MATHEMATICAL ENGINEERING TECHNICAL REPORTS

Bayes Admissible Estimation of the Means in Poisson Decomposable Graphical Models

Hisayuki Hara and Akimichi Takemura

METR 2005–22 August 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

Bayes Admissible Estimation 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

August, 2005

Abstract

We investigate the Bayes estimation of the means in Poisson decomposable graphical models. Some classes of Bayes estimators are provided which improve on the maximum likelihood estimator under the normalized squared error loss. Both proper and improper priors are included in the proposed class of priors. Concerning the generalized Bayes estimators with respect to the improper priors we address their admissibility.

Keywords and phrases : admissibility, Bayes, chordal graph, contingency table, decomposable graphical Poisson model, inadmissibility, perfect sequence, shrinkage estimation.

1 Introduction

In this article we study the simultaneous Bayes estimation of the means in Poisson decomposable graphical models. Consider a J-way contingency table. Let $\Delta = \{1,\dots,J\}$ be the set of variables which corresponds to the set of vertices in the conditional independence graph $\mathcal{G}$. $\mathcal{G}$ is assumed to be chordal(decomposable). Let I_δ denote the number of levels for $\delta \in \Delta$. We express the set of levels of δ by $\mathcal{I}_\delta = \{1,\dots,I_\delta\}$. 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. $\boldsymbol{x} = \{x(i)\}_{i\in\mathcal{I}} \in \mathbb{Z}_+^{|\mathcal{I}|}$ denotes the vector of the cell frequencies.

Then the Poisson decomposable graphical model is expressed as follows,

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

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

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

where $\mathcal{C}$ is the set of cliques of $\mathcal{G}$ 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 assume that $\mathcal{G}$ is chordal but not complete and address the problem of the simultaneous estimation of $\boldsymbol{\lambda} = \{\lambda(i)\}_{i \in \mathcal{I}}$ from the decision theoretic viewpoint under the following normalized squared error loss function

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

We derive the estimators which dominate the MLE under the loss (2).

When $\mathcal{G}$ is complete, the model (1) corresponds to the saturated model. A series of results on the shrinkage estimation of multivariate Poisson means which were inspired by Clevenson and Zidek[5] correspond to this setting. In the saturated model $\boldsymbol{x} = \{x(i)\}_{i \in \mathcal{I}}$ is the maximum likelihood, the uniformly minimum variance unbiased and a minimax estimator. Clevenson and Zidek[5] presented a class of Bayes estimators improving on $\boldsymbol{x}$ for $|\mathcal{I}| \geq 2$. Their class includes both proper and generalized Bayes estimators. Ghosh and Parsian[7] enlarged the class of Clevenson and Zidek[7] and presented another class of proper Bayes estimators. Brown and Farrel[2] and Johnstone[14] provided some complete class theorems for generalized Bayes estimators of $\boldsymbol{\lambda}$ in the saturated model. Brown and Hwang[3] first showed that the generalized Bayes estimators in the class of Clevenson and Zidek[5] are admissible. Johnstone[14] proved the admissibility of a wider class of specific estimators which includes the class of Clevenson and Zidek[5]. Other important results on the simultaneous estimation in the saturated model may be found, for example, in Chou[4], Ghosh and Yang[8] Hwang[12], Ghosh, Hwang and Tsui[6], Tsui[18], Tsui and Press[19] etc.

The model (1) with $\mathcal{S} = \{\emptyset\}$ and $\mathcal{C} = \{\{1\}, \ldots, \{J\}\}$ is called Poisson multiplicative model. Recently Hara and Takemura[9] applied the argument in the saturated model to the Poisson multiplicative models and presented two classes of estimators which improves on the MLE. One is the class of Bayes estimator which is the extension of the one in Clevenson and Zidek[5]. The other corresponds to the class of Chou[4]. Hara and Takemura[10] showed that the results with respect to the Chou-type estimators in Hara and Takemura[9] can be extended to the general decomposable model (1) and provided some classes of estimators which dominate the MLE.

The main purpose in this article is to extend the results on the Bayes estimation in Ghosh and Parsian[7] and Clevenson and Zidek[5] and Hara and Takemura[9] to the general Poisson decomposable models (1) and provide some classes of Bayes admissible estimators improving on the MLE under the loss function (2).

The paper is organized as follows. In Section 2 we summarize some basic facts on Poisson decomposable graphical models and decomposable graphs and prepare some lemmas which we utilize in the following arguments. In Section 3 we introduce two types of priors and the corresponding Bayes estimators which we discuss in this article. Section 4 provides some class of improved Bayes estimators for the model with only one minimal vertex separator. In Section 5 we extend the results obtained in Section 4 to the general decomposable graphical models. The proposed class of improved estimators includes both proper Bayes and generalized Bayes with respect to improper priors. In Section 6 we show that some of the generalized Bayes estimators proposed in Section 4 and 5 are admissible by extending the argument of Johnstone[14]. Section 7 gives Monte Carlo studies which confirm the theoretical results of the dominance relationship.

2 Preliminary facts of the Poisson decomposable graphical models and the decomposable graph

2.1 Poisson decomposable models and the MLE of the means

In this section we first define some notations which we use in the following arguments. Mostly we follow the notations of Lauritzen[16] and Hara and Takemura[10].

As is mentioned in the previous section, we assume that the conditional independence graph $\mathcal{G}$ is chordal and not complete. For a subset of vertices $V \subset \Delta$ let $\mathcal{G}(V)$ denote the subgraph induced by V. $\mathcal{C}(V)$ and $\mathcal{S}(V)$ represent the set of the cliques and the set of the minimal vertex separators in $\mathcal{G}(V)$, respectively. Define I_V for $V \subset \Delta$ as $I_V = \prod_{\delta \in V} I_\delta$ and $I_{\{\emptyset\}} \equiv 1$. For a set of cliques $\mathcal{C}^* \subset \mathcal{C}$, define $\Delta(\mathcal{C}^*) \subset \Delta$ by $\Delta(\mathcal{C}^*) = \bigcup_{C \in \mathcal{C}^*} C$.

Suppose $|\mathcal{C}| = K$. Let $C_1, C_2, \ldots, C_K$ denote 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. Define $S_1 = H_0 = S_2$. It is easy to show that $R_k = C_k \setminus H_{k-1} = C_k \setminus S_k \neq \emptyset$ for $k = 1, \ldots, K$. Then the model (1) is rewritten by

$$\lambda(i) = \lambda \alpha(i_{S_1}) \prod_{k=1}^{K} \beta(i_{C_k}), \quad \beta(i_{C_k}) = \frac{\alpha(i_{C_k})}{\alpha(i_{S_k})}, \tag{3}$$

$$\sum_{i_{S_1} \in \mathcal{I}_{S_1}} \alpha(i_{S_1}) = 1, \qquad \sum_{j_{C_k} : j_{S_k} = i_{S_k}} \beta(j_{C_k}) = \sum_{j_{R_k} \in \mathcal{I}_{R_k}} \beta(i_{S_k}, j_{R_k}) = 1, \quad \forall i_{S_k} \in \mathcal{I}_{S_k}, \ k = 1, \ldots, K.$$

For our discussion in Section 5, we briefly consider variables with just 1 level. Let $\Delta_1 \subset \Delta$ be the set of all vertices such that $I_\delta = 1$, $\delta \in \Delta$. If $\Delta_1 \neq \emptyset$, the model (3) is essentially equivalent to the decomposable model for $\mathcal{G}(\Delta \setminus \Delta_1)$. We note that if we assume $I_\delta \geq 1$, the definition (3) includes the model for all of the induced subgraphs of $\mathcal{G}$. For avoiding triviality, we assume $I_\delta \geq 2$ for all $\delta \in \Delta$, i.e. $\Delta_1 = \emptyset$ unless otherwise noted.

For $V \subset \Delta$, let $x(i_V)$ denote the marginal frequency. Then $\boldsymbol{x}_{\mathcal{C}} = \{x(i_C), i_C \in \mathcal{I}_C, C \in \mathcal{C}\}$ 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[16], the marginal probability function of $\boldsymbol{x}_{\mathcal{C}}$ is expressed by

$$\Pr(\boldsymbol{x}_{\mathcal{C}}) = e^{-\lambda} \lambda^{x^+} \cdot t(\boldsymbol{x}_{\mathcal{C}}) \cdot \prod_{i_{S_1} \in \mathcal{I}_{S_1}} \alpha(i_{S_1})^{x(i_{S_1})} \cdot \prod_{k=1}^{K} \prod_{i_{C_k} \in \mathcal{I}_{C_k}} \beta(i_{C_k})^{x(i_{C_k})}, \tag{4}$$

where

$$t(\boldsymbol{x}_\mathcal{C}) = \frac{\prod_{k=2}^K \prod_{i_{S_k} \in \mathcal{I}_{S_k}} x(i_{S_k})!}{\prod_{k=1}^K \prod_{i_{C_k} \in \mathcal{I}_{C_k}} x(i_{C_k})!}$$

and the MLE of $\boldsymbol{\lambda}$ is given by

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

$$\hat{\lambda}^{ML}(i) = \begin{cases} \dfrac{\prod_{k=1}^K x(i_{C_k})}{\prod_{k=2}^K x(i_{S_k})}, & \text{if } \ x(i_{S_k}) \neq 0 \ \text{ for } \ k = 2, \ldots, K, \\ \\ 0, & \text{otherwise.} \end{cases} \tag{5}$$

The following identity obtained by Hara and Takemura[10] which is an extension of the one by Hudson[11] and Hwang[12] is useful when we evaluate the estimators in the following argument.

Lemma 2.1 (Hara and Takemura[10]). *Suppose that $\boldsymbol{x}_\mathcal{C}$ has the probability function (4). 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 $x(i_C) < 1$, then*

$$E\left[\frac{g(\boldsymbol{x}_\mathcal{C})}{\lambda(i)}\right] = E\left[\frac{t(\boldsymbol{x}_\mathcal{C} + \boldsymbol{e}^i_\mathcal{C})}{t(\boldsymbol{x}_\mathcal{C})} \cdot g(\boldsymbol{x}_\mathcal{C} + \boldsymbol{e}^i_\mathcal{C})\right] = E\left[\frac{\prod_{k=2}^K (x(i_{S_k}) + 1)}{\prod_{k=1}^K (x(i_{C_k}) + 1)} \cdot g(\boldsymbol{x}_\mathcal{C} + \boldsymbol{e}^i_\mathcal{C})\right], \tag{6}$$

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

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

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

2.2 Some preparations on decomposable graphs

In this section we prepare a proposition on decomposable graphs needed later.

For a minimal vertex separator $S \in \mathcal{S}$, let $\mathcal{C}_S = \{C \in \mathcal{C} \mid C \supset S\}$ denote the set of cliques which includes S. From Proposition 2.1 in Hara and Takemura[10], S decomposes $\mathcal{G}(\Delta(\mathcal{C}_S))$ into $\nu(S) + 1$ connected components. Denote them by $\Gamma^S_1, \Gamma^S_2, \ldots, \Gamma^S_{\nu(S)+1}$. Let $\mathbb{C}(S, \mathcal{G})$ be the class of $\nu(S) + 1$ cliques in $\mathcal{C}_S$ such that

$$\mathbb{C}(S, \mathcal{G}) = \Big\{\{C_{(1)}, \ldots, C_{(\nu(S)+1)}\} \mid C_{(1)} \in \mathcal{C}(\Gamma^S_1 \cup S), \ldots, C_{(\nu(S)+1)} \in \mathcal{C}(\Gamma^S_{\nu(S)+1} \cup S)\Big\}.$$

Then we have the following proposition.

Proposition 2.1. *For any $\tilde{\mathcal{C}}_S \in \mathbb{C}(S, \mathcal{G})$, there exists a perfect sequence of cliques in $\mathcal{G}$ such that the first $\nu(S) + 1$ elements of the sequence are $\tilde{\mathcal{C}}_S$.*

We use the following lemma in order to prove Proposition 2.1.

Lemma 2.2. *Let δ be a simplicial vertex in $\mathcal{G}$ and C_δ be the clique which includes δ. Suppose $|\mathcal{C}| = K$. Then $|\mathcal{C}(\Delta \setminus \delta)| = K - 1$ if and only if there exists a clique $C' \in \mathcal{C} \setminus \{C_\delta\}$ such that $C' \supset (C_\delta \setminus \delta)$.*

Proof. From Lemma A.1 in Hara and Takemura[10], we note that C_δ is unique in $\mathcal{C}$. Suppose that $|\mathcal{C}(\Delta \setminus \delta)| = K - 1$. From the fact that $\Delta \setminus \{\delta\} \supseteq \Delta(\mathcal{C} \setminus \{C_\delta\})$, it follows that $\mathcal{C}(\Delta \setminus \delta) = \mathcal{C} \setminus \{C_\delta\}$. Hence

$$C_\delta \setminus \delta = \bigcup_{C \in \mathcal{C} \setminus \{C_\delta\}} C \cap C_\delta.$$

$\mathcal{G}(C_\delta \setminus \delta)$ is complete. Thus from the maximality of cliques there exists $C' \in \mathcal{C} \setminus \{C_\delta\}$ such that $C' \supset C_\delta \setminus \delta$.

If we suppose that there exists $C' \in \mathcal{C} \setminus \{C_\delta\}$ such that $C' \supset C_\delta \setminus \delta$, then $\mathcal{C}(\Delta \setminus \{\delta\}) = \mathcal{C} \setminus \{C_\delta\}$. Hence $|\mathcal{C}(\Delta \setminus \{\delta\})| = |\mathcal{C} \setminus \{C_\delta\}| = K - 1$. Then we can complete the proof. □

Proof of Proposition 2.1. We prove this lemma by induction on the number of vertices $J = |\Delta|$. For $J \leq 3$ the statement is trivial. Suppose then $J > 4$ and assume that the lemma holds up to J.

Suppose that $\mathcal{G}$ has $J + 1$ vertices and that $|\mathcal{C}| = K$. If $\mathcal{G}(\Delta(\tilde{\mathcal{C}}_S)) = \mathcal{G}$, the lemma is obvious. Thus we assume that $\mathcal{G}(\Delta(\tilde{\mathcal{C}}_S)) \neq \mathcal{G}$. From Lemma B.1 in Hara and Takemura[10], if $\mathcal{G}(\Delta(\tilde{\mathcal{C}}_S)) \neq \mathcal{G}$, $\Delta \setminus \Delta(\tilde{\mathcal{C}}_S)$ includes at least one simplicial vertex in $\mathcal{G}$. Denote it by $\delta \in \Delta \setminus \Delta(\tilde{\mathcal{C}}_S)$. From Lemma A.1 in Hara and Takemura[9], δ must be a member of exactly one clique in $\mathcal{C} \setminus \mathcal{C}(\Delta(\tilde{\mathcal{C}}_S))$, say C_δ. We note that $\mathcal{C}(\Delta \setminus \delta)$ satisfies

$$\mathcal{C}(\Delta \setminus \delta) = \mathcal{C} \setminus \{C_\delta\} \quad \text{or} \quad \mathcal{C}(\Delta \setminus \delta) = (\mathcal{C} \setminus \{C_\delta\}) \cup \{C_\delta \setminus \delta\}.$$

We first consider the case $\mathcal{C}(\Delta \setminus \delta) = \mathcal{C} \setminus \{C_\delta\}$. The inductive assumption implies that there exists a perfect sequence of the cliques of $\mathcal{G}(\Delta \setminus \delta)$ such that the first $\nu(S) + 1$ elements of the sequence are $\tilde{\mathcal{C}}_S$. Denote it by $C_1, \ldots, C_{K-1}$. From Lemma 2.2 $C_1, \ldots, C_{K-1}, C_\delta$ satisfies the running intersection property and thus it is perfect.

Next we consider the case where $\mathcal{C}(\Delta \setminus \delta) = (\mathcal{C} \setminus \{C_\delta\}) \cup \{C_\delta \setminus \delta\}$. The inductive assumption implies that the cliques in $\mathcal{G}(\Delta(\mathcal{C} \setminus \{C_\delta\}))$ admit a perfect sequence $C_1, \ldots, C_K$ with $\{C_1, \ldots, C_{\nu(S)+1}\} = \tilde{\mathcal{C}}_S$. There exists a clique C_k, $\nu(S) + 1 < k \leq K$ such that $C_k = C_\delta \setminus \delta$. If we replace C_k with C_δ, we can obtain a perfect sequence of the cliques in $\mathcal{G}$ with the desired property. □

3 Bayes estimators of the means in the decomposable models

3.1 Bayes estimators in the saturated model

In this section we first summarize briefly the Bayes estimation in the saturated model. The saturated model can be expressed as follows,

$$x(i) \sim \mathrm{Po}(\lambda(i)), \quad \lambda(i) = \lambda\alpha(i), \quad i = 1, \ldots, I,$$

$$\sum_{i=1}^{I} \alpha(i) = 1, \quad \alpha(i) \geq 0.$$

We write $\boldsymbol{\alpha} = (\alpha(1), \ldots, \alpha(I))'$. Clevenson and Zidek[5] and Ghosh and Parsian[7] introduced the class of priors as follows,

$$\pi(\lambda, \boldsymbol{\alpha})\mathrm{d}\lambda\mathrm{d}\boldsymbol{\alpha} = m(\lambda)\mathrm{d}\lambda\mathrm{d}\boldsymbol{\alpha},$$

where

$$m(\lambda) = \int_0^\infty t^I \lambda^{I-1} \exp(-t\lambda)\eta(t)\mathrm{d}t$$

and $\mathrm{d}\boldsymbol{\alpha}$ denotes the uniform probability measure on the simplex $\{\boldsymbol{\alpha} \mid \sum_{i=1}^{I} \alpha(i) = 1, \alpha(i) \geq 0\}$. The resulting Bayes estimator for the loss (2) is expressed by

$$\hat{\boldsymbol{\lambda}} = \begin{cases} \Psi(x^+) \cdot \dfrac{\boldsymbol{x}}{x^+ + I - 1}, & \text{for} \quad x^+ > 0 \\ \mathbf{0}, & \text{for} \quad x^+ = 0 \end{cases} \tag{7}$$

where $x^+ = \sum_{i=1}^{I} x(i)$ and $\Psi(\cdot)$ is defined on $x > 0$ as

$$\Psi(x) = \frac{\int_0^\infty \exp(-\lambda)\lambda^x m(\lambda)\mathrm{d}\lambda}{\int_0^\infty \exp(-\lambda)\lambda^{x-1} m(\lambda)\mathrm{d}\lambda}. \tag{8}$$

We note that $\Psi(x^+)$ is the Bayes estimator of λ with respect to the prior $m(\lambda)$. Clevenson and Zidek[5] and Ghosh and Parsian[7] derived conditions on $\eta(\cdot)$ to dominate $\boldsymbol{x}$ under the loss (2). They also introduced a convenient explicit form of $\eta(\cdot)$ like

$$\eta(t) \propto t^{\alpha-1}(1+t)^{-(\alpha+\beta)}. \tag{9}$$

$\Psi(\cdot)$ for this estimator is

$$\Psi(x) = \frac{x^+ + \beta - 1}{x^+ + \alpha + \beta + I - 1} \cdot (x^+ + I - 1).$$

Clevenson and Zidek[5] considered the estimator with $-1 \leq \alpha$ and $\beta = 1$. Then $\eta(\cdot)$ and $\Psi(\cdot)$ are expressed by

$$\eta(t) = t^{\alpha-1}(1+t)^{-(\alpha+1)}, \quad \Psi(x) = \frac{x}{x + \alpha + I} \cdot (x + I - 1).$$

They showed that this class with $\alpha \leq I - 2$ and $\beta = 1$ dominates $\boldsymbol{x}$. The prior is proper for $\alpha > 0$ but improper for $-1 \leq \alpha \leq 0$. Brown and Hwang[3] showed that the estimators with $-1 \leq \alpha \leq 0$ are also admissible.

Ghosh and Parsian[7] enlarged the class of Clevenson and Zidek[5] and showed that the estimator with $0 < \alpha \leq I - 2$ and $\beta > 0$ is proper Bayes and dominates $\boldsymbol{x}$. Johnstone[14] showed that the prior with $-1 < \alpha \leq 0$ and $\beta > 0$ is improper but the resulting estimator improves on $\boldsymbol{x}$.

3.2 Bayes estimators in the decomposable model

In this section we generalize the argument on the Bayes estimation in the saturated model to decomposable models and introduce two types of prior measures and the corresponding Bayes estimators. Let $\mathrm{d}\boldsymbol{\alpha}(S_1)$ and $\mathrm{d}\boldsymbol{\beta}(i_{S_k})$ denote the uniform probability measure on the simplex $\{\alpha(i_{S_1}), i_{S_1} \in \mathcal{I}_{S_1}\}$ and $\{\beta(i'_{C_k}), i'_{C_k} \in \mathcal{I}_{C_k}, i'_{S_k} = i_{S_k}\}$, respectively. Define $\lambda(i_{S_1})$ as $\lambda(i_{S_1}) = \lambda\alpha(i_{S_1})$. Then the prior measures which we consider in this article are expressed as the following form,

$$\pi(\lambda, \boldsymbol{\alpha}(S_1), \boldsymbol{\beta})\mathrm{d}\lambda\mathrm{d}\boldsymbol{\alpha}(S_1)\mathrm{d}\boldsymbol{\beta} = m(\lambda)\mathrm{d}\lambda\mathrm{d}\boldsymbol{\alpha}(S_1) \cdot \prod_{k=1}^{K} \prod_{i_{S_k} \in \mathcal{I}_{S_k}} \mathrm{d}\boldsymbol{\beta}(i_{S_k}), \tag{10}$$

$$\pi_{S_1}(\boldsymbol{\lambda}(S_1), \boldsymbol{\beta})\mathrm{d}\boldsymbol{\lambda}(S_1)\mathrm{d}\boldsymbol{\beta} = \prod_{i_{S_1} \in \mathcal{I}_{S_1}} m(\lambda(i_{S_1}))\mathrm{d}\lambda(i_{S_1}) \cdot \prod_{k=1}^{K} \prod_{i_{S_k} \in \mathcal{I}_{S_k}} \mathrm{d}\boldsymbol{\beta}(i_{S_k}), \tag{11}$$

where

$$\boldsymbol{\lambda}(S_1) = \{\lambda(i_{S_1}), i_{S_1} \in \mathcal{I}_{S_1}\}, \quad \mathrm{d}\boldsymbol{\beta} = \{\mathrm{d}\boldsymbol{\beta}(i_{S_k}), i_{S_k} \in \mathcal{I}_{S_k}, k = 1, \dots, K\},$$

$$m(\lambda) = \int_0^\infty t^\mu \lambda^{\mu-1} \exp(-t\lambda)\eta(t)\mathrm{d}t, \qquad \mu > 0. \tag{12}$$

In the saturated model $\mu = I$, which is the degrees of freedom of the model. Here we take μ to be a positive constant. Denote the corresponding Bayes estimators with respect to the prior measure (10) and (11) by $\hat{\boldsymbol{\lambda}}(\pi) = \{\hat{\lambda}(i, \pi)\}$ and $\hat{\boldsymbol{\lambda}}(\pi_{S_1}) = \{\hat{\lambda}(i, \pi_{S_1})\}$, respectively. Then $\hat{\lambda}(i, \pi)$ and $\hat{\lambda}(i, \pi_{S_1})$ are obtained by minimizing the posterior expected loss as follows,

$$\hat{\lambda}(i, \pi) = \hat{\lambda}^{ML}(i) \cdot \frac{\Psi(x^+)}{x^+ + I_{S_1} - 1} \cdot \prod_{k=1}^{K} \frac{x(i_{S_k})}{x(i_{S_k}) + I_{R_k} - 1}, \tag{13}$$

$$\hat{\lambda}(i, \pi_{S_1}) = \hat{\lambda}^{ML}(i) \cdot \frac{\Psi(x(i_{S_1}))}{x(i_{S_1})} \prod_{k=1}^{K} \frac{x(i_{S_k})}{x(i_{S_k}) + I_{R_k} - 1}, \tag{14}$$

where $x^+ = x(i_\emptyset) = \sum_{i \in \mathcal{I}} x(i)$ and $\Psi(x)$ is defined in (8) with $m(\cdot)$ in (12) for $x > 0$ and we define $\Psi(0) \equiv 0$. In Section 4 and Section 5 we derive conditions on $\eta(\cdot)$ for $\hat{\boldsymbol{\lambda}}(\pi)$ and $\hat{\boldsymbol{\lambda}}(\pi_{S_1})$ to improve on $\hat{\boldsymbol{\lambda}}^{ML}$.

We consider the explicit form of $\eta(\cdot)$ in (9) also for decomposable models. Now we assume that $\alpha + \mu > 0$ and $\beta > 0$. Denote the corresponding Bayes estimators by $\hat{\boldsymbol{\lambda}}^{\alpha,\beta}(\pi)$ and $\hat{\boldsymbol{\lambda}}^{\alpha,\beta}(\pi_{S_1})$, respectively. Then $\Psi(\cdot)$ for $\hat{\boldsymbol{\lambda}}^{\alpha,\beta}(\pi)$ and $\hat{\boldsymbol{\lambda}}^{\alpha,\beta}(\pi_{S_1})$ is expressed by

$$\Psi^{\alpha,\beta}(x) = \begin{cases} \dfrac{x + \beta - 1}{x + \alpha + \beta + \mu - 1} \cdot (x + \mu - 1), & \text{for} \quad x > 0 \\ 0, & \text{for} \quad x = 0. \end{cases}$$

When $\alpha > 0$ and $\beta > 0$, the estimators correspond to the one by Ghosh and Parsian[7]. Denote the estimators by $\hat{\boldsymbol{\lambda}}^{GP}(\pi)$ and $\hat{\boldsymbol{\lambda}}^{GP}(\pi_{S_1})$. In this case the priors are proper. Hence $\hat{\boldsymbol{\lambda}}^{GP}(\pi)$ and $\hat{\boldsymbol{\lambda}}^{GP}(\pi_{S_1})$ are admissible.

When $\alpha \geq -1$ and $\beta = 1$, the estimators correspond to the one by Clevenson and Zidek[5]. Denote the estimators by $\hat{\boldsymbol{\lambda}}^{CZ}(\pi)$ and $\hat{\boldsymbol{\lambda}}^{CZ}(\pi_{S_1})$. Then $\Psi(\cdot)$ for $\hat{\boldsymbol{\lambda}}^{CZ}(\pi)$ and $\hat{\boldsymbol{\lambda}}^{CZ}(\pi_{S_1})$ is expressed by

$$\Psi(x) = \frac{x}{x+\alpha+\mu} \cdot (x+\mu-1).$$

The priors are proper for $\alpha > 0$. Thus $\hat{\boldsymbol{\lambda}}^{CZ}(\pi)$ and $\hat{\boldsymbol{\lambda}}^{CZ}(\pi_{S_1})$ are admissible. However the priors become improper for $-1 \leq \alpha \leq 0$.

We also consider the estimator with $\beta = -\alpha$ and $\gamma = -(\alpha+1)$. Then $m(\cdot)$ and $\Psi(\cdot)$ are

$$m(\lambda) = \lambda^{\gamma}, \quad \Psi(x) = x + \gamma,$$

respectively. In this case the priors become improper. Denote the resulting Bayes estimators by $\hat{\boldsymbol{\lambda}}^{\gamma}(\pi)$ and $\hat{\boldsymbol{\lambda}}^{\gamma}(\pi_{S_1})$.

In the saturated model the generalized Bayes estimator (7) with $\beta = -\alpha$ and $\gamma = -(\alpha+1)$ belongs to the class of estimators improving on $\boldsymbol{x}$ derived by Clevenson and Zidek[5] when $-1 < \gamma < I-1$. Following Johnstone[14], the estimator is admissible when $-1 < \gamma \leq 0$. This class of priors is a subclass of the one discussed in Komaki[15] for the setting of simultaneous prediction. He showed that the Bayesian predictive distribution with respect to this prior with $-1 < \gamma \leq 0$ is admissible and dominates the one with respect to the Jeffreys prior under the Kullback-Leibler loss.

In the following sections sufficient conditions on α, β, γ and μ are presented for $\boldsymbol{\lambda}^{GP}$, $\boldsymbol{\lambda}^{CZ}$ and $\boldsymbol{\lambda}^{\gamma}$ to improve on $\boldsymbol{\lambda}^{ML}$.

4 Improved Bayes estimation of the Means in the decomposable models with one minimal vertex separator

In the following sections we derive some sufficient conditions on the class of estimators introduced in the previous section to dominate $\boldsymbol{\lambda}^{ML}$. In this section we assume that the model has only one minimal vertex separator S_1. In the next section we extend the results obtained here to the general Poisson decomposable graphical models (3).

The model (3) with one minimal vertex separator S_1 is rewritten by

$$x(i) \sim \mathrm{Po}(\lambda(i)), \quad \lambda(i) = \lambda\alpha(i_{S_1}) \prod_{k=1}^{\nu(S_1)+1} \beta(i_{C_k}), \quad \beta(i_{C_k}) = \frac{\alpha(i_{C_k})}{\alpha(i_{S_1})}, \tag{15}$$

and $\hat{\lambda}^{ML}(i)$ in (5), $\hat{\lambda}(i,\pi)$ in (13) and $\hat{\lambda}(i,\pi_{S_1})$ in (14) for the model (15) are rewritten by

$$\hat{\lambda}^{ML}(i) = \frac{\prod_{k=1}^{\nu(S_1)+1} x(i_{C_k})}{x(i_{S_1})^{\nu(S_1)}},$$

$$\hat{\lambda}(i,\pi) = \hat{\lambda}^{ML}(i) \cdot \frac{\Psi(x^+)}{x^+ + I_{S_1} - 1} \prod_{k=1}^{\nu(S_1)+1} \frac{x(i_{S_1})}{x(i_{S_1}) + I_{R_k} - 1},$$

$$\hat{\lambda}(i,\pi_{S_1}) = \hat{\lambda}^{ML}(i)\cdot\frac{\Psi(x(i_{S_1}))}{x(i_{S_1})}\prod_{k=1}^{\nu(S_1)+1}\frac{x(i_{S_1})}{x(i_{S_1})+I_{R_k}-1},$$

respectively.

We first derive the condition on $\hat{\boldsymbol{\lambda}}(\pi)$ to improve on $\hat{\boldsymbol{\lambda}}^{ML}$. By using the identity (6), we can obtain the UMVUE of the difference of the risks between $\hat{\boldsymbol{\lambda}}^{ML}$ and $\hat{\boldsymbol{\lambda}}(\pi)$. By evaluating it, we can obtain the following results.

Theorem 4.1. *Suppose that $\Psi(\cdot)$ satisfies*

$$0 \le \Psi(x^+ + 1)\cdot\frac{(x^+ + I_{S_1})^{\nu(S_1)}}{\prod_{k=1}^{\nu(S_1)+1}(x^+ + I_{S_1}I_{R_k})} \le 1, \tag{16}$$

$$\begin{aligned} x(i_{S_1})&\left(1-\frac{\Psi(x^+)}{x^+ + I_{S_1}-1}\cdot\prod_{k=1}^{\nu(S_1)+1}\frac{x(i_{S_1})}{x(i_{S_1})+I_{R_k}-1}\right)\\ &\le (x(i_{S_1})+1)\left(1-\frac{\Psi(x^+ + 1)}{x^+ + I_{S_1}}\cdot\prod_{k=1}^{\nu(S_1)+1}\frac{x(i_{S_1})+1}{x(i_{S_1})+I_{R_k}}\right) \end{aligned} \tag{17}$$

and that $\eta(\cdot)$ satisfies

(i) $\overline{\lim}_{t\to\infty} t^{1+\epsilon}\eta(t) < \infty$ *for some* $\epsilon > 0$*;*

(ii) $\eta(t)$ *is differentiable on* $t > 0$*;*

(iii) $\sup_{t>0} t\eta'(t)/\eta(t) \le \sum_{k=1}^{\nu(S_1)+1} I_{S_1}(I_{R_k}-1) - (I_{S_1}+1)$*;*

Then $\hat{\boldsymbol{\lambda}}(\pi)$ improves on $\hat{\boldsymbol{\lambda}}^{ML}$ under the loss (2).

We use the following lemma to prove the theorem.

Lemma 4.1. *Let a_k, $k = 1,\ldots,K$ be positive constants. Define $g_1(y)$ and $g_2(y)$ by*

$$g_1(y) = \frac{\prod_{k=1}^{K}(y+a_k)}{y^{K-1}}, \quad g_2(y) = \frac{y^{K+1}}{\prod_{k=1}^{K}(y+a_k)},$$

respectively. Then $g_1(y)$ and $g_2(y)$ are convex on $y > 0$.

Proof. $g_1(y)$ can be rewritten by

$$g_1(y) = y + b_0 + \sum_{k=1}^{K-1}\frac{b_k}{y^k}, \quad b_k > 0, \quad k = 1,\ldots,K.$$

Hence

$$\frac{\mathrm{d}^2 g_1(y)}{\mathrm{d}y^2} = \sum_{k=1}^{K-1}\frac{k(k+1)b_k}{y^{k+2}} > 0.$$

for $y > 0$. Thus $g_1(y)$ is convex on $y > 0$.

$\mathrm{d}g_2(y)/\mathrm{d}y$ is

$$\frac{\mathrm{d}g_2(y)}{\mathrm{d}y} = \frac{y^K}{\prod_{k=1}^K (y+a_k)} \left(1 + \sum_{k=1}^K \frac{a_k}{y+a_k}\right).$$

Hence we have

$$\frac{\mathrm{d}^2 g_2(y)}{\mathrm{d}y^2} = \frac{y^{K-1}}{\prod_{k=1}^K (y+a_k)} \left\{\sum_{k=1}^K \frac{a_k^2}{(y+a_k)^2} + \left(\sum_{k=1}^K \frac{a_k}{y+a_k}\right)^2\right\} > 0$$

for $y > 0$. Thus $g_2(y)$ is also convex on $y > 0$. □

Proof of Theorem 4.1. By using the identity (6), the difference between two risk functions of $\hat{\boldsymbol{\lambda}}^{ML}$ and $\hat{\boldsymbol{\lambda}}(\pi)$ is expressed by

$$\begin{aligned}
&\mathrm{R}(\boldsymbol{\lambda}, \hat{\boldsymbol{\lambda}}^{ML}) - \mathrm{R}(\boldsymbol{\lambda}, \hat{\boldsymbol{\lambda}}(\pi)) \\
&\quad = \mathrm{E}[\mathrm{L}(\boldsymbol{\lambda}, \hat{\boldsymbol{\lambda}}^{ML}) - \mathrm{L}(\boldsymbol{\lambda}, \hat{\boldsymbol{\lambda}}(\pi))] \\
&\quad = \sum_{i\in\mathcal{I}} \mathrm{E}\left[\frac{1}{\lambda(i)}\left\{(\hat{\lambda}^{ML}(i) - \lambda(i))^2 - (\hat{\lambda}(i,\pi) - \lambda(i))^2\right\}\right] \\
&\quad = \sum_{i\in\mathcal{I}} \mathrm{E}\left[\frac{2\hat{\lambda}^{ML}(i)}{\lambda(i)}\left(1 - \frac{\Psi(x^+)}{x^+ + I_{S_1} - 1}\prod_{k=1}^{\nu(S_1)+1}\frac{x(i_{S_1})}{x(i_{S_1}) + I_{R_k} - 1}\right)\left(\hat{\lambda}^{ML}(i) - \lambda(i)\right)\right. \\
&\qquad\qquad \left. - \frac{1}{\lambda(i)}\left\{\hat{\lambda}^{ML}(i)\left(1 - \frac{\Psi(x^+)}{x^+ + I_{S_1} - 1}\prod_{k=1}^{\nu(S_1)+1}\frac{x(i_{S_1})}{x(i_{S_1}) + I_{R_k} - 1}\right)\right\}^2\right] \\
&\quad = \sum_{i\in\mathcal{I}} \mathrm{E}\left[\frac{2\prod_{k=1}^{\nu(S_1)+1}(x(i_{C_k})+1)}{(x(i_{S_1})+1)^{\nu(S_1)}}\left(1 - \frac{\Psi(x^+ + 1)}{x^+ + I_{S_1}}\prod_{k=1}^{\nu(S_1)+1}\frac{x(i_{S_1})+1}{x(i_{S_1}) + I_{R_k}}\right)\right. \\
&\qquad\qquad - \frac{2\prod_{k=1}^{\nu(S_1)+1} x(i_{C_k})}{x(i_{S_1})^{\nu(S_1)}}\left(1 - \frac{\Psi(x^+)}{x^+ + I_{S_1} - 1}\prod_{k=1}^{\nu(S_1)+1}\frac{x(i_{S_1})}{x(i_{S_1}) + I_{R_k} - 1}\right) \\
&\qquad\qquad \left. - \frac{\prod_{k=1}^{\nu(S_1)+1}(x(i_{C_k})+1)}{(x(i_{S_1})+1)^{\nu(S_1)}}\left(1 - \frac{\Psi(x^+ + 1)}{x^+ + I_{S_1}}\prod_{k=1}^{\nu(S_1)+1}\frac{x(i_{S_1})+1}{x(i_{S_1}) + I_{R_k}}\right)^2\right] \\
&\quad = \sum_{i_{S_1}\in\mathcal{I}_{S_1}} \mathrm{E}\left[\frac{\prod_{k=1}^{\nu(S_1)+1}(x(i_{S_1}) + I_{R_k})}{(x(i_{S_1})+1)^{\nu(S_1)}} - \left(\frac{\Psi(x^+ + 1)}{x^+ + I_{S_1}}\right)^2 \frac{(x(i_{S_1})+1)^{\nu(S_1)+2}}{\prod_{k=1}^{\nu(S_1)+1}(x(i_{S_1}) + I_{R_k})}\right. \\
&\qquad\qquad \left. - 2x(i_{S_1})\left(1 - \frac{\Psi(x^+)}{x^+ + I_{S_1} - 1}\prod_{k=1}^{\nu(S_1)+1}\frac{x(i_{S_1})}{x(i_{S_1}) + I_{R_k} - 1}\right)\right]. \qquad (18)
\end{aligned}$$

This implies that

$$
\begin{aligned}
&\hat{\mathrm{Rd}}(\hat{\boldsymbol{\lambda}}(\pi)) \\
&= \sum_{i_{S_1}\in\mathcal{I}_{S_1}} \left\{ \frac{\prod_{k=1}^{\nu(S_1)+1}(x(i_{S_1})+I_{R_k})}{(x(i_{S_1})+1)^{\nu(S_1)}} - \left(\frac{\Psi(x^++1)}{x^++I_{S_1}}\right)^2 \frac{(x(i_{S_1})+1)^{\nu(S_1)+2}}{\prod_{k=1}^{\nu(S_1)+1}(x(i_{S_1})+I_{R_k})} \right. \\
&\qquad \left. -2x(i_{S_1})\left(1-\frac{\Psi(x^+)}{x^++I_{S_1}-1}\prod_{k=1}^{\nu(S_1)+1}\frac{x(i_{S_1})}{x(i_{S_1})+I_{R_k}-1}\right)\right\} \qquad (19)
\end{aligned}
$$

is the UMVUE of $\mathrm{R}(\boldsymbol{\lambda},\hat{\boldsymbol{\lambda}}^{ML})-\mathrm{R}(\boldsymbol{\lambda},\hat{\boldsymbol{\lambda}}(\pi))$.

From Lemma 4.1 with $y=x(i_{S_1})+1$, $a_k=I_{R_k}$ and $K=\nu(S_1)+1$ both

$$
\frac{\prod_{k=1}^{\nu(S_1)+1}(x(i_{S_1})+I_{R_k})}{(x(i_{S_1})+1)^{\nu(S_1)}} \quad \text{and} \quad \frac{(x(i_{S_1})+1)^{\nu(S_1)+2}}{\prod_{k=1}^{\nu(S_1)+1}(x(i_{S_1})+I_{R_k})}
$$

are convex on $x(i_{S_1})$. Thus from Jensen's inequality

$$
\sum_{i_{S_1}\in\mathcal{I}_{S_1}} \frac{\prod_{k=1}^{\nu(S_1)+1}(x(i_{S_1})+I_{R_k})}{(x(i_{S_1})+1)^{\nu(S_1)}} \geq \frac{\prod_{k=1}^{\nu(S_1)+1}(x^++I_{S_1}I_{R_k})}{(x^++I_{S_1})^{\nu(S_1)}}, \qquad (20)
$$

$$
\sum_{i_{S_1}\in\mathcal{I}_{S_1}} \frac{(x(i_{S_1})+1)^{\nu(S_1)+2}}{\prod_{k=1}^{\nu(S_1)+1}(x(i_{S_1})+I_{R_k})} \geq \frac{(x^++I_{S_1})^{\nu(S_1)+2}}{\prod_{k=1}^{\nu(S_1)+1}(x^++I_{S_1}I_{R_k})}. \qquad (21)
$$

Hence we have from (17), (20) and (21)

$$
\begin{aligned}
\mathrm{Rd}(\hat{\boldsymbol{\lambda}}(\pi)) &\geq \frac{\prod_{k=1}^{\nu(S_1)+1}(x^++I_{S_1}I_{R_k})}{(x^++I_{S_1})^{\nu(S_1)}} - 2(x^++I_{S_1}) \\
&\quad + \frac{(x^++I_{S_1})^{\nu(S_1)+2}}{\prod_{k=1}^{\nu(S_1)+1}(x^++I_{S_1}I_{R_k})}\cdot\frac{\Psi(x^++1)}{x^++I_{S_1}}\cdot\left(2-\frac{\Psi(x^++1)}{x^++I_{S_1}}\right) \\
&= \left(\frac{\prod_{k=1}^{\nu(S_1)+1}(x^++I_{S_1}I_{R_k})}{(x^++I_{S_1})^{\nu(S_1)}} - \Psi(x^++1)\right) \\
&\quad \times\left\{1-\frac{(x^++I_{S_1})^{\nu(S_1)+1}}{\prod_{k=1}^{\nu(S_1)+1}(x^++I_{S_1}I_{R_k})}\left(2-\frac{\Psi(x^++1)}{x^++I_{S_1}}\right)\right\}. \qquad (22)
\end{aligned}
$$

From (8) and the assumptions (i), (ii) and (iii) of the theorem

$$
\begin{aligned}
\frac{\Psi(x^++1)}{x^++\mu} &= \frac{\int_0^\infty (1+t)^{-(x^++\mu+1)}t^\mu\eta(t)\mathrm{d}t}{\int_0^\infty (1+t)^{-(x^++\mu)}t^\mu\eta(t)\mathrm{d}t} \\
&= \frac{x^+-1}{x^++\mu} - \frac{\int_0^\infty (t\eta'(t)/\eta(t))(1+t)^{-(x^++\mu)}t^\mu\eta(t)\mathrm{d}t}{\int_0^\infty (1+t)^{-(x^++\mu)}t^\mu\eta(t)\mathrm{d}t}\cdot\frac{1}{x^++\mu} \\
&\geq \frac{x^+-1}{x^++\mu} - \frac{\sum_{k=1}^{\nu(S_1)+1} I_{S_1}(I_{R_k}-1)-(I_{S_1}+1)}{x^++\mu}.
\end{aligned}
$$

Hence we have

$$
\begin{aligned}
2 - \frac{\Psi(x^+ + 1)}{x^+ + I_{S_1}} &\le 2 - \frac{x^+ - 1}{x^+ + I_{S_1}} + \frac{\sum_{k=1}^{\nu(S_1)+1} I_{S_1}(I_{R_k} - 1) - (I_{S_1} + 1)}{x^+ + I_{S_1}} \\
&= \frac{1}{x^+ + I_{S_1}} \left(x^+ + I_{S_1} + \sum_{k=1}^{\nu(S_1)+1} I_{S_1}(I_{R_k} - 1) \right) \\
&\le \frac{\prod_{k=1}^{\nu(S_1)+1} (x^+ + I_{S_1} I_{R_k})}{(x^+ + I_{S_1})^{\nu(S_1)+1}}. \qquad (23)
\end{aligned}
$$

From (16) and (23) the right hand side of (22) is always nonnegative. Then we can complete the proof. □

By utilizing Theorem 4.1, we can derive the condition on α, β and μ of $\hat{\boldsymbol{\lambda}}^{\alpha,\beta}(\pi)$ to improve on $\hat{\boldsymbol{\lambda}}^{ML}$. We use the following lemma to derive the conditions.

Lemma 4.2. *Let $\psi(y)$ be a nondecreasing function such that $\psi(y) \ge 0$ and $y(1 - \psi(y))$ is nondecreasing on $y \ge 1$. Suppose that $a_k \ge b_k \ge 0$, $k = 1, \dots, K$ and define $\phi_k(y)$, $k = 0, 1, \dots, K$ as*

$$
\phi_0(y) = y(1 - \psi(y)), \quad \phi_k(y) = y \left(1 - \psi(y) \prod_{j=1}^{k} \frac{y + b_j}{y + a_j} \right) \quad \textit{for} \quad k = 1, \dots, K.
$$

Then $\phi_k(y)$ is nondecreasing on $y \ge 1$.

Proof. We prove this lemma by induction on k. When $k = 0$, the lemma is obvious. Suppose $k > 0$ and assume that the lemma holds up to k. Then

$$
\phi_{k+1}(y) = y \left(1 - \psi(y) \prod_{j=1}^{k} \frac{y + b_j}{y + a_j} \right) + \psi(y) \prod_{j=1}^{k} \frac{y + b_j}{y + a_j} \cdot \frac{(a_{k+1} - b_{k+1}) y}{y + a_{k+1}}
$$

From the inductive assumption the first term is nondecreasing. The second term is obviously nondecreasing. Thus the proof is completed. □

Define $\psi^{\alpha,\beta}(x)$ on $x \ge 1$ by

$$
\psi^{\alpha,\beta}(x) \equiv \frac{\Psi^{\alpha,\beta}(x)}{x + I_{S_1} - 1} = \frac{x + \beta - 1}{x + \alpha + \beta + \mu - 1} \cdot \frac{x + \mu - 1}{x + I_{S_1} - 1}.
$$

Since $\Psi^{\alpha,\beta}(0)$ is defined to be 0 and $x^+ = 0$ implies $x(i_{S_1}) = 0$, if $\psi^{\alpha,\beta}(x^+)$ is nondecreasing and $0 \le \psi^{\alpha,\beta}(x^+) \le 1$ on $x^+ \ge 1$, then (16) and (17) with $\Psi(\cdot) = \Psi^{\alpha,\beta}(\cdot)$ are satisfied from Lemma 4.2.

$\eta(t)$ for $\hat{\boldsymbol{\lambda}}^{\alpha,\beta}(\pi)$ is (9). Since $\alpha - 1 < \alpha + \beta$, the conditions (i) and (ii) of Theorem 4.1 are satisfied. $h(t) = t\eta'(t)/\eta(t)$ is expressed by

$$
h(t) = \alpha - 1 - \frac{t}{1 + t}(\alpha + \beta).
$$

Thus $\sup_{t>0} h(t)$ is

$$\sup_{t>0} h(t) = \begin{cases} \alpha - 1, & \text{if} \quad \alpha + \beta \geq 0, \\ -\beta - 1, & \text{if} \quad \alpha + \beta < 0. \end{cases}$$

If $\alpha + \beta \geq 0$, the condition (iii) is rewritten by $\alpha \leq \sum_{k=1}^{\nu(S_1)+1} I_{S_1}(I_{R_k} - 1) - I_{S_1}$. Since β is assumed to be positive, the condition (iii) is always satisfied when $\alpha + \beta < 0$. Noting that $\alpha + \beta < 0$ implies $\alpha < 0$, we can obtain the following results.

Corollary 4.1. *If α, β and μ satisfy*

(i) $\psi^{\alpha,\beta}(x)$ *is nondecreasing on* $x \geq 1$,

(ii) $0 \leq \psi^{\alpha,\beta}(x) \leq 1$ *for* $x \geq 1$,

(iii)

$$\alpha \leq \sum_{k=1}^{\nu(S_1)+1} I_{S_1}(I_{R_k} - 1) - I_{S_1}, \tag{24}$$

then $\hat{\boldsymbol{\lambda}}^{\alpha,\beta}(\pi)$ *improves on* $\hat{\boldsymbol{\lambda}}^{ML}$ *under the loss (2). If* $\alpha > 0$, $\hat{\boldsymbol{\lambda}}^{\alpha,\beta}(\pi)$ *is admissible.*

Remark 4.1. By using Corollary 4.1, we can derive conditions on $\hat{\boldsymbol{\lambda}}^{GP}(\pi)$, $\hat{\boldsymbol{\lambda}}^{CZ}(\pi)$ and $\hat{\boldsymbol{\lambda}}^{\gamma}(\pi)$ to improve on $\hat{\boldsymbol{\lambda}}^{ML}$. With respect to $\hat{\boldsymbol{\lambda}}^{GP}(\pi)$, both of α and β are positive. Hence $\min(\beta, \mu) \leq I_{S_1}$ and (24) is a sufficient condition on $\hat{\boldsymbol{\lambda}}^{GP}(\pi)$ to satisfy the conditions of Corollary 4.1 and then $\hat{\boldsymbol{\lambda}}^{GP}(\pi)$ improves on $\hat{\boldsymbol{\lambda}}^{ML}$. Since $\alpha \geq -1$ and $\beta = 1$ for $\hat{\boldsymbol{\lambda}}^{CZ}(\pi)$, (i) and (ii) of Corollary 4.1 are always satisfied. Then if (24) holds, $\hat{\boldsymbol{\lambda}}^{CZ}(\pi)$ improves on $\hat{\boldsymbol{\lambda}}^{ML}$. $\alpha + \beta = 0$ and $\gamma = -(\alpha + 1)$ for $\hat{\boldsymbol{\lambda}}^{\gamma}(\pi)$. Hence if $-1 < \gamma \leq I_{S_1} - 1$, the conditions of Corollary 4.1 hold and then $\hat{\boldsymbol{\lambda}}^{\gamma}(\pi)$ improves on $\hat{\boldsymbol{\lambda}}^{ML}$.

Next we derive the condition on $\hat{\boldsymbol{\lambda}}(\pi_{S_1})$ to improve on $\hat{\boldsymbol{\lambda}}^{ML}$. In the similar way to Theorem 4.1 we can obtain the following results.

Theorem 4.2. *Suppose that* $\Psi(\cdot)$ *satisfies*

$$0 \leq \Psi(x(i_{S_1}) + 1) \cdot \frac{(x(i_{S_1}) + 1)^{\nu(S_1)}}{\prod_{k=1}^{\nu(S_1)+1}(x(i_{S_1}) + I_{R_k})} \leq 1, \tag{25}$$

$$\begin{aligned} x(i_{S_1}) &\left(1 - \frac{\Psi(x(i_{S_1}))}{x(i_{S_1})} \cdot \prod_{k=1}^{\nu(S_1)+1} \frac{x(i_{S_1})}{x(i_{S_1}) + I_{R_k} - 1}\right) \\ &\leq (x(i_{S_1}) + 1) \left(1 - \frac{\Psi(x(i_{S_1}) + 1)}{x(i_{S_1}) + 1} \cdot \prod_{k=1}^{\nu(S_1)+1} \frac{x(i_{S_1}) + 1}{x(i_{S_1}) + I_{R_k}}\right) \end{aligned} \tag{26}$$

and that $\eta(\cdot)$ *satisfies*

(i) $\overline{\lim}_{t\to\infty} t^{1+\epsilon}\eta(t) < \infty$ *for some* $\epsilon > 0$;

(ii) $\eta(t)$ *is differentiable on* $t > 0$;

(iii) $\sup_{t>0} t\eta'(t)/\eta(t) \le \sum_{k=1}^{\nu(S_1)+1}(I_{R_k}-1)-2$;

Then $\hat{\boldsymbol{\lambda}}(\pi_{S_1})$ *improves on* $\hat{\boldsymbol{\lambda}}^{ML}$ *under the loss (2).*

Proof. By using the similar argument to the proof of Theorem 4.1, the UMVUE of the difference of the risks between $\hat{\boldsymbol{\lambda}}^{ML}$ and $\hat{\boldsymbol{\lambda}}(\pi_{S_1})$ is expressed by

$$
\begin{aligned}
&\hat{\mathrm{Rd}}(\hat{\boldsymbol{\lambda}}(\pi_{S_1}))\\
&= \sum_{i_{S_1}\in\mathcal{I}_{S_1}} \left\{ \frac{\prod_{k=1}^{\nu(S_1)+1}(x(i_{S_1})+I_{R_k})}{(x(i_{S_1})+1)^{\nu(S_1)}} - \Psi^2(x(i_{S_1})+1)\cdot\frac{(x(i_{S_1})+1)^{\nu(S_1)}}{\prod_{k=1}^{\nu(S_1)+1}(x(i_{S_1})+I_{R_k})} \right.\\
&\qquad \left. - 2x(i_{S_1})\left(1-\frac{\Psi(x(i_{S_1}))}{x(i_{S_1})}\prod_{k=1}^{\nu(S_1)+1}\frac{x(i_{S_1})}{x(i_{S_1})+I_{R_k}-1}\right)\right\}\\
&\ge \sum_{i_{S_1}\in\mathcal{I}_{S_1}} \left\{ \frac{\prod_{k=1}^{\nu(S_1)+1}(x(i_{S_1})+I_{R_k})}{(x(i_{S_1})+1)^{\nu(S_1)}} \right.\\
&\qquad - \Psi^2(x(i_{S_1})+1)\cdot\frac{(x(i_{S_1})+1)^{\nu(S_1)}}{\prod_{k=1}^{\nu(S_1)+1}(x(i_{S_1})+I_{R_k})}\\
&\qquad \left. - 2(x(i_{S_1})+1)\left(1-\frac{\Psi(x(i_{S_1})+1)}{x(i_{S_1})+1}\prod_{k=1}^{\nu(S_1)+1}\frac{(x(i_{S_1})+1)}{x(i_{S_1})+I_{R_k}}\right)\right\}\\
&= \sum_{i_{S_1}\in\mathcal{I}_{S_1}} \left(\frac{\prod_{k=1}^{\nu(S_1)+1}(x(i_{S_1})+I_{R_k})}{(x(i_{S_1})+1)^{\nu(S_1)}} - \Psi(x(i_{S_1})+1)\right)\\
&\qquad \times\left\{1-\frac{(x(i_{S_1})+1)^{\nu(S_1)+1}}{\prod_{k=1}^{\nu(S_1)+1}(x(i_{S_1})+I_{R_k})}\left(2-\frac{\Psi(x(i_{S_1})+1)}{x(i_{S_1})+1}\right)\right\}. \qquad (27)
\end{aligned}
$$

The inequality follows from (26). We have in the same way as (23)

$$
\begin{aligned}
2-\frac{\Psi(x(i_{S_1})+1)}{x(i_{S_1})+1} &\le 2-\frac{x(i_{S_1})-1}{x(i_{S_1})+1}+\frac{\sum_{k=1}^{\nu(S_1)+1}(I_{R_k}-1)-2}{x(i_{S_1})+1}\\
&= \frac{1}{x(i_{S_1})+1}\left(x(i_{S_1})+1+\sum_{k=1}^{\nu(S_1)+1}(I_{R_k}-1)\right)\\
&\le \frac{\prod_{k=1}^{\nu(S_1)+1}(x(i_{S_1})+I_{R_k})}{(x(i_{S_1})+1)^{\nu(S_1)+1}}, \qquad (28)
\end{aligned}
$$

From (25) and (28) the right hand side of (27) is always nonnegative under the conditions of the theorem. □

Corresponding to Corollary 4.1, we can also obtain the condition on $\hat{\boldsymbol{\lambda}}^{\alpha,\beta}(\pi_{S_1})$ to improve on $\hat{\boldsymbol{\lambda}}^{ML}(\pi_{S_1})$.

Corollary 4.2. *Suppose that the conditions (25) and (26) with $\Psi(\cdot) = \Psi^{\alpha,\beta}(\cdot)$ hold. If α satisfies*

$$\alpha \leq \sum_{k=1}^{\nu(S_1)+1} (I_{R_k} - 1) - 1, \tag{29}$$

$\hat{\boldsymbol{\lambda}}^{\alpha,\beta}(\pi)$ improves on $\hat{\boldsymbol{\lambda}}^{ML}$ under the loss (2). If $\alpha > 0$, $\hat{\boldsymbol{\lambda}}^{\alpha,\beta}(\pi)$ is admissible.

Remark 4.2. We can also obtain the improved estimators from the class $\hat{\boldsymbol{\lambda}}^{GP}(\pi_{S_1})$, $\hat{\boldsymbol{\lambda}}^{CZ}(\pi_{S_1})$ and $\hat{\boldsymbol{\lambda}}^{\gamma}(\pi_{S_1})$. Define I_R^* as $I_R^* = \max_{1\leq k\leq \nu(S_1)+1} I_{R_k}$. Then $\min(\beta,\mu) \leq I_R^*$ is a sufficient condition on $\Psi(\cdot)$ for $\hat{\boldsymbol{\lambda}}^{GP}(\pi_{S_1})$ to satisfy (25) and (26) from Lemma 4.2. Thus if $\min(\beta,\mu) \leq I_R^*$ and (29) is satisfied, $\hat{\boldsymbol{\lambda}}^{GP}(\pi_{S_1})$ improves on $\hat{\boldsymbol{\lambda}}^{ML}$. Since (25) and (26) are always satisfied for $\hat{\boldsymbol{\lambda}}^{CZ}(\pi_{S_1})$, (29) is a sufficient condition on $\hat{\boldsymbol{\lambda}}^{CZ}(\pi_{S_1})$ to improve on $\hat{\boldsymbol{\lambda}}^{ML}$. In the same way if $-1 < \gamma \leq I_R^* - 1$, $\hat{\boldsymbol{\lambda}}^{\gamma}(\pi_{S_1})$ improves on $\hat{\boldsymbol{\lambda}}^{ML}$.

Remark 4.3. Hara and Takemura[9] considered $\hat{\boldsymbol{\lambda}}^{CZ}(\pi_{S_1})$ in the model (15) with $|R_k| = 1$ and $|S_1| = 1$ or $S_1 = \emptyset$ and derived some conditions to improve the MLE. We note that $\lambda(i_{S_1}) = \lambda$ when $S_1 = \emptyset$. The results in Hara and Takemura[9] coincide with the one for $\hat{\boldsymbol{\lambda}}^{CZ}(\pi_{S_1})$ in Remark 4.2.

5 Extension to general decomposable graphical models

In this section we extend the results for the model with one minimal vertex separator (15) obtained in the previous section to general Poisson decomposable graphical models (3). For a perfect sequence $C_1, \ldots, C_K$ of the cliques in $\mathcal{G}$ and the corresponding minimax vertex separators $S_1 \equiv S_2, \ldots, S_K$, we can assume from Proposition 2.1 that $S_1 = \cdots = S_{\nu(S_1)+1}$ and $\{C_1, \ldots, C_{\nu(S_1)+1}\} \in \mathbb{C}(S_1, \mathcal{G})$. We first state a generalization of Theorem 4.1.

Theorem 5.1. *If $\Psi(\cdot)$ and $\eta(\cdot)$ satisfy*

$$\frac{\Psi(x^+)}{x^+ + I_{S_1} - 1} \leq \frac{\Psi(x^+ + 1)}{x^+ + I_{S_1}} \quad \textit{for} \quad x^+ \geq 1, \tag{30}$$

$$0 \leq \frac{\Psi(x^+ + 1)}{x^+ + I_{S_1}} \cdot \prod_{j=1}^{k} \frac{x(i_{S_j}) + 1}{x(i_{S_j}) + I_{R_j}} \leq 1, \tag{31}$$

for $\nu(S_1) + 1 \leq k \leq K$, $\hat{\boldsymbol{\lambda}}(\pi)$ improves on $\hat{\boldsymbol{\lambda}}^{ML}$ under the loss (2).

In order to prove this theorem, we use the following lemma.

Lemma 5.1. *For $I_{R_k} \geq 1$, $k = 1, \ldots, K$ and $0 \leq c \leq I_R^* - 1$, $I_R^* = \max_{1\leq k\leq \nu(S_1)+1} I_{R_k}$,*

$$\frac{(x(i_{S_1}) + c)x(i_{S_K})}{(x(i_{S_K}) + I_{R_K} - 1)^2} \cdot \prod_{k=1}^{K-1} \left(\frac{x(i_{C_k})}{x(i_{S_k}) + I_{R_k} - 1} \right)$$
$$\leq \frac{(x(i_{S_1}) + c + 1)(x(i_{S_K}) + 1)}{(x(i_{S_K}) + I_{R_K})^2} \cdot \prod_{k=1}^{K-1} \left(\frac{x(i_{C_k}) + 1}{x(i_{S_k}) + I_{R_k}} \right).$$

Proof. From the running intersection property of $C_1, C_2, \ldots, C_K$ there exists a sequence of the indices $k_1 < \cdots < k_J < k_{J+1} = K$ such that

$$S_{k_1} = S_1, \quad S_{k_2} \neq S_1, \quad C_{k_J} \supset S_K,$$
$$C_{k_j} \supset S_{k_{j+1}} \text{ for } \ k = 1, \ldots, J-1.$$

Define $\mathcal{K}_0$, $\mathcal{K}_1$ and $\mathcal{K}_2$ by

$$\mathcal{K}_0 = \{k_1, \ldots, k_J\}, \quad \mathcal{K}_1 = \{k \mid k \notin \mathcal{J}_0, k \le \nu(S_1)+1\}, \quad \mathcal{K}_2 = \{k \mid k \notin \mathcal{J}_0, k > \nu(S_1)+1\}.$$

Then by shifting the indices in $\mathcal{K}_0$

$$\begin{aligned}
&\frac{(x(i_{S_1})+c)x(i_{S_K})}{(x(i_{S_K})+I_{R_K}-1)^2} \cdot \prod_{k=1}^{K-1} \left(\frac{x(i_{C_k})}{x(i_{S_k})+I_{R_k}-1} \right) \\
&\quad = \frac{x(i_{S_K})}{x(i_{S_K})+I_{R_K}-1} \cdot \frac{x(i_{S_1})+c}{x(i_{S_1})+I_{R_{k_1}}-1} \cdot \prod_{j=1}^{J} \left(\frac{x(i_{C_{k_j}})}{x(i_{S_{k_{j+1}}})+I_{R_{k_{j+1}}}-1} \right) \\
&\qquad\qquad \times \prod_{k \notin \mathcal{K}_0} \left(\frac{x(i_{C_k})}{x(i_{S_k})+I_{R_k}-1} \right) \\
&\quad = \frac{x(i_{S_K})}{x(i_{S_K})+I_{R_K}-1} \cdot \frac{(x(i_{S_1})+c)\prod_{k\in\mathcal{K}_1} x(i_{C_k})}{\prod_{k=1}^{\nu(S_1)+1}(x(i_{S_1})+I_{R_k}-1)} \\
&\qquad\qquad \times \prod_{j=1}^{J} \left(\frac{x(i_{C_{k_j}})}{x(i_{S_{k_{j+1}}})+I_{R_{k_{j+1}}}-1} \right) \cdot \prod_{k\in\mathcal{K}_2} \left(\frac{x(i_{C_k})}{x(i_{S_k})+I_{R_k}-1} \right)
\end{aligned}$$

Hence for $0 \le c \le I_R^* - 1$,

$$\begin{aligned}
&\frac{x(i_{S_K})}{x(i_{S_K})+I_{R_K}-1} \cdot \frac{(x(i_{S_1})+c)\prod_{k\in\mathcal{K}_1} x(i_{C_k})}{\prod_{k=1}^{\nu(S_1)+1}(x(i_{S_1})+I_{R_k}-1)} \\
&\qquad \times \prod_{j=1}^{J} \left(\frac{x(i_{C_{k_j}})}{x(i_{S_{k_{j+1}}})+I_{R_{k_{j+1}}}-1} \right) \cdot \prod_{k\in\mathcal{K}_2} \left(\frac{x(i_{C_k})}{x(i_{S_k})+I_{R_k}-1} \right) \\
&\quad \le \frac{x(i_{S_K})+1}{x(i_{S_K})+I_{R_K}} \cdot \frac{(x(i_{S_1})+c+1)\prod_{k\in\mathcal{K}_1}(x(i_{C_k})+1)}{\prod_{k=1}^{\nu(S_1)+1}(x(i_{S_1})+I_{R_k})} \\
&\qquad \times \prod_{j=1}^{J} \left(\frac{(x(i_{C_{k_j}})+1)}{x(i_{S_{k_{j+1}}})+I_{R_{k_{j+1}}}} \right) \cdot \prod_{k\in\mathcal{K}_2} \left(\frac{x(i_{C_k})+1}{x(i_{S_k})+I_{R_k}} \right) \\
&\quad = \frac{(x(i_{S_1})+c+1)(x(i_{S_K})+1)}{(x(i_{S_K})+I_{R_K})^2} \cdot \prod_{k=1}^{K-1} \left(\frac{x(i_{C_k})+1}{x(i_{S_k})+I_{R_k}} \right),
\end{aligned}$$

which completes the proof. □

Proof of Theorem 5.1. In the same way as (18) and (19) the UMVUE of the difference of the risks between $\hat{\boldsymbol{\lambda}}^{ML}$ and $\hat{\boldsymbol{\lambda}}(\pi)$ for the model (3) can be obtained by

$$\hat{\mathrm{Rd}}(\hat{\boldsymbol{\lambda}}(\pi), \mathcal{G})$$

$$= \sum_{i \in \mathcal{I}} \Bigg\{ (x(i_{S_1})+1) \cdot \prod_{k=1}^{K} \frac{x(i_{C_k})+1}{x(i_{S_k})+1} - 2x(i_{S_1}) \prod_{k=1}^{K} \frac{x(i_{C_k})}{x(i_{S_k})}$$

$$+2x(i_{S_1}) \cdot \frac{\Psi(x^+)}{x^+ + I_{S_1} - 1} \cdot \prod_{k=1}^{K} \frac{x(i_{C_k})}{x(i_{S_k}) + I_{R_k} - 1}$$

$$- (x(i_{S_1})+1) \cdot \left(\frac{\Psi(x^+ + 1)}{x^+ + I_{S_1}} \right)^2 \cdot \prod_{k=1}^{K} \frac{(x(i_{C_k})+1)(x(i_{S_k})+1)}{(x(i_{S_k}) + I_{R_k})^2} \Bigg\}$$

$$= \sum_{i_{H_{K-1}} \in \mathcal{I}_{H_{K-1}}} \Bigg\{ (x(i_{S_1})+1) \cdot \prod_{k=1}^{K-1} \frac{(x(i_{C_k})+1)}{(x(i_{S_k})+1)} \cdot \frac{(x(i_{S_K}) + I_{R_K})}{(x(i_{S_K})+1)}$$

$$-2x(i_{S_1}) \prod_{k=1}^{K-1} \frac{x(i_{C_k})}{x(i_{S_k})}$$

$$+\frac{2x(i_{S_1})\Psi(x^+)}{x^+ + I_{S_1} - 1} \cdot \prod_{k=1}^{K-1} \frac{x(i_{C_k})}{x(i_{S_k}) + I_{R_k} - 1} \cdot \frac{x(i_{S_K})}{x(i_{S_K}) + I_{R_K} - 1}$$

$$-(x(i_{S_1})+1) \cdot \left(\frac{\Psi(x^+ + 1)}{x^+ + I_{S_1}} \right)^2 \cdot \frac{x(i_{S_K})+1}{x(i_{S_K}) + I_{R_K}}$$

$$\times \prod_{k=1}^{K-1} \frac{(x(i_{C_k})+1)(x(i_{S_k})+1)}{(x(i_{S_k}) + I_{R_k})^2} \Bigg\}.$$

Differentiating partially $\hat{\mathrm{Rd}}(\hat{\boldsymbol{\lambda}}(\pi), \mathcal{G})$ with respect to I_{R_K},

$$\frac{\partial \hat{\mathrm{Rd}}(\hat{\boldsymbol{\lambda}}(\pi), \mathcal{G})}{\partial I_{R_K}}$$

$$= \sum_{i_{H_{K-1}} \in \mathcal{I}_{H_{K-1}}} \Bigg\{ (x(i_{S_1})+1) \prod_{k=1}^{K-1} \frac{(x(i_{C_k})+1)}{(x(i_{S_k})+1)} \cdot \frac{1}{(x(i_{S_K})+1)}$$

$$-\frac{2x(i_{S_1})\Psi(x^+)}{x^+ + I_{S_1} - 1} \cdot \prod_{k=1}^{K-1} \frac{x(i_{C_k})}{x(i_{S_k}) + I_{R_k} - 1} \cdot \frac{x(i_{S_K})}{(x(i_{S_K}) + I_{R_K} - 1)^2}$$

$$-(x(i_{S_1})+1) \cdot \left(\frac{\Psi(x^+ + 1)}{x^+ + I_{S_1}} \right)^2 \cdot \frac{x(i_{S_K})+1}{(x(i_{S_K}) + I_{R_K})^2}$$

$$\times \prod_{k=1}^{K-1} \frac{(x(i_{C_k})+1)(x(i_{S_k})+1)}{(x(i_{S_k}) + I_{R_k})^2} \Bigg\}$$

$$\geq \sum_{i_{H_{K-1}} \in \mathcal{I}_{H_{K-1}}} \Bigg\{ (x(i_{S_1})+1) \prod_{k=1}^{K-1} \frac{(x(i_{C_k})+1)}{(x(i_{S_k})+1)} \cdot \frac{1}{(x(i_{S_K})+1)}$$

$$-2(x(i_{S_1})+1) \cdot \frac{\Psi(x^+ + 1)}{x^+ + I_{S_1}} \cdot \prod_{k=1}^{K-1} \frac{x(i_{C_k})+1}{x(i_{S_k}) + I_{R_k}} \cdot \frac{x(i_{S_K})+1}{(x(i_{S_K}) + I_{R_K})^2}$$

$$-(x(i_{S_1})+1)\cdot\left(\frac{\Psi(x^++1)}{x^++I_{S_1}}\right)^2\cdot\frac{x(i_{S_K})+1}{(x(i_{S_K})+I_{R_K})^2}$$
$$\left.\times\prod_{k=1}^{K-1}\frac{(x(i_{C_k})+1)(x(i_{S_k})+1)}{(x(i_{S_k})+I_{R_k})^2}\right\}$$
$$=\sum_{i_{H_{K-1}}\in\mathcal{I}_{H_{K-1}}}(x(i_{S_1})+1)\prod_{k=1}^{K-1}\frac{(x(i_{C_k})+1)}{(x(i_{S_k})+1)}\cdot\frac{1}{(x(i_{S_K})+1)}$$
$$\times\left\{1-\frac{\Psi(x^++1)}{x^++I_{S_1}}\cdot\prod_{k=1}^{K-1}\frac{x(i_{S_k})+1}{x(i_{S_k})+I_{R_k}}\right.$$
$$\left.\times\left(2-\frac{\Psi(x^++1)}{x^++I_{S_1}}\cdot\prod_{k=1}^{K-1}\frac{x(i_{S_k})+1}{x(i_{S_k})+I_{R_k}}\right)\left(\frac{x(i_{S_K})+1}{x(i_{S_K})+I_{R_K}}\right)^2\right\}. \quad (32)$$

The inequality follows from Lemma 5.1 with $c=0$ and (30). From (31) with $k=K-1$ the right hand side of (32) is always nonnegative for $I_{R_K}\geq 1$. Thus $\hat{\mathrm{Rd}}(\hat{\boldsymbol{\lambda}}(\pi),\mathcal{G})$ is nondecreasing on $I_{R_K}\geq 1$. Hence we have

$$\hat{\mathrm{Rd}}(\hat{\boldsymbol{\lambda}}(\pi),\mathcal{G})$$
$$\geq\sum_{i_{H_{K-1}}\in\mathcal{I}_{H_{K-1}}}\left\{(x(i_{S_1})+1)\prod_{k=1}^{K-1}\frac{(x(i_{C_k})+1)}{(x(i_{S_k})+1)}-2x(i_{S_1})\cdot\prod_{k=1}^{K-1}\frac{x(i_{C_k})}{x(i_{S_k})}\right.$$
$$+2x(i_{S_1})\cdot\frac{\Psi(x^+)}{x^++I_{S_1}-1}\cdot\prod_{k=1}^{K-1}\frac{x(i_{C_k})}{x(i_{S_k})+I_{R_k}-1}$$
$$\left.-(x(i_{S_1})+1)\cdot\left(\frac{\Psi(x^++1)}{x^++I_{S_1}}\right)^2\cdot\prod_{k=1}^{K-1}\frac{(x(i_{C_k})+1)(x(i_{S_k})+1)}{(x(i_{S_k})+I_{R_k})^2}\right\}$$
$$=\hat{\mathrm{Rd}}(\hat{\boldsymbol{\lambda}}(\pi),\mathcal{G}(H_{K-1})). \quad (33)$$

We note that the right hand side of (33) is equal to $\hat{\mathrm{Rd}}(\hat{\boldsymbol{\lambda}}(\pi),\mathcal{G})$ with $I_{R_K}=1$, which corresponds to the induced subgraph $\mathcal{G}(H_{K-1})$. By iterating this procedure we obtain

$$\hat{\mathrm{Rd}}(\hat{\boldsymbol{\lambda}}(\pi),\mathcal{G})\geq\hat{\mathrm{Rd}}(\hat{\boldsymbol{\lambda}}(\pi),\mathcal{G}(H_{K-1}))\geq\cdots\geq\hat{\mathrm{Rd}}(\hat{\boldsymbol{\lambda}}(\pi),\mathcal{G}(H_{\nu(S_1)+1})).$$

From Lemma 4.2, (30) and (31) with $k=\nu(S_1)+1$ imply (16) and (17). Hence

$$\hat{\mathrm{Rd}}(\hat{\boldsymbol{\lambda}}(\pi),\mathcal{G}(H_{\nu(S_1)+1}))\geq 0$$

under the condition of the theorem. Then we can complete the proof. □

In the same way we can obtain the following result corresponding to Theorem 4.2.

Theorem 5.2. *If $\Psi(\cdot)$ satisfies the condition of Theorem 4.2 and*

$$\frac{\Psi(x)}{x+I_R^*-1}\leq\frac{\Psi(x+1)}{x+I_R^*}, \quad (34)$$

$$0 \leq \frac{\Psi(x(i_{S_1})+1)}{x(i_{S_1})+1} \cdot \prod_{j=1}^{k} \frac{x(i_{S_j})+1}{x(i_{S_j})+I_{R_j}} \leq 1,$$

for $\nu(S_1)+1 \leq k \leq K$, $\hat{\boldsymbol{\lambda}}(\pi_{S_1})$ *improves on* $\hat{\boldsymbol{\lambda}}^{ML}$ *under the loss (2).*

Theorem 5.1 and 5.2 suggest that the dominance relationship between the proposed Bayes estimators and the MLE for general decomposable models can be reduced to the one in the model with one minimal vertex separator discussed in the previous section. Then we can obtain the following results corresponding to Corollary 4.1 and 4.2.

Corollary 5.1. *If* α, β *and* μ *satisfy the conditions in Corollary 4.1,* $\hat{\boldsymbol{\lambda}}^{\alpha,\beta}(\pi)$ *improves on* $\hat{\boldsymbol{\lambda}}^{ML}$ *under the loss (2). If* $\alpha > 0$, $\hat{\boldsymbol{\lambda}}^{\alpha,\beta}(\pi)$ *is admissible.*

Corollary 5.2. *If* α, β *and* μ *satisfy the conditions in Corollary 4.2 and (34),* $\hat{\boldsymbol{\lambda}}^{\alpha,\beta}(\pi_{S_1})$ *improves on* $\hat{\boldsymbol{\lambda}}^{ML}$ *under the loss (2). If* $\alpha > 0$, $\hat{\boldsymbol{\lambda}}^{\alpha,\beta}(\pi_{S_1})$ *is admissible.*

Remark 5.1. The sufficient conditions on $\hat{\boldsymbol{\lambda}}^{GP}(\pi)$, $\hat{\boldsymbol{\lambda}}^{GP}(\pi_{S_1})$, $\hat{\boldsymbol{\lambda}}^{CZ}(\pi)$, $\hat{\boldsymbol{\lambda}}^{CZ}(\pi_{S_1})$, $\hat{\boldsymbol{\lambda}}^{\gamma}(\pi)$ and $\hat{\boldsymbol{\lambda}}^{\gamma}(\pi_{S_1})$ obtained in Remark 4.1 and 4.2 are applicable for general decomposable models without any change.

6 Admissibility of the generalized Bayes estimators

In this section we discuss the admissibility of $\hat{\boldsymbol{\lambda}}(\pi)$ with respect to improper priors. Johnstone[14] showed that the admissibility of the estimators of $\boldsymbol{\lambda}$ in the saturated model can be reduced to the admissibility of one dimensional parameter λ. Here we generalize the argument to the decomposable graphical models and obtain a sufficient condition for the admissibility of $\hat{\boldsymbol{\lambda}}(\pi)$.

Let $\pi_n(\lambda, \boldsymbol{\alpha}(S_1), \boldsymbol{\beta})$ be a sequence of proper priors of form (10) with $m(\cdot) = m_n(\cdot)$ and denote the resulting Bayes estimator and $\Psi(\cdot)$ with $m(\cdot) = m_n(\cdot)$ by $\hat{\boldsymbol{\lambda}}(\pi_n)$ and $\Psi_n(\cdot)$, respectively. We write the Bayes risk of $\hat{\boldsymbol{\lambda}}(\pi)$ with respect to π_n as $\mathrm{B}(\hat{\boldsymbol{\lambda}}(\pi), \pi_n)$. Following Blyth[1], $\hat{\boldsymbol{\lambda}}(\pi)$ is admissible if

$$\lim_{n\to\infty} \Big(\mathrm{B}(\hat{\boldsymbol{\lambda}}(\pi), \pi_n) - \mathrm{B}(\hat{\boldsymbol{\lambda}}(\pi_n), \pi_n) \Big) = 0$$

for some sequence of proper priors π_n.

We note that $\Psi_n(x^+)$ is the Bayes estimator of λ with respect to the prior $m_n(\lambda)$ for the loss function $\mathrm{L}(\lambda, \hat{\lambda}) = (\hat{\lambda} - \lambda)^2/\lambda$. Define $M_n(x)$ by

$$M_n(x) = \int_0^\infty \exp(-\lambda)\lambda^x m_n(\lambda) \mathrm{d}\lambda.$$

The following result is a generalization of Theorem 4.1 in Johnstone[14]

Theorem 6.1. *If* $\Psi(x^+)$ *is admissible as an estimator of* λ *in the one-dimensional problem under the loss function* $\mathrm{L}(\lambda, \hat{\lambda}) = (\hat{\lambda} - \lambda)^2/\lambda$, $\hat{\boldsymbol{\lambda}}(\pi)$ *is also admissible.*

Proof. Following Johnstone[13], the difference of the Bayes risks between $\Psi(x^+)$ and $\Psi_n(x^+)$ with respect to the prior $m_n(\lambda)$ is written by

$$\begin{aligned}
&\mathrm{B}(\Psi(x^+),\pi_n(\lambda)) - \mathrm{B}(\Psi_n(x^+),\pi_n(\lambda))\\
&= \int \Big(\mathrm{R}(\lambda,\Psi(x^+)) - \mathrm{R}(\lambda,\Psi_n(x^+))\Big)\pi_n(\lambda)\mathrm{d}\lambda\\
&= \sum_{x^+\geq 1} (\Psi(x^+) - \Psi_n(x^+))^2 \cdot \frac{M_n(x^+-1)}{x^+!}.
\end{aligned}$$

Define $\boldsymbol{x}_{\mathcal{C}}(i)$ be $\boldsymbol{x}_{\mathcal{C}}(i) = \{x(i_{C_k}), k=1,\dots,K\}$. Let $\boldsymbol{x}_{\mathcal{C}}(i) \geq \mathbf{1}$ denote that $x(i_{C_k}) \geq 1$, $\forall k = 1,\dots,K$. Noting that $\hat{\lambda}(i,\pi) = \hat{\lambda}(i,\pi_n) = 0$ if $x(i_{C_k}) = 0$ for some k, the difference of the Bayes risks between $\hat{\boldsymbol{\lambda}}(\pi)$ and $\hat{\boldsymbol{\lambda}}(\pi_n)$ with respect to the prior π_n is

$$\begin{aligned}
&\mathrm{B}(\hat{\boldsymbol{\lambda}}(\pi),\pi_n) - \mathrm{B}(\hat{\boldsymbol{\lambda}}(\pi_n),\pi_n)\\
&= \int \Big(\mathrm{R}(\boldsymbol{\lambda},\hat{\boldsymbol{\lambda}}(\pi)) - \mathrm{R}(\boldsymbol{\lambda},\hat{\boldsymbol{\lambda}}(\pi_n))\Big)\,\pi_n(\lambda,\boldsymbol{\alpha}(S_1),\boldsymbol{\beta})\mathrm{d}\lambda\mathrm{d}\boldsymbol{\alpha}(S_1)\mathrm{d}\boldsymbol{\beta}\\
&= \int \sum_{i\in\mathcal{I}} \sum_{\boldsymbol{x}_{\mathcal{C}}:\boldsymbol{x}_{\mathcal{C}}(i)\geq\mathbf{1}} \Bigg\{ \frac{\Psi^2(x^+) - \Psi_n^2(x^+)}{\lambda(i)} \left(\frac{x(i_{S_1})}{x^+ + I_{S_1} - 1} \prod_{k=1}^{K} \frac{x(i_{C_k})}{x(i_{S_k}) + I_{R_k} - 1} \right)^2\\
&\qquad -2\left(\Psi(x^+) - \Psi_n(x^+)\right) \frac{x(i_{S_1})}{x^+ + I_{S_1} - 1} \prod_{k=1}^{K} \frac{x(i_{C_k})}{x(i_{S_k}) + I_{R_k} - 1} \Bigg\}\\
&\qquad \times \Pr(\boldsymbol{x}_{\mathcal{C}}) m_n(\lambda)\mathrm{d}\lambda\mathrm{d}\boldsymbol{\alpha}(S_1)\mathrm{d}\boldsymbol{\beta}\\
&= \sum_{i\in\mathcal{I}} \sum_{\boldsymbol{x}_{\mathcal{C}}:\boldsymbol{x}_{\mathcal{C}}(i)\geq\mathbf{1}} \Big\{ \left(\Psi^2(x^+) - \Psi_n^2(x^+)\right) \cdot M_n(x^+-1) - 2(\Psi(x^+) - \Psi_n(x^+)) \cdot M_n(x^+) \Big\}\\
&\qquad \times \frac{x(i_{S_1})}{x^+ + I_{S_1} - 1} \cdot \prod_{k=1}^{K} \frac{x(i_{C_k})}{x(i_{S_k}) + I_{R_k} - 1} \cdot \frac{\prod_{i_{S_1}\in\mathcal{I}_{S_1}} \Gamma(x(i_{S_1})+1)}{\Gamma(x^+ + I_{S_1})}\\
&\qquad \times \frac{\prod_{k=1}^{K} \prod_{i_{C_k}\in\mathcal{I}_{C_k}} \Gamma(x(i_{C_k})+1)}{\prod_{k=1}^{K} \prod_{i_{S_k}\in\mathcal{I}_{S_k}} \Gamma(x(i_{S_k}) + I_{R_k})} \cdot t(\boldsymbol{x}_{\mathcal{C}})\\
&= \sum_{i\in\mathcal{I}} \sum_{\boldsymbol{x}_{\mathcal{C}}:\boldsymbol{x}_{\mathcal{C}}(i)\geq\mathbf{1}} (\Psi(x^+) - \Psi_n(x^+))^2 \cdot \frac{M_n(x^+-1)}{x^+!} \cdot C(i,\boldsymbol{x}_{\mathcal{C}}),
\end{aligned}$$

where

$$\begin{aligned}
C(i,\boldsymbol{x}_{\mathcal{C}}) = &\frac{x(i_{S_1})}{x^+ + I_{S_1} - 1} \cdot \frac{x^+!}{(x^+ + I_{S_1} - 1)!} \cdot \prod_{k=1}^{K} \frac{x(i_{C_k})}{x(i_{S_k}) + I_{R_k} - 1}\\
&\times \prod_{k=1}^{K} \prod_{i_{S_k}\in\mathcal{I}_{S_k}} \frac{x(i_{S_k})!}{(x(i_{S_k}) + I_{R_k} - 1)!}.
\end{aligned}$$

$C(i,\boldsymbol{x}_{\mathcal{C}})$ satisfies $0 < C(i,\boldsymbol{x}_{\mathcal{C}}) < 1$ for $\boldsymbol{x}_{\mathcal{C}}$ such that $\boldsymbol{x}_{\mathcal{C}}(i) \geq \mathbf{1}$. Hence

$$\lim_{n\to\infty} \sum_{i\in\mathcal{I}} \sum_{\boldsymbol{x}_{\mathcal{C}}:\boldsymbol{x}_{\mathcal{C}}(i)\geq\mathbf{1}} (\Psi(x^+) - \Psi_n(x^+))^2 \cdot \frac{M_n(x^+-1)}{x^+!} \cdot C(i,\boldsymbol{x}_{\mathcal{C}}) = 0$$

is equivalent to

$$\lim_{n\to\infty}\sum_{x^+\geq 1}(\Psi(x^+)-\Psi_n(x^+))^2\cdot\frac{M_n(x^+-1)}{x^+!}=0.$$

Thus the admissibility of $\Psi(x^+)$ implies the admissibility of $\hat{\boldsymbol{\lambda}}(\pi)$. □

By using Corollary 5.2 in Johnstone[13] and Corollary 4.3 in Johnstone[14], we can obtain a sufficient condition for admissibility of $\hat{\boldsymbol{\lambda}}(\pi)$.

Theorem 6.2. *Suppose that* $\Psi(\cdot)$ *satisfies*

$$|\Psi(x)-x|\leq M(1+\sqrt{x}),\quad \textit{for}\quad \exists M>0$$

$$\Psi(x+1)-\Psi(x)=O(1).$$

and there exist $c\leq 1$ *and* $x_0>0$ *such that* $\Psi(x)\leq x+c/\log x$ *for* $x>x_0$*, then* $\hat{\boldsymbol{\lambda}}(\pi)$ *is admissible.*

Corollary 6.1. *If* $\alpha\geq -1$*,* $\hat{\boldsymbol{\lambda}}^{\alpha,\beta}(\pi)$ *is admissible.*

By using the similar argument to the proof of Theorem 6.1, we can show that the admissibility of $\hat{\boldsymbol{\lambda}}(\pi_{S_1})$ can be reduced to the admissibility of $\{\Psi(x(i_{S_1})), i_{S_1}\in\mathcal{I}_{S_1}\}$ as the estimator of I_{S_1} dimensional parameter $\boldsymbol{\lambda}(i_{S_1})$ under the loss $\sum_{i_{S_1}\in\mathcal{I}_{S_1}}(\Psi(x(i_{S_1}))-\lambda(i_{S_1}))^2/\lambda(i_{S_1})$. Since $\boldsymbol{\lambda}(i_{S_1})$ is multidimensional, it seems difficult to derive sufficient conditions on $\hat{\boldsymbol{\lambda}}(\pi_{S_1})$ with respect to an improper prior to be admissible at this point.

7 Monte Carlo Studies

We study the risk performance of $\hat{\boldsymbol{\lambda}}^{\alpha,\beta}(\pi)$ and $\hat{\boldsymbol{\lambda}}^{\alpha,\beta}(\pi_{S_1})$ for the 5-way decomposable model corresponding to the graph in Figure 1 through Monte Carlo studies with 100,000 replications.

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

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

We set $I_\delta=3$ for all $\delta\in\Delta=\{1,2,3,4,5\}$ and $\alpha(i_C)=\prod_{\delta\in C}\alpha(i_\delta)$ for all $C\in\mathcal{C}$.

In Tables 1 to 6 we present the risks of $\hat{\boldsymbol{\lambda}}^{\alpha,\beta}(\pi)$ and $\hat{\boldsymbol{\lambda}}^{\alpha,\beta}(\pi_{S_1})$ with

- $S_1=\{2\}$ and $\mathbb{C}(S_1)=\{\{1,2\},\{2,3,5\}\}$,
- $S_1=\{2,5\}$ and $\mathbb{C}(S_1)=\{\{2,3,5\},\{2,4,5\}\}$.

for some combination of λ and (α,β,μ) which satisfies the conditions of Corollary 4.1 and 4.2. With respect to $\alpha(i_\delta)$ we consider the following two cases,

- $\alpha(i_\delta)=1/3$ for all $i_\delta\in\mathcal{I}_\delta$, $\delta\in\Delta$,
- $\alpha(i_{S_1})$ are unbalanced and $\alpha(i_\delta)=1/3$ for all $i_\delta\in\mathcal{I}_\delta$, $\delta\in\Delta\setminus\{S_1\}$.

The summary of 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 toward zero, we can see considerable amount of risk reduction when λ is small.
- The improvement is in the inverse proportion to λ.
- The difference among estimators is small.
- When $\alpha(i_\delta)$ are unbalanced, the dominance between $\hat{\boldsymbol{\lambda}}^{\alpha,\beta}(\pi)$ and $\hat{\boldsymbol{\lambda}}^{\alpha,\beta}(\pi_{S_1})$ seems to depend on λ. When $\lambda = 1$ and 10000, $\hat{\boldsymbol{\lambda}}^{\alpha,\beta}(\pi)$ shows larger risk reduction than $\hat{\boldsymbol{\lambda}}^{\alpha,\beta}(\pi_{S_1})$. However $\hat{\boldsymbol{\lambda}}^{\alpha,\beta}(\pi_{S_1})$ seems to be better for $\lambda = 100$.

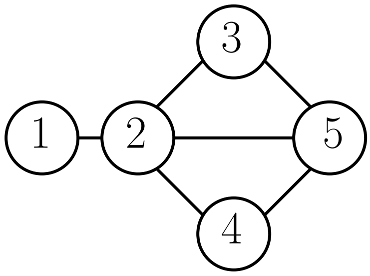

Figure 1: 5-way decomposable model

References

[1] Blyth, C. R. (1951). On minimax statistical decision procedures and their admissibility, *Ann. Math. Statist.*, **22**, 22-42.

[2] Brown, L. D. and Farrell, R. H.(1985). Complete class theorems for estimation of multivariate Poisson means and related problems, *Ann. Statist.*, **13**, 706-726.

[3] 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.

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

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

[6] 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.

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

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

[9] 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`), to appear in *J. Japan. Statist. Soc.*

[10] Hara, H. and Takemura, A.(2005). Improving on the maximum likelihood estimators of the means in Poisson decomposable graphical models, Technical Reports, *METR*, 2005-08, 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] Johnstone, I. (1986). Admissible estimation, Dirichlet principles and recurrence of birth-death chains on Z_+^p, *Probability Theory and Related Fields*, **71**, 231-269.

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

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

[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.

Table 1: Risks of $\hat{\boldsymbol{\lambda}}^{ML}$ and $\hat{\boldsymbol{\lambda}}^{\alpha,\beta}(\pi)$ for the model in Figure 1 with balanced $\alpha(i_\delta)$.

(1) $S_1 = \{2\}$

	λ					
(α, β, μ)	0.1	1.0	10	10^2	10^3	10^4
$(27, 1, 3)$	0.100	0.998	9.513	40.643	50.132	50.928
$(27/2, 1, 3)$	0.100	0.997	9.391	38.917	49.628	50.871
$(0, 1, 3)$	0.103	1.008	9.383	38.222	49.441	50.851
$(-1, 1, 3)$	0.106	1.015	9.425	38.240	49.440	50.850
$(27, 1, 3/2)$	0.100	0.998	9.541	40.709	50.134	50.928
$(27/2, 1, 3/2)$	0.100	0.998	9.412	38.940	49.629	50.871
$(0, 1, 3/2)$	0.102	1.006	9.380	38.222	49.441	50.851
$(-3/4, 3/4, 3)$	0.104	1.011	9.411	38.234	49.440	50.850
$(-3/2, 3/2, 3)$	0.112	1.025	9.457	38.253	49.441	50.851
$\hat{\boldsymbol{\lambda}}^{ML}$	239.31	214.44	111.08	56.919	51.505	51.060

(2) $S_1 = \{2, 5\}$

	λ					
(α, β, μ)	0.1	1.0	10	10^2	10^3	10^4
$(27, 1, 9)$	0.100	0.997	9.448	41.001	50.296	50.946
$(27/2, 1, 9)$	0.101	0.997	9.366	39.508	49.800	50.890
$(0, 1, 9)$	0.105	1.009	9.445	39.003	49.617	50.870
$(-1, 1, 9)$	0.106	1.013	9.482	39.029	49.617	50.870
$(27, 1, 9/2)$	0.100	0.997	9.501	41.172	50.302	50.947
$(27/2, 1, 9/2)$	0.100	0.997	9.396	39.567	49.802	50.890
$(0, 1, 9/2)$	0.104	1.007	9.438	39.002	49.617	50.870
$(-9/4, 9/4, 9)$	0.124	1.040	9.581	39.081	49.619	50.870
$(-9/2, 9/2, 9)$	0.191	1.134	9.844	39.224	49.631	50.871
$\hat{\boldsymbol{\lambda}}^{ML}$	239.31	214.44	111.08	56.919	51.505	51.060

Table 2: Risks of $\hat{\boldsymbol{\lambda}}^{ML}$ and $\hat{\boldsymbol{\lambda}}^{\alpha,\beta}(\pi_{S_1})$ for the model in Figure 1 with balanced $\alpha(i_\delta)$.

(1) $S_1 = \{2\}$

	λ					
(α, β, μ)	0.1	1.0	10	10^2	10^3	10^4
$(9, 1, 9)$	0.109	1.008	9.319	40.015	50.197	50.939
$(9/2, 1, 9)$	0.116	1.021	9.334	38.779	49.675	50.877
$(0, 1, 9)$	0.134	1.057	9.497	38.271	49.451	50.852
$(-1, 1, 9)$	0.142	1.073	9.584	38.313	49.445	50.851
$(9, 1, 9/2)$	0.104	1.000	9.366	40.481	50.219	50.939
$(9/2, 1, 9/2)$	0.108	1.009	9.332	38.969	49.682	50.878
$(0, 1, 9/2)$	0.128	1.047	9.469	38.280	49.452	50.852
$(-9/4, 9/4, 9)$	0.299	1.267	9.983	38.497	49.460	50.852
$(-9/2, 9/2, 9)$	0.879	1.963	11.331	39.285	49.554	50.862
$\hat{\boldsymbol{\lambda}}^{ML}$	239.31	214.44	111.08	56.919	51.505	51.060

(2) $S_1 = \{2, 5\}$

	λ					
(α, β, μ)	0.1	1.0	10	10^2	10^3	10^4
$(3, 1, 3)$	0.166	1.083	9.479	41.353	50.749	51.003
$(3/2, 1, 3)$	0.207	1.143	9.646	40.162	50.103	50.925
$(0, 1, 3)$	0.302	1.279	10.065	39.553	49.747	50.884
$(-1, 1, 3)$	0.457	1.492	10.699	39.708	49.682	50.876
$(3, 1, 3/2)$	0.128	1.035	9.459	42.048	50.785	51.003
$(3/2, 1, 3/2)$	0.152	1.074	9.560	40.552	50.119	50.926
$(0, 1, 3/2)$	0.232	1.193	9.928	39.664	49.751	50.884
$(-3/4, 3/4, 3)$	0.306	1.311	10.369	39.644	49.686	50.877
$(-3/2, 3/2, 3)$	0.886	1.999	11.620	40.035	49.704	50.878
$\hat{\boldsymbol{\lambda}}^{ML}$	239.31	214.44	111.08	56.919	51.505	51.060

Table 3: Risks of $\hat{\boldsymbol{\lambda}}^{ML}$ and $\hat{\boldsymbol{\lambda}}^{\alpha,\beta}(\pi)$ for the model in Figure 1 with unbalanced $\alpha(i_2)$.

(1) $S_1 = \{2\}$, $\lambda = 1.0$

	$\boldsymbol{\alpha}(2) = \{\alpha(i_2), i_2 \in \mathcal{I}_2\}$			
(α, β, μ)	(0.1,0.2,0.7)	(0.1,0.3,0.6)	(0.2,0.3,0.5)	(0.3,0.3,0.4)
$(27, 1, 3)$	0.997	0.997	0.998	0.998
$(27/2, 1, 3)$	0.996	0.996	0.997	0.997
$(0, 1, 3)$	1.004	1.006	1.007	1.008
$(-1, 1, 3)$	1.011	1.013	1.014	1.015
$(27, 1, 3/2)$	0.997	0.998	0.998	0.998
$(27/2, 1, 3/2)$	0.996	0.997	0.997	0.998
$(0, 1, 3/2)$	1.003	1.004	1.005	1.006
$(-3/4, 3/4, 3)$	1.008	1.009	1.010	1.011
$(-3/2, 3/2, 3)$	1.021	1.023	1.024	1.025
$\hat{\boldsymbol{\lambda}}^{ML}$	215.75	215.13	214.55	214.54

(2) $S_1 = \{2\}$, $\lambda = 100.0$

	$\boldsymbol{\alpha}(2)$			
(α, β, μ)	(0.1,0.2,0.7)	(0.1,0.3,0.6)	(0.2,0.3,0.5)	(0.3,0.3,0.4)
$(27, 1, 3)$	35.800	37.016	39.600	40.499
$(27/2, 1, 3)$	33.844	35.118	37.822	38.768
$(0, 1, 3)$	32.956	34.280	37.086	38.073
$(-1, 1, 3)$	32.962	34.288	37.100	38.090
$(27, 1, 3/2)$	35.873	37.088	39.667	40.564
$(27/2, 1, 3/2)$	33.871	35.145	37.847	38.792
$(0, 1, 3/2)$	32.957	34.280	37.086	38.073
$(-3/4, 3/4, 3)$	32.960	34.285	37.096	38.084
$(-3/2, 3/2, 3)$	32.970	34.298	37.113	38.103
$\hat{\boldsymbol{\lambda}}^{ML}$	62.380	61.377	57.905	57.068

(3) $S_1 = \{2\}$, $\lambda = 10000.0$

	$\boldsymbol{\alpha}(2)$			
(α, β, μ)	(0.1,0.2,0.7)	(0.1,0.3,0.6)	(0.2,0.3,0.5)	(0.3,0.3,0.4)
$(27, 1, 3)$	50.749	50.731	50.887	50.943
$(27/2, 1, 3)$	50.692	50.675	50.831	50.887
$(0, 1, 3)$	50.671	50.654	50.810	50.866
$(-1, 1, 3)$	50.671	50.654	50.810	50.866
$(27, 1, 3/2)$	50.749	50.731	50.887	50.943
$(27/2, 1, 3/2)$	50.692	50.675	50.831	50.887
$(0, 1, 3/2)$	50.671	50.654	50.810	50.866
$(-3/4, 3/4, 3)$	50.671	50.654	50.810	50.866
$(-3/2, 3/2, 3)$	50.671	50.654	50.810	50.866
$\hat{\boldsymbol{\lambda}}^{ML}$	51.055	51.004	51.052	51.080

Table 4: Risks of $\hat{\boldsymbol{\lambda}}^{ML}$ and $\hat{\boldsymbol{\lambda}}^{\alpha,\beta}(\pi_{S_1})$ for the model in Figure 1 with unbalanced $\alpha(i_2)$.

(1) $S_1 = \{2\}$, $\lambda = 1.0$

	$\boldsymbol{\alpha}(2)$			
(α, β, μ)	(0.1,0.2,0.7)	(0.1,0.3,0.6)	(0.2,0.3,0.5)	(0.3,0.3,0.4)
$(9, 1, 9)$	1.004	1.006	1.007	1.008
$(9/2, 1, 9)$	1.017	1.019	1.020	1.021
$(0, 1, 9)$	1.054	1.055	1.056	1.057
$(-1, 1, 9)$	1.070	1.071	1.072	1.073
$(9, 1, 9/2)$	0.997	0.998	1.000	1.000
$(9/2, 1, 9/2)$	1.005	1.007	1.008	1.009
$(0, 1, 9/2)$	1.044	1.045	1.046	1.047
$(-9/4, 9/4, 9)$	1.265	1.266	1.265	1.268
$(-9/2, 9/2, 9)$	1.961	1.962	1.959	1.965
$\hat{\boldsymbol{\lambda}}^{ML}$	215.75	215.13	214.55	214.54

(2) $S_1 = \{2\}$, $\lambda = 100.0$

	$\boldsymbol{\alpha}(2)$			
(α, β, μ)	(0.1,0.2,0.7)	(0.1,0.3,0.6)	(0.2,0.3,0.5)	(0.3,0.3,0.4)
$(9, 1, 9)$	34.020	35.544	38.691	39.833
$(9/2, 1, 9)$	33.197	34.607	37.562	38.615
$(0, 1, 9)$	32.941	34.282	37.118	38.119
$(-1, 1, 9)$	33.014	34.346	37.169	38.163
$(9, 1, 9/2)$	34.382	35.915	39.140	40.297
$(9/2, 1, 9/2)$	33.347	34.758	37.747	38.804
$(0, 1, 9/2)$	32.940	34.283	37.126	38.128
$(-9/4, 9/4, 9)$	33.277	34.587	37.372	38.351
$(-9/2, 9/2, 9)$	34.276	35.521	38.214	39.150
$\hat{\boldsymbol{\lambda}}^{ML}$	62.380	61.377	57.905	57.068

(3) $S_1 = \{2\}$, $\lambda = 10000.0$

	$\boldsymbol{\alpha}(2)$			
(α, β, μ)	(0.1,0.2,0.7)	(0.1,0.3,0.6)	(0.2,0.3,0.5)	(0.3,0.3,0.4)
$(9, 1, 9)$	50.827	50.796	50.910	50.956
$(9/2, 1, 9)$	50.718	50.697	50.840	50.894
$(0, 1, 9)$	50.673	50.655	50.811	50.867
$(-1, 1, 9)$	50.672	50.654	50.810	50.867
$(9, 1, 9/2)$	50.828	50.797	50.911	50.956
$(9/2, 1, 9/2)$	50.719	50.697	50.840	50.894
$(0, 1, 9/2)$	50.673	50.655	50.811	50.867
$(-9/4, 9/4, 9)$	50.675	50.657	50.812	50.868
$(-9/2, 9/2, 9)$	50.693	50.673	50.823	50.878
$\hat{\boldsymbol{\lambda}}^{ML}$	51.055	51.004	51.052	51.080

Table 5: Risks of $\hat{\boldsymbol{\lambda}}^{ML}$ and $\hat{\boldsymbol{\lambda}}^{\alpha,\beta}(\pi)$ for the model in Figure 1 with unbalanced $\alpha(i_{25})$.

(1) $S_1 = \{2, 5\}$, $\lambda = 1.0$

	$\boldsymbol{\alpha}(2) = \boldsymbol{\alpha}(5) = \{\alpha(i_5), i_5 \in \mathcal{I}_5\}$			
(α, β, μ)	(0.1,0.2,0.7)	(0.1,0.3,0.6)	(0.2,0.3,0.5)	(0.3,0.3,0.4)
$(27, 1, 9)$	0.994	0.995	0.997	0.997
$(27/2, 1, 9)$	1.003	1.006	1.008	1.009
$(0, 1, 9)$	1.006	1.009	1.012	1.013
$(-1, 1, 9)$	0.996	0.997	0.997	0.998
$(27, 1, 9/2)$	0.995	0.996	0.997	0.997
$(27/2, 1, 9/2)$	1.001	1.004	1.006	1.007
$(0, 1, 9/2)$	1.006	1.009	1.012	1.013
$(-9/4, 9/4, 9)$	1.032	1.036	1.039	1.040
$(-9/2, 9/2, 9)$	1.124	1.128	1.134	1.134
$\hat{\boldsymbol{\lambda}}^{ML}$	216.19	215.07	214.90	214.60

(2) $S_1 = \{2, 5\}$, $\lambda = 100.0$

	$\boldsymbol{\alpha}(2) = \boldsymbol{\alpha}(5)$			
(α, β, μ)	(0.1,0.2,0.7)	(0.1,0.3,0.6)	(0.2,0.3,0.5)	(0.3,0.3,0.4)
$(27, 1, 9)$	34.674	36.352	40.135	41.643
$(27/2, 1, 9)$	32.054	33.868	37.952	39.574
$(0, 1, 9)$	30.668	32.628	37.036	38.780
$(-1, 1, 9)$	30.657	32.625	37.054	38.805
$(27, 1, 9/2)$	34.946	36.612	40.369	41.868
$(27/2, 1, 9/2)$	32.164	33.971	38.040	39.657
$(0, 1, 9/2)$	30.669	32.629	37.036	38.779
$(-9/4, 9/4, 9)$	30.658	32.639	37.094	38.855
$(-9/2, 9/2, 9)$	30.716	32.718	37.219	38.996
$\hat{\boldsymbol{\lambda}}^{ML}$	65.520	63.636	58.511	57.134

(3) $S_1 = \{2, 5\}$, $\lambda = 10000.0$

	$\boldsymbol{\alpha}(2) = \boldsymbol{\alpha}(5)$			
(α, β, μ)	(0.1,0.2,0.7)	(0.1,0.3,0.6)	(0.2,0.3,0.5)	(0.3,0.3,0.4)
$(27, 1, 9)$	50.667	50.704	50.905	51.000
$(27/2, 1, 9)$	50.584	50.620	50.821	50.917
$(0, 1, 9)$	50.554	50.590	50.792	50.887
$(-1, 1, 9)$	50.554	50.590	50.791	50.887
$(27, 1, 9/2)$	50.667	50.704	50.905	51.000
$(27/2, 1, 9/2)$	50.584	50.620	50.822	50.917
$(0, 1, 9/2)$	50.554	50.590	50.792	50.887
$(-9/4, 9/4, 9)$	50.554	50.590	50.792	50.887
$(-9/2, 9/2, 9)$	50.555	50.591	50.793	50.888
$\hat{\boldsymbol{\lambda}}^{ML}$	51.086	51.042	51.032	51.085

Table 6: Risks of $\hat{\boldsymbol{\lambda}}^{ML}$ and $\hat{\boldsymbol{\lambda}}^{\alpha,\beta}(\pi_{S_1})$ for the model in Figure 1 with unbalanced $\alpha(i_{25})$.

(1) $S_1 = \{2,5\}$, $\lambda = 1.0$

	$\boldsymbol{\alpha}(2) = \boldsymbol{\alpha}(5)$			
(α, β, μ)	(0.1,0.2,0.7)	(0.1,0.3,0.6)	(0.2,0.3,0.5)	(0.3,0.3,0.4)
$(3, 1, 3)$	1.075	1.079	1.083	1.083
$(3/2, 1, 3)$	1.135	1.139	1.143	1.144
$(0, 1, 3)$	1.271	1.274	1.280	1.280
$(-1, 1, 3)$	1.485	1.488	1.495	1.493
$(3, 1, 3/2)$	1.028	1.031	1.035	1.035
$(3/2, 1, 3/2)$	1.066	1.070	1.074	1.074
$(0, 1, 3/2)$	1.186	1.189	1.194	1.194
$(-3/4, 3/4, 3)$	1.304	1.307	1.313	1.312
$(-3/2, 3/2, 3)$	1.992	1.994	2.004	2.001
$\hat{\boldsymbol{\lambda}}^{ML}$	216.19	215.07	214.90	214.60

(2) $S_1 = \{2,5\}$, $\lambda = 100.0$

	$\boldsymbol{\alpha}(2) = \boldsymbol{\alpha}(5)$			
(α, β, μ)	(0.1,0.2,0.7)	(0.1,0.3,0.6)	(0.2,0.3,0.5)	(0.3,0.3,0.4)
$(3, 1, 3)$	31.218	33.526	38.763	41.036
$(3/2, 1, 3)$	30.671	32.853	37.798	39.881
$(0, 1, 3)$	30.591	32.680	37.389	39.302
$(-1, 1, 3)$	31.041	33.081	37.667	39.475
$(3, 1, 3/2)$	31.602	33.954	39.371	41.721
$(3/2, 1, 3/2)$	30.863	33.075	38.139	40.267
$(0, 1, 3/2)$	30.602	32.708	37.477	39.412
$(-3/4, 3/4, 3)$	30.849	32.913	37.564	39.407
$(-3/2, 3/2, 3)$	31.672	33.651	38.091	39.815
$\hat{\boldsymbol{\lambda}}^{ML}$	65.520	63.636	58.511	57.134

(3) $S_1 = \{2,5\}$, $\lambda = 10000.0$

	$\boldsymbol{\alpha}(2) = \boldsymbol{\alpha}(5)$			
(α, β, μ)	(0.1,0.2,0.7)	(0.1,0.3,0.6)	(0.2,0.3,0.5)	(0.3,0.3,0.4)
$(3, 1, 3)$	50.945	50.918	50.962	51.023
$(3/2, 1, 3)$	50.712	50.723	50.862	50.944
$(0, 1, 3)$	50.584	50.616	50.808	50.901
$(-1, 1, 3)$	50.560	50.596	50.798	50.894
$(3, 1, 3/2)$	50.953	50.925	50.963	51.024
$(3/2, 1, 3/2)$	50.716	50.726	50.862	50.944
$(0, 1, 3/2)$	50.585	50.617	50.808	50.901
$(-3/4, 3/4, 3)$	50.562	50.597	50.799	50.894
$(-3/2, 3/2, 3)$	50.567	50.602	50.801	50.896
$\hat{\boldsymbol{\lambda}}^{ML}$	51.086	51.042	51.032	51.085