2\}, \{2, 3\}\}$, $\mathcal{S} = \{\{2\}\}$. Then the model (1) is expressed by

$$\lambda(i) = \lambda \frac{\alpha(i_{12})\alpha(i_{23})}{\alpha(i_2)}.$$

Consider the class of estimators (10) with $V = \{2\}$ and apply Theorem 3.2 to this model. Since $\nu(\{2\}) = 1$ and $I_{\mathbb{C}(\{2\},\mathcal{G})} = I_1 + I_3$, the condition (17) on $\phi_V(\cdot) = \phi_2(\cdot)$ is written by

$$0 \le \phi_2(x) \le \min\Big(\ 2(I_1 + I_3) - 2\gamma - 2, \quad 2\beta - 2\gamma, \quad (x + \beta)^{\gamma+1}\ \Big). \tag{24}$$

Next we set $V = \{\emptyset\}$ and apply Theorem 3.4 to this model. Since $I^* = I_2(I_1 + I_3 - 1)$, the class (23) is

$$0 \le \phi(x) \le \min\Big(\ 2I_2(I_1 + I_3 - 1) - 2\gamma - 2, \quad 2\beta - 2\gamma, \quad (x + \beta)^{\gamma+1}\ \Big). \tag{25}$$

(24) and (25) coincide with the results of Theorem 4.1 and 4.3 in Hara and Takemura[10], respectively.

Consider the class of estimators (10) with $V = \{1\}$. Since $\{1\}$ is simplicial, we can apply Theorem 3.1 and obtain the following condition on $\phi_V(\cdot) = \phi_1(\cdot)$ corresponding to (14),

$$0 \le \phi_1(x) \le \min\Big(\ 2I_2 - 2\gamma - 2,\ \ 2\beta - 2\gamma,\ \ (x+\beta)^{\gamma+1}\ \Big).$$

This class is not included in that of Hara and Takemura[10].

4.2 5-way model of Figure 3

For the graph in Figure 3, $\mathcal{C} = \{\{1,2\},\{2,3,5\},\{2,4,5\}\}$, $\mathcal{S} = \{\{2\},\{2,5\}\}$ and $\nu(\{2\}) = \nu(\{2,5\}) = 1$. The model (1) is expressed by

$$\lambda(i) = \lambda \frac{\alpha(i_{12})\alpha(i_{235})\alpha(i_{245})}{\alpha(i_2)\alpha(i_{25})}. \tag{26}$$

From (3) $\hat{\lambda}^{ML}(i, \boldsymbol{x}_{\mathcal{C}})$ is

$$\hat{\lambda}^{ML}(i, \boldsymbol{x}_{\mathcal{C}}) = \begin{cases} \dfrac{x(i_{12})x(i_{235})x(i_{245})}{x(i_2)x(i_{25})}, & \text{if } \ x(i_2) \neq 0 \ \text{ and } \ x(i_{25}) \neq 0, \\ 0, & \text{otherwise.} \end{cases}$$

Consider the class of estimators (10) with $V = \{2\}$. $\{2\}$ satisfies $\{2\} \in \mathcal{S}$ and $\{2\} \subset \{2,5\} \in \mathcal{S}$. Thus this case corresponds to the condition (ii) and we can apply Theorem 3.3. From (16) and (20) $I_V^* = I_2^* = I_5(I_3 + I_4 - 1)$. Hence the condition (21) on $\phi_V(\cdot) = \phi_2(\cdot)$ is written by

$$0 \le \phi_2(x) \le \min\Big(\ 2I_5(I_3 + I_4 - 1) - 2\gamma - 2,\ \ 2\beta - 2\gamma,\ \ (x+\beta)^{\gamma+1}\ \Big).$$

Next we consider to take $V = \emptyset$. From (22) I^* in this model is

$$I^* = \max\Big(\ I_{25}(I_3 + I_4 - 1),\ \ I_2(I_1 + I_5 + \max(I_3, I_4) - 1)\ \Big).$$

By applying Theorem 3.4 with this I^*, we can obtain another class.

4.3 The case where the graph is disconnected

When $\mathcal{G}$ is disconnected, we suppose that $\emptyset \in \mathcal{S}$, $\alpha(i_\emptyset) \equiv 1$, $i_{V\cup\emptyset} \equiv i_V$ for $V \subset \Delta$ and that $\nu(\emptyset) = \nu_{\mathcal{G}} - 1$, where $\nu_{\mathcal{G}}$ is the number of connected components. We may consider $\mathcal{G}$ as if $\emptyset$ were adjacent to all the vertices in Δ.

The model (1) with $\mathcal{C} = \{\{1\},\ldots,\{J\}\}$, $|\Delta| = J$, and $\mathcal{S} = \{\emptyset\}$ is called J-way Poisson multiplicative model. The corresponding conditional independent graph is disconnected set of vertices as presented in Figure 5.

The model (1) corresponding to the Figure 5 is written by

$$\lambda(i) = \lambda \prod_{j=1}^{J} \alpha(i_j), \quad \sum_{i_j \in \mathcal{I}_j} \alpha(i_j) = 1.$$

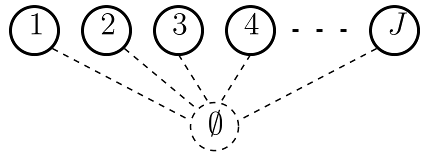

Figure 5: J-way multiplicative model

From (3) the MLE of $\boldsymbol{\lambda}$ is

$$\hat{\lambda}^{ML}(i, \boldsymbol{x}_{\mathcal{C}}) = \begin{cases} \dfrac{\prod_{j=1}^{J} x(i_j)}{(x^+)^{J-1}}, & \text{if } \ x^+ \neq 0, \\ 0, & \text{otherwise.} \end{cases}$$

Hara and Takemura[10] derived the Chou[3]-type estimator improving on the MLE in this model. Consider the class of improved estimators (10) with $V = \emptyset$ and apply Theorem 3.2 to this model. Since $\nu(\emptyset) = J$ and $I_{\mathbb{C}(\emptyset, \mathcal{G})} = \sum_{j=1}^{J} I_j$, then the condition (17) on $\phi_V(\cdot) = \phi(\cdot)$ is expressed by

$$0 \leq \phi(x) \leq \min\Big(\ 2\sum_{j=1}^{J} I_j - 2\gamma - 2J, \ \ 2\beta - 2\gamma, \ \ (x+\beta)^{\gamma+1} \ \Big).$$

This class coincides with the class of Hara and Takemura[10]. By applying Theorem 3.4 we can also obtain the above class.

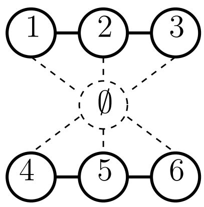

Figure 6: 6-way disconnected model

Next we consider the model in Figure 6. This model is composed of two disconnected 3-way models which is considered in Section 4.1. In this model $\mathcal{C} = \{\{1,2\},\{2,3\},\{4,5\},\{5,6\}\}$, $\mathcal{S} = \{\{2\},\{5\},\emptyset\}$ and $\nu(\{2\}) = \nu(\{5\}) = \nu(\{\emptyset\}) = 1$. Thus the model (1) is written by

$$\lambda(i) = \lambda \frac{\alpha(i_{12})\alpha(i_{23})\alpha(i_{45})\alpha(i_{56})}{\alpha(i_2)\alpha(i_5)}.$$

The MLE of $\boldsymbol{\lambda}$ is

$$\hat{\lambda}^{ML}(i, \boldsymbol{x}_{\mathcal{C}}) = \begin{cases} \dfrac{x(i_{12})x(i_{23})x(i_{45})x(i_{56})}{x(i_2)x(i_5)x^+}, & \text{if } \ x^+ \neq 0, \\ 0, & \text{otherwise.} \end{cases}$$

Consider the class of estimators (10) with $V = \emptyset$. From (22) I^* is

$$I^* = \max\Big(\ I_{12} + I_{23} - I_2,\quad I_{45} + I_{56} - I_5,\quad \max(I_{12}, I_{23}) + \max(I_{45}, I_{56}) - 1\ \Big).$$

By applying Theorem 3.4 with this I^*, we can obtain a class of improved estimators.

5 Monte Carlo Studies

We study the risk performance of the proposed estimators for the 5-way decomposable model corresponding to the graph in Figure 3 through Monte Carlo studies with 100,000 replications. We consider the following class of estimators

$$\hat{\boldsymbol{\lambda}}^{\beta,V} = \hat{\boldsymbol{\lambda}}^{ML}\left(1 - \frac{\beta}{x(i_V) + \beta}\right).$$

This is the class (10) with $\phi_V(x(i_V)) = \beta$ and $\gamma = 0$. We set $I_\delta = I$ for all $\delta \in \Delta = \{1, 2, 3, 4, 5\}$ and $V = \emptyset$, $\{2\}$ and $\{2, 5\}$, which satisfy the conditions (iii), (ii) and (i), respectively. By applying the corresponding theorem in Section 3, we can obtain the upper bound of β for each V in order to improve on $\hat{\boldsymbol{\lambda}}^{ML}$. Denote the upper bound by β_V^U. The theorem which is applied and β_V^U for each V is summarized in Table 1.

The model is expressed by (26). We set $\alpha(i_C) = \prod_{\delta \in C} \alpha(i_\delta)$ for all cliques in this model. With respect to $\alpha(i_\delta)$ we considered the following three cases,

- $\alpha(i_\delta) = 1/I$ for all δ
- $\alpha(i_2)$ are unbalanced
- $\alpha(i_2)$ and $\alpha(i_5)$ are unbalanced, i.e. $\alpha(i_{25})$ are unbalanced.

We write $\boldsymbol{\alpha}(\delta) = \{\alpha(i_\delta)\}_{i_\delta \in \mathcal{I}_\delta}$.

In Table 2 to Table 6 we present the risks of $\hat{\boldsymbol{\lambda}}^{ML}$ and $\hat{\boldsymbol{\lambda}}^{\beta,V}$ with $\beta = \beta_V^U$, $\beta_V^U/2$ and $\beta_V^U/4$ for some λ and $I = 2, 3$. The summary of the experiments is as follows.

- We can confirm the dominance of the proposed estimators over the MLE. As can be expected from the fact that the proposed estimators shrink the MLE towards zero, we can see considerable amount of risk reduction when λ is small.
- The improvement is in the inverse proportion to λ.
- When $\alpha(i_\delta)$ are balanced, $\hat{\boldsymbol{\lambda}}^{\beta,\emptyset}$ shows larger risk reduction.
- When $\alpha(i_V)$, $V \in \{\{2\}, \{2, 5\}\}$ vary widely, $\hat{\boldsymbol{\lambda}}^{\beta,V}$ shows larger risk reduction as λ gets large.

Table 1: The upper bound of β for each V

V	Condition	Theorem	β_V^U
$\emptyset$	(iii)	3.4	$2I^2(2I-1)-2$
$\{2\}$	(ii)	3.3	$2I(2I-1)-2$
$\{2,5\}$	(i)	3.2	$2(2I-1)-2$

Table 2: Risks of $\hat{\boldsymbol{\lambda}}^{ML}$ and $\hat{\boldsymbol{\lambda}}^{\beta,V}$ for the model in Figure 3 with balanced $\alpha(i_\delta)$.

(1) $I=2$ and $\alpha(i_\delta)=1/2$

		λ						
V	β	0.1	0.5	1.0	5.0	10.0	50.0	100.0
	22	0.165	0.574	1.058	4.146	6.630	12.097	13.161
$\emptyset$	11	0.349	0.849	1.421	4.804	7.162	11.782	12.807
	5.5	0.949	1.706	2.526	6.780	9.102	12.607	13.271
	10	0.369	0.773	1.247	4.311	6.796	12.236	13.247
$\{2\}$	5	1.017	1.502	2.052	5.369	7.702	12.113	13.010
	2.5	2.780	3.400	4.069	7.732	9.824	12.892	13.431
	4	1.380	1.747	2.182	5.073	7.470	12.603	13.447
$\{2,5\}$	2	3.656	4.007	4.409	7.077	9.098	12.771	13.391
	1	8.062	8.268	8.478	10.181	11.454	13.425	13.718
$\hat{\boldsymbol{\lambda}}^{ML}$		31.567	30.106	28.323	20.972	17.790	14.717	14.362

(2) $I=3$ and $\alpha(i_\delta)=1/3$

		λ						
V	β	0.1	0.5	1.0	5.0	10.0	50.0	100.0
	88	0.136	0.554	1.073	4.997	9.295	28.822	38.297
$\emptyset$	44	0.244	0.735	1.352	5.995	10.799	29.241	37.126
	22	0.654	1.418	2.399	9.652	16.362	35.593	41.760
	28	0.396	0.813	1.334	5.230	9.491	29.048	38.553
$\{2\}$	14	1.212	1.715	2.351	6.979	11.727	30.239	38.006
	7	4.000	4.748	5.717	12.402	18.533	36.900	42.676
	8	3.090	3.416	3.843	7.246	11.184	30.328	39.706
$\{2,5\}$	4	9.782	9.964	10.268	13.157	16.818	33.744	40.819
	2	26.919	26.546	26.318	26.585	28.679	41.082	45.424
$\hat{\boldsymbol{\lambda}}^{ML}$		239.31	226.97	214.44	148.82	111.08	63.33	56.92

Table 3: Risks of $\hat{\boldsymbol{\lambda}}^{ML}$ and $\hat{\boldsymbol{\lambda}}^{\beta,V}$ for the model in Figure 3 with unbalanced $\alpha(i_2)$

(1) $I = 2$ and $\lambda = 1.0$

		$\boldsymbol{\alpha}(2)$			
V	β	(0.1,0.9)	(0.2,0.8)	(0.3,0.7)	(0.4,0.6)
	22	1.066	1.064	1.062	1.059
$\emptyset$	11	1.445	1.441	1.433	1.424
	5.5	2.586	2.578	2.556	2.533
	10	1.221	1.237	1.245	1.247
$\{2\}$	5	2.015	2.042	2.053	2.051
	2.5	4.013	4.061	4.078	4.070
	4	2.169	2.185	2.189	2.184
$\{2,5\}$	2	4.410	4.433	4.432	4.416
	1	8.518	8.548	8.530	8.492
$\hat{\boldsymbol{\lambda}}^{ML}$		28.700	28.699	28.533	28.379

(2) $I = 2$ and $\lambda = 10.0$

		$\boldsymbol{\alpha}(2)$			
V	β	(0.1,0.9)	(0.2,0.8)	(0.3,0.7)	(0.4,0.6)
	22	6.978	6.791	6.692	6.643
$\emptyset$	11	7.928	7.518	7.297	7.190
	5.5	10.453	9.735	9.342	9.154
	10	5.636	6.169	6.523	6.726
$\{2\}$	5	6.369	7.003	7.400	7.624
	2.5	8.299	9.059	9.500	9.740
	4	6.371	6.873	7.201	7.402
$\{2,5\}$	2	8.007	8.511	8.830	9.030
	1	10.713	11.051	11.261	11.404
$\hat{\boldsymbol{\lambda}}^{ML}$		20.889	19.267	18.351	17.916

(3) $I = 2$ and $\lambda = 100.0$

		$\boldsymbol{\alpha}(2)$			
V	β	(0.1,0.9)	(0.2,0.8)	(0.3,0.7)	(0.4,0.6)
	22	13.603	13.291	13.200	13.161
$\emptyset$	11	13.352	12.970	12.861	12.814
	5.5	13.881	13.454	13.335	13.282
	10	11.583	12.654	13.039	13.195
$\{2\}$	5	11.714	12.538	12.843	12.970
	2.5	12.611	13.143	13.333	13.407
	4	11.929	12.933	13.276	13.404
$\{2,5\}$	2	12.380	13.049	13.277	13.363
	1	13.228	13.558	13.668	13.705
$\hat{\boldsymbol{\lambda}}^{ML}$		15.047	14.569	14.437	14.378

Table 4: Risks of $\hat{\boldsymbol{\lambda}}^{ML}$ and $\hat{\boldsymbol{\lambda}}^{\beta,V}$ for the model in Figure 3 with unbalanced $\alpha(i_2)$

(1) $I = 3$ and $\lambda = 1.0$

		$\boldsymbol{\alpha}(2)$			
V	β	(0.1,0.2,0.7)	(0.1,0.3,0.6)	(0.2,0.3,0.5)	(0.3,0.3,0.4)
	88	1.076	0.545	0.482	0.419
$\emptyset$	44	1.364	0.730	0.654	0.579
	22	2.440	1.423	1.302	1.182
	28	1.326	0.802	0.741	0.679
$\{2\}$	14	2.342	1.708	1.635	1.557
	7	5.701	4.752	4.644	4.521
	8	3.854	3.416	3.361	3.302
$\{2,5\}$	4	10.313	9.986	9.943	9.878
	2	26.447	26.622	26.651	26.587
$\hat{\boldsymbol{\lambda}}^{ML}$		215.75	227.83	229.77	230.25

(2) $I = 3$ and $\lambda = 10.0$

		$\boldsymbol{\alpha}(2)$			
V	β	(0.1,0.2,0.7)	(0.1,0.3,0.6)	(0.2,0.3,0.5)	(0.3,0.3,0.4)
	88	9.569	5.050	4.437	3.822
$\emptyset$	44	11.681	6.386	5.559	4.744
	22	18.762	11.008	9.507	8.045
	28	8.811	4.741	4.315	3.857
$\{2\}$	14	10.718	6.217	5.777	5.293
	7	16.730	10.886	10.396	9.837
	8	10.673	6.942	6.520	6.080
$\{2,5\}$	4	16.386	13.077	12.649	12.218
	2	28.964	27.589	27.164	26.796
$\hat{\boldsymbol{\lambda}}^{ML}$		131.95	176.81	176.70	177.95

(3) $I = 3$ and $\lambda = 100.0$

		$\boldsymbol{\alpha}(2)$			
V	β	(0.1,0.2,0.7)	(0.1,0.3,0.6)	(0.2,0.3,0.5)	(0.3,0.3,0.4)
	88	39.818	34.014	29.736	26.142
$\emptyset$	44	39.739	40.731	34.317	29.450
	22	45.414	55.387	46.353	39.747
	28	32.944	19.898	20.145	20.076
$\{2\}$	14	33.378	20.829	21.347	21.556
	7	38.899	26.176	27.296	28.004
	8	34.081	21.464	21.644	21.532
$\{2,5\}$	4	36.413	25.378	25.654	25.738
	2	42.505	34.945	34.957	35.024
$\hat{\boldsymbol{\lambda}}^{ML}$		62.380	105.56	93.761	86.945

Table 5: Risks of $\hat{\boldsymbol{\lambda}}^{ML}$ and $\hat{\boldsymbol{\lambda}}^{\beta,V}$ for the model in Figure 3 with unbalanced $\alpha(i_{25})$

(1) $I = 2$ and $\lambda = 1.0$

		$\boldsymbol{\alpha}(2) = \boldsymbol{\alpha}(5)$			
V	β	(0.1,0.9)	(0.2,0.8)	(0.3,0.7)	(0.4,0.6)
	22	1.066	1.062	1.062	1.059
$\emptyset$	11	1.444	1.433	1.433	1.424
	5.5	2.581	2.556	2.558	2.534
	10	1.220	1.234	1.247	1.248
$\{2\}$	5	2.010	2.035	2.059	2.054
	2.5	3.996	4.042	4.090	4.074
	4	2.120	2.154	2.181	2.182
$\{2, 5\}$	2	4.327	4.377	4.422	4.412
	1	8.389	8.455	8.519	8.488
$\hat{\boldsymbol{\lambda}}^{ML}$		28.508	28.511	28.550	28.390

(2) $I = 2$ and $\lambda = 10.0$

		$\boldsymbol{\alpha}(2) = \boldsymbol{\alpha}(5)$			
V	β	(0.1,0.9)	(0.2,0.8)	(0.3,0.7)	(0.4,0.6)
	22	7.034	6.819	6.705	6.647
$\emptyset$	11	8.051	7.580	7.331	7.200
	5.5	10.668	9.846	9.406	9.172
	10	5.752	6.216	6.544	6.732
$\{2\}$	5	6.572	7.092	7.447	7.639
	2.5	8.590	9.197	9.576	9.765
	4	5.525	6.394	6.989	7.345
$\{2, 5\}$	2	7.185	8.030	8.617	8.971
	1	9.986	10.608	11.066	11.350
$\hat{\boldsymbol{\lambda}}^{ML}$		21.377	19.521	18.504	17.957

(3) $I = 2$ and $\lambda = 100.0$

		$\boldsymbol{\alpha}(2) = \boldsymbol{\alpha}(5)$			
V	β	(0.1,0.9)	(0.2,0.8)	(0.3,0.7)	(0.4,0.6)
	22	13.906	13.385	13.214	13.171
$\emptyset$	11	13.715	13.082	12.877	12.826
	5.5	14.280	13.578	13.352	13.295
	10	11.753	12.719	13.057	13.205
$\{2\}$	5	11.959	12.629	12.862	12.981
	2.5	12.926	13.253	13.351	13.420
	4	10.754	12.394	13.113	13.383
$\{2, 5\}$	2	11.308	12.628	13.145	13.347
	1	12.374	13.280	13.579	13.699
$\hat{\boldsymbol{\lambda}}^{ML}$		15.490	14.706	14.454	14.393

Table 6: Risks of $\hat{\boldsymbol{\lambda}}^{ML}$ and $\hat{\boldsymbol{\lambda}}^{\beta,V}$ for the model in Figure 3 with unbalanced $\alpha(i_{25})$

(1) $I = 3$ and $\lambda = 1.0$

		$\boldsymbol{\alpha}(2) = \boldsymbol{\alpha}(5)$			
V	β	(0.1,0.2,0.7)	(0.1,0.3,0.6)	(0.2,0.3,0.5)	(0.3,0.3,0.4)
	88	1.076	1.074	1.073	1.073
$\emptyset$	44	1.365	1.359	1.356	1.353
	22	2.445	2.423	2.412	2.402
	28	1.328	1.329	1.333	1.334
$\{2\}$	14	2.348	2.345	2.352	2.353
	7	5.716	5.702	5.722	5.722
	8	3.852	3.843	3.853	3.847
$\{2,5\}$	4	10.321	10.281	10.301	10.279
	2	26.480	26.363	26.397	26.344
$\hat{\boldsymbol{\lambda}}^{ML}$		216.19	215.07	214.90	214.60

(2) $I = 3$ and $\lambda = 10.0$

		$\boldsymbol{\alpha}(2) = \boldsymbol{\alpha}(5)$			
V	β	(0.1,0.2,0.7)	(0.1,0.3,0.6)	(0.2,0.3,0.5)	(0.3,0.3,0.4)
	88	9.595	9.518	9.367	9.305
$\emptyset$	44	11.755	11.504	11.034	10.839
	22	18.951	18.265	17.003	16.481
	28	8.901	9.119	9.350	9.472
$\{2\}$	14	10.945	11.227	11.543	11.714
	7	17.197	17.661	18.230	18.532
	8	10.037	10.482	10.936	11.155
$\{2,5\}$	4	15.543	16.034	16.538	16.801
	2	27.776	28.174	28.479	28.704
$\hat{\boldsymbol{\lambda}}^{ML}$		133.55	127.57	116.56	112.01

(3) $I = 3$ and $\lambda = 100.0$

		$\boldsymbol{\alpha}(2) = \boldsymbol{\alpha}(5)$			
V	β	(0.1,0.2,0.7)	(0.1,0.3,0.6)	(0.2,0.3,0.5)	(0.3,0.3,0.4)
	88	40.760	40.201	38.756	38.387
$\emptyset$	44	41.287	40.362	37.898	37.250
	22	47.539	46.266	42.831	41.916
	28	33.548	34.920	37.436	38.440
$\{ 2 \}$	14	34.427	35.394	37.262	37.942
	7	40.480	41.025	42.276	42.652
	8	30.539	32.689	37.378	39.422
$\{2,5\}$	4	33.111	34.971	39.008	40.613
	2	39.391	40.916	44.151	45.292
$\hat{\boldsymbol{\lambda}}^{ML}$		65.520	63.636	58.511	57.134

Appendix

A Characterization of simplicial vertex

Lemma A.1. *The following conditions are equivalent for a decomposable graph $\mathcal{G}$.*

(i) $\delta \in \Delta$ *is simplicial ;*

(ii) there is only one clique which includes δ *;*

(iii) $\delta \notin S$ *for all* $S \in \mathcal{S}$*.*

Proof. First we show that (i) $\Leftrightarrow$ (ii). Suppose $\delta \in \Delta$ is simplicial. From the definition of simplicial vertex the union of δ and $\mathrm{adj}(\delta, \mathcal{G})$ is a complete subset of Δ that is maximal with respect to the inclusive relation, i.e. a clique. Thus there is only one clique which includes δ. Conversely if we assume (ii) and denote the clique by C_δ, $\mathrm{adj}(\delta, \mathcal{G}) = C_\delta \setminus \{\delta\}$ is complete. Thus (i) $\Leftrightarrow$ (ii).

Next that (ii) $\Leftrightarrow$ (iii). Suppose (ii) and let C_δ be the clique. Since $\mathcal{G}$ is decomposable, there exists a perfect sequence of cliques $C_1, \ldots, C_K$ and the corresponding minimal vertex separators $S_2, \ldots, S_K$ such that

$$C_1 = C_\delta, \quad S_k = (\bigcup_{j=1}^{k-1} C_j) \cap C_k = \bigcup_{j=1}^{k-1} (C_j \cap C_k), \quad k = 2, \ldots, K$$

(See Lemma 2.18 in Lauritzen[15] for example). We note that $\delta \cap C_k = \emptyset$ for $k \geq 2$. Thus $\delta \notin S_k$ for all $S_k \in \mathcal{S}$. Conversely suppose that there exists two cliques C_a and C_b such that $\delta \subset C_a \cap C_b$. Then δ has $S \in \mathcal{S}$ such that $\delta \subset S$ from the running intersection property of perfect sequences. Thus the proof is completed. □

B The proof of Lemma 2.9

In order to prove Lemma 2.9, we use the following lemma.

Lemma B.1. *If* $\mathcal{G} \neq \mathcal{G}(\Delta(\tilde{\mathcal{C}}))$ *for* $\tilde{\mathcal{C}} \in \mathbb{C}(S, \mathcal{G})$ *and* $S \in \mathcal{S}$*,* $\Delta \setminus \Delta(\tilde{\mathcal{C}})$ *includes at least one simplicial vertex in* $\mathcal{G}$*.*

Proof. The proof is by induction on the number of vertices J. The lemma is trivial if $J \leq 4$. Suppose $J > 4$ and assume that the lemma holds for all decomposable graphs with fewer than J vertices.

Suppose that $\Delta \setminus \Delta(\tilde{\mathcal{C}})$ does not include simplicial vertex in $\mathcal{G}$. Let $\delta \in \Delta(\tilde{\mathcal{C}})$ be a simplicial vertex in $\mathcal{G}$. We write $\mathcal{G}^\delta = \mathcal{G}(\Delta \setminus \{\delta\})$. Then there exists a $\tilde{\mathcal{C}}' \in \mathbb{C}(S, \mathcal{G}^\delta)$ such that $\mathcal{G}^\delta(\Delta(\tilde{\mathcal{C}}')) = \mathcal{G}(\Delta(\tilde{\mathcal{C}}) \setminus \{\delta\})$. Since there is only one clique which includes δ from Lemma A.1, $\mathrm{adj}(\delta, \mathcal{G}) \cap (\Delta \setminus \Delta(\tilde{\mathcal{C}})) = \emptyset$. Thus $\mathrm{adj}(v, \mathcal{G}) = \mathrm{adj}(v, \mathcal{G}(\Delta \setminus \{\delta\}))$ for any $v \in \Delta \setminus \Delta(\tilde{\mathcal{C}})$, which implies $v \in \Delta$ is not simplicial in $\mathcal{G}^\delta$ either. This contradicts the inductive assumption from the fact that $\Delta \setminus \Delta(\tilde{\mathcal{C}}) = (\Delta \setminus \{\delta\}) \setminus (\Delta(\tilde{\mathcal{C}}) \setminus \{\delta\}) = (\Delta \setminus \{\delta\}) \setminus \Delta(\tilde{\mathcal{C}}')$. This proves the lemma. □

Proof of Lemma 2.9 The proof is by induction on the number of vertices J. The lemma is trivial if $J = 2$. Suppose $J > 2$ and assume that the lemma holds for all decomposable graph with fewer than J vertices.

Suppose that $\mathcal{G}$ has J vertices. If $\mathcal{G}(\Delta(\tilde{\mathcal{C}}))$ includes all of the simplicial vertices in $\mathcal{G}$, $\mathcal{G}(\Delta(\tilde{\mathcal{C}})) = \mathcal{G}$ from Lemma B.1.

In the case where $\mathcal{G}(\Delta(\tilde{\mathcal{C}})) \neq \mathcal{G}$, there exists at least one simplicial vertex δ such that $\delta \notin \Delta(\tilde{\mathcal{C}})$. Since $\mathcal{G}(\Delta \setminus \delta)$ has $J-1$ vertices, there exist a perfect elimination scheme $\sigma_{\Delta\setminus\{\delta\}} : \{1, 2, \ldots, J-1\} \mapsto \Delta \setminus \{\delta\}$ and $j < J-1$ such that

$$\Delta(\tilde{\mathcal{C}}) = \{\sigma_{\Delta\setminus\{\delta\}}(j-1), \ldots, \sigma_{\Delta\setminus\{\delta\}}(J-1)\}$$

from the inductive assumption. If we set $\sigma(1) = \delta$ and $\sigma(j) = \sigma_{\Delta\setminus\{\delta\}}(j-1)$, $j = 2, \ldots, J$, then σ satisfies (9), which completes the proof. □

C The proof of Lemma 2.2 and Lemma 2.3

Proof of Lemma 2.2 As we mentioned in Section 2, the MLE of $\alpha(i_C)$ is $x(i_C)/x^+$ for all $C \in \mathcal{C}$. Since any cliques and minimal vertex separators in $\mathcal{G}(\Delta(\sigma, j))$ are included in some cliques in $\mathcal{C}$, the MLE of $\alpha(i_{C'})$ and $\alpha(i_{S'})$ for all $C' \in \mathcal{C}(\Delta(\sigma, j))$ and $S' \in \mathcal{S}(\Delta(\sigma, j))$ are $x(i_{C'})/x^+$ and $x(i_{S'})/x^+$, respectively. Thus it suffices to show that the conditional independent graph of $\Delta(\sigma, j)$ is $\mathcal{G}(\Delta(\sigma, j))$.

We prove this by induction on j. The lemma is trivial if $j = 1$. Suppose $j_0 \geq 2$ and assume that the lemma holds for $j \leq j_0$.

$\sigma(j_0)$ is a simplicial vertex in $\mathcal{G}(\Delta(\sigma, j_0))$. From the Lemma A.1 $\sigma(j_0)$ satisfies $\sigma(j_0) \notin S$ for all $S \in \mathcal{S}(\Delta(\sigma, j_0))$. Thus any two of vertices in $\Delta(\sigma, j_0 + 1) = \Delta(\sigma, j_0) \setminus \{\sigma(j_0)\}$ is not separated by $\sigma(j_0)$. From the global Markov property the conditional independent graph of $\Delta(\sigma, j_0 + 1)$ is $\mathcal{G}(\Delta(\sigma, j_0 + 1))$, which completes the proof. □

Proof of Lemma 2.3. Let C_1 be the clique which includes $\sigma(1)$. Then we have

$$\begin{aligned}
\sum_{i':i'_{\Delta(\sigma,2)}=i_{\Delta(\sigma,2)}} \hat{\lambda}^{ML}(i', \boldsymbol{x}_C + \boldsymbol{e}_{\mathcal{C}}^{i'}) &= \sum_{i':i'_{\Delta(\sigma,2)}=i_{\Delta(\sigma,2)}} \frac{\prod_{C\in\mathcal{C}}(x(i'_C)+1)}{\prod_{S\in\mathcal{S}}(x(i'_S)+1)^{\nu(S)}} \\
&= \frac{(x(i_{C_1\setminus\{\sigma(1)\}}) + I_{\sigma(1)})\prod_{C\neq C_1}(x(i_C)+1)}{\prod_{S\in\mathcal{S}}(x(i_S)+1)^{\nu(S)}} \\
&\geq \frac{(x(i_{C_1\setminus\{\sigma(1)\}}) + 1)\prod_{C\neq C_1}(x(i_C)+1)}{\prod_{S\in\mathcal{S}}(x(i_S)+1)^{\nu(S)}} \\
&= \frac{\prod_{C\in\mathcal{C}(\Delta(\sigma,2))}(x(i_C)+1)}{\prod_{S\in\mathcal{S}(\Delta(\sigma,2))}(x(i_S)+1)^{\nu(S,\Delta(\sigma,2))}} \\
&= \hat{\lambda}^{ML}(i_{\Delta(\sigma,2)}, \boldsymbol{x}_{\mathcal{C}} + \boldsymbol{e}_{\mathcal{C}}^{i})
\end{aligned}$$

By iterating this operation for $\sigma(k)$, $k = 2, \ldots, j$, in sequence, we obtain as a consequence

$$\begin{aligned}\sum_{i'_{\Delta(\sigma,j-1)}:i'_{\Delta(\sigma,j)}=i_{\Delta(\sigma,j)}} \cdots \sum_{i':i'_{\Delta(\sigma,2)}=i_{\Delta(\sigma,2)}} \hat{\lambda}^{ML}(i', \boldsymbol{x}_C + \boldsymbol{e}_{\mathcal{C}}^{i'}) &\geq \frac{\prod_{C\in\mathcal{C}(\Delta(\sigma,j))}(x(i_C)+1)}{\prod_{S\in\mathcal{S}(\Delta(\sigma,j))}(x(i_S)+1)^{\nu(S,\Delta(\sigma,j))}} \\ &= \hat{\lambda}^{ML}(i_{\Delta(\sigma,j)}, \boldsymbol{x}_{\mathcal{C}} + \boldsymbol{e}_{\mathcal{C}}^{i}).\end{aligned}$$

□

References

[1] Blair, J. R. S. and Peyton, B. (1993). An introduction to chordal graphs and clique trees. In *Graph Theory and Sparse Matrix Computations*(J. R. Gilbert and J. W. H. Liu, eds.), IMA Vol. Math. Appl., **56**, 1-30. Springer, Berlin.

[2] Brown, L. D. and Hwang, J. T. (1982). A unified admissibility proof. In *Statistical Decision Theory and Related Topics III*(S. S. Gupta and J. O. Berger, eds.), **1**, 205-230. Academic, New York.

[3] Chou, J. P. (1991). Simultaneous estimation in discrete multivariate exponential families, *Ann. Statist.*, **19**, 314-328.

[4] Clevenson, M. L. and Zidek, J. V. (1975). Simultaneous estimation of the means of independent Poisson laws, *J. Amer. Statist. Assoc.*, **70**, 698-705.

[5] Darroch, J. N., Lauritzen, S. L. and Speed, T. P. (1980). Markov fields and log-linear interaction models for contingency tables, *Ann. Statist.*, **8**, 522-539.

[6] Dirac, G. A. (1961). On rigid circuit graphs, *Abhandlungen Mathematisches Seminar Hamburg*, **25**, 71-76.

[7] Ghosh, M., Hwang, J. T. and Tsui, K. W. (1983). Construction of improved estimators in multiparameter estimation for discrete exponential families, *Ann. Statist.*, **11**, 351-367.

[8] Ghosh, M. and Parsian, A. (1981). Bayes minimax estimation of multiple Poisson parameters, *J. Multivariate Analysis*, **11**, 280-288.

[9] Ghosh, M. and Yang, M. C. (1988). Simultaneous estimation of Poisson means under entropy loss, *Ann. Statist.*, **16**, 278-291.

[10] Hara, H. and Takemura, A. (2004). Simultaneous estimation of the means in some Poisson log linear models, Technical Reports, *METR*, 2004-47, Department of Mathematical Informatics, University of Tokyo(available from `http://www.keisu.t.u-tokyo.ac.jp/Research/techrep.0.html`).

[11] Hudson, H. M. (1978). A natural identity for exponential families with applications in multiparameter estimation. *Ann. Statist.*, **6**, 478-484.

[12] Hwang, J. T. (1982). Improving upon standard estimators in discrete exponential families with applications to Poisson and negative binomial cases. *Ann. Statist.*, **10**, 857-867.

[13] Johnstone, I. (1984). Admissibility, difference equations and recurrence in estimating a Poisson mean. *Ann. Statist.*, **12**, 1173-1198.

[14] Komaki, F. (2004). Simultaneous prediction of independent Poisson observables. *Ann. Statist.*, **32**, 1744–1769.

[15] Lauritzen, S. L. (1996). Graphical Models, Clarendon Press, Oxford.

[16] Letac, G. and Massam, H. (2003). Wishart distributions for decomposable graphs, Working Paper(available from `http://www.lsp.ups-tlse.fr/Fp/Letac/`).

[17] Sundberg, R. (1975). Some results about decomposable (or Markov-type) models for multidimensional contingency tables: Distribution of marginals and partitioning of tests. *Scandinavian Journal of Statistics*, **2**, 71-79.

[18] Tsui, K. W. (1984). Robustness of Clevenson-Zidek-type estimators, *J. Amer. Statist. Assoc.*, **79**, 152-157.

[19] Tsui, K. W. and Press, S. J. (1982). Simultaneous Estimation of Several Poisson Parameters Under K-Normalized Squared Error Loss, *Ann. Statist.*, **10**, 93-100.