MATHEMATICAL ENGINEERING TECHNICAL REPORTS

Simultaneous Estimation of the Means in Some Poisson Log Linear Models

Hisayuki Hara and Akimichi Takemura

METR 2004–47 December 2004

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

Simultaneous Estimation of the Means in Some Poisson Log Linear Models

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

December, 2004

Abstract

In this article we study the simultaneous estimation of the means in J-way multiplicative models and a decomposable model for three-way layouts. The estimators which improve on the maximum likelihood estimators under the normalized squared error losses are provided for each model. The proposed estimators correspond to the ones by Clevenson and Zidek[2], Tsui and Press[11] and Chou[1].

Keywords and phrases : multiplicative Poisson model, decomposable Poisson model, shrinkage estimation, Bayes estimation, unbiased estimation of risk difference.

1 Introduction

Consider a two-way contingency table $x_{i_1i_2}$, $i_1 = 1,\dots,I_1$, $i_2 = 1,\dots,I_2$, where $x_{i_1i_2}$ are independent Poisson random variables with means $\lambda_{i_1i_2}$. The conditional distribution of $x_{i_1i_2}$ given x_{++} is the multinomial distribution $\mathrm{Mn}(x_{++}, p_{11},\dots,p_{I_1I_2})$ with $p_{i_1i_2} = \lambda_{i_1i_2}/\lambda$, $\lambda = \lambda_{++}$, where $+$ denotes summation over index. The independent model is described as

$$p_{i_1i_2} = p_{i_1+}p_{+i_2} \quad \text{or} \quad \lambda_{i_1i_2} = \frac{\lambda_{i_1+}\lambda_{+i_2}}{\lambda}.$$

This is equivalent to

$$\lambda_{i_1i_2} = \lambda\alpha_{1i_1}\alpha_{2i_2}, \tag{1}$$

$$\sum_{i_j=1}^{I_j}\alpha_{ji_j} = 1 \quad \text{for } j = 1,2.$$

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

The model (1) is generally known as the multiplicative Poisson model. In this article we first address the simultaneous estimation of $\boldsymbol{\lambda} = (\lambda_{11}, \dots, \lambda_{I_1 I_2})'$ in this model from the decision theoretic viewpoint.

A series of results on the shrinkage estimation of multivariate Poisson means which has its origin in Clevenson and Zidek[2] corresponds to the estimation of $\boldsymbol{\lambda}$ in the saturated model in which there are no constraints on $\lambda_{i_1 i_2}$. Clevenson and Zidek[2] proved the inadmissibility of the maximum likelihood estimator(MLE) when $I_1 \cdot I_2 \geq 2$ and provided a class of estimators which dominate the MLE under the normalized squared loss

$$\mathrm{L}(\boldsymbol{\lambda}, \hat{\boldsymbol{\lambda}}) = \sum_{i_1, i_2} \frac{1}{\lambda_{i_1 i_2}} (\hat{\lambda}_{i_i i_2} - \lambda_{i_1 i_2})^2. \tag{2}$$

Their class includes proper Bayes, i.e. admissible estimators. Since then, considerable research has been devoted to this problem.

Tsui and Press[11] enlarged the class of Clevenson and Zidek and also derived improved estimators under k-normalized squared error loss, $\mathrm{L}_k(\boldsymbol{\lambda}, \hat{\boldsymbol{\lambda}}) = \sum_{i_1, i_2} (\hat{\lambda}_{i_i i_2} - \lambda_{i_1 i_2})^2 / \lambda_{i_1 i_2}^k$. Hwang[7] extended their results to the estimation of the mean parameters of a subclass of discrete exponential families. He generalized the identity of Hudson[6] and derived an unbiased estimator of the difference of two risk functions. Chou[1] gave a class of improved estimators for a wider class of discrete exponential families. In the setting of simultaneous prediction of Poisson random variables, Komaki[9] derived fundamental results on admissibility under the Kullback-Leibler loss. Other important results in this field may be found, for example, in Ghosh and Parsian[4], Ghosh, Hwang and Tsui[3], Ghosh and Yang[4], Johnstone[8], Tsui[10].

We consider to apply these arguments to the estimation of $\boldsymbol{\lambda}$ for the multiplicative Poisson model (1). In section 2 we provide two classes of estimators which dominate the MLE under the loss (2). The one corresponds exactly to the class of Chou[1]. The other is an extension of the Bayes estimator of Clevenson and Zidek. In section 3 we extend the results in section 2 to the multiplicative Poisson models for multi-way contingency tables and prove that Chou-type estimators dominate the MLE.

These results suggest that we may apply the argument in section 2 and 3 to the estimation of the means of more general Poisson log linear models. In section 4 we take up the decomposable Poisson model for three-way layouts,

$$x_{i_1 i_2 i_3} \sim \mathrm{Po}\left(\lambda_{i_1 i_2 i_3}\right), \quad \lambda_{i_1 i_2 i_3} = \lambda \frac{\alpha_{i_1 i_2} \beta_{i_2 i_3}}{\gamma_{i_2}},$$

$$i_j = 1, \dots, I_j, \quad I_j \geq 2, \quad j = 1, 2, 3,$$

$$\sum_{i_1} \alpha_{i_1 i_2} = \sum_{i_3} \beta_{i_2 i_3} = \gamma_{i_2}, \quad \sum_{i_2} \gamma_{i_2} = 1.$$

We provide three classes of improved estimators also in this model.

Section 5 gives some Monte Carlo studies which confirm these theoretical results of the dominance relationship. In section 6 we give some concluding remarks.

2 Estimation of the means in the multiplicative Poisson model for two-way contingency tables

In this section we consider the simultaneous estimation of $\boldsymbol{\lambda} = (\lambda_{11}, \ldots, \lambda_{I_1 I_2})'$ in two-way multiplicative model (1) under the loss function (2). In this model $\boldsymbol{x}_1 = (x_{1+}, \ldots, x_{I_1+})'$ and $\boldsymbol{x}_2 = (x_{+1}, \ldots, x_{+I_2})'$ are the complete sufficient statistics. The joint probability function of $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ is easily obtained by

$$\Pr(\boldsymbol{x}_1, \boldsymbol{x}_2) = \exp(-\lambda)\lambda^{x_{++}} \prod_{i_1} \alpha_{1i_1}^{x_{i_1+}} \prod_{i_2} \alpha_{2i_2}^{x_{+i_2}} \cdot t(\boldsymbol{x}_1, \boldsymbol{x}_2), \tag{3}$$

where

$$t(\boldsymbol{x}_1, \boldsymbol{x}_2) = \frac{x_{++}!}{\prod_{i_1} x_{i_1+}! \prod_{i_2} x_{+i_2}!}.$$

From (3) the MLE of $\lambda_{i_1 i_2}$ is

$$\boldsymbol{\delta}^{ML} = (\delta_{11}^{ML}, \ldots, \delta_{I_1 I_2}^{ML})', \quad \delta_{i_1 i_2}^{ML} = \begin{cases} \dfrac{x_{i_1+} x_{+i_2}}{x_{++}} & \text{if } x_{++} \neq 0, \\ 0 & \text{if } x_{++} = 0. \end{cases}$$

The following lemma corresponds to the identity of Hudson[6] and Hwang[7].

Lemma 2.1 *Let $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ have the probability function (3). If g is a real valued function with $E|g(\boldsymbol{x}_1, \boldsymbol{x}_2)| < \infty$ and $g(\boldsymbol{x}_1, \boldsymbol{x}_2) = 0$ whenever $x_{i_1+} < m$ or $x_{+i_2} < m$, then*

$$\begin{aligned} &E\left[\frac{1}{\lambda_{i_1 i_2}^m} \cdot g(\boldsymbol{x}_1 + m\boldsymbol{e}_{i_1}, \boldsymbol{x}_2 + m\boldsymbol{e}_{i_2})\right] \\ &\quad = E\left[\frac{t(\boldsymbol{x}_1 + m\boldsymbol{e}_{i_1}, \boldsymbol{x}_2 + m\boldsymbol{e}_{i_2})}{t(\boldsymbol{x}_1, \boldsymbol{x}_2)} \cdot g(\boldsymbol{x}_1 + m\boldsymbol{e}_{i_1}, \boldsymbol{x}_2 + m\boldsymbol{e}_{i_2})\right], \end{aligned} \tag{4}$$

where $\boldsymbol{e}_{i_j}$ are the $I_j \times 1$ vectors with 1 as the i_j-th component and 0 for others.

The proof is the same as in Hudson[6] and Hwang[7] and is omitted here. From this lemma with $m = -1$ and $g(\boldsymbol{x}_1, \boldsymbol{x}_2) \equiv 1$, $\boldsymbol{\delta}^{ML}$ is found to be the uniformly minimum variance unbiased estimator(UMVUE) of $\boldsymbol{\lambda}$.

We address the estimators which dominate $\boldsymbol{\delta}^{ML}$ under the loss function (2) from the following class

$$\boldsymbol{\delta}^{\psi} = \boldsymbol{\delta}^{ML} - \boldsymbol{\Psi}(\boldsymbol{x}_1, \boldsymbol{x}_2),$$

where $\boldsymbol{\Psi}(\boldsymbol{x}_1, \boldsymbol{x}_2) = (\Psi_{11}(\boldsymbol{x}_1, \boldsymbol{x}_2), \ldots, \Psi_{I_1 I_2}(\boldsymbol{x}_1, \boldsymbol{x}_2))'$. By using (4) in Lemma 2.1 with $m = 1$, the difference between two risk functions of $\boldsymbol{\delta}^{ML}$ and $\boldsymbol{\delta}^{\psi}$ is expressed as

$$\begin{aligned} &\mathrm{R}(\boldsymbol{\lambda}, \boldsymbol{\delta}^{ML}) - \mathrm{R}(\boldsymbol{\lambda}, \boldsymbol{\delta}^{\psi}) \\ &\quad = \mathrm{E}[\mathrm{L}(\boldsymbol{\lambda}, \boldsymbol{\delta}^{ML}) - \mathrm{L}(\boldsymbol{\lambda}, \boldsymbol{\delta}^{\psi})] \\ &\quad = \sum_{i_1, i_2} \mathrm{E}\left[\frac{1}{\lambda_{i_1 i_2}} \Psi_{i_1 i_2}(\boldsymbol{x}_1, \boldsymbol{x}_2)(2\delta_{i_1 i_2}^{ML} - \Psi_{i_1 i_2}(\boldsymbol{x}_1, \boldsymbol{x}_2)) - 2\Psi_{i_1 i_2}(\boldsymbol{x}_1, \boldsymbol{x}_2)\right] \end{aligned}$$

$$= \sum_{i_1,i_2} \mathrm{E}\left[2\left(\Psi_{i_1i_2}(\boldsymbol{x}_1+\boldsymbol{e}_{i_1},\boldsymbol{x}_2+\boldsymbol{e}_{i_2}) - \Psi_{i_1i_2}(\boldsymbol{x}_1,\boldsymbol{x}_2)\right)\right]$$
$$- \sum_{i_1,i_2} \mathrm{E}\left[\frac{x_{++}+1}{(x_{i_1+}+1)(x_{+i_2}+1)}\Psi^2_{i_1i_2}(\boldsymbol{x}_1+\boldsymbol{e}_{i_1},\boldsymbol{x}_2+\boldsymbol{e}_{i_2})\right].$$

The last term implies that

$$\begin{aligned}\hat{\mathrm{Rd}}(\boldsymbol{\delta}^{\psi}) &= 2\sum_{i_1,i_2}\left(\Psi_{i_1i_2}(\boldsymbol{x}_1+\boldsymbol{e}_{i_1},\boldsymbol{x}_2+\boldsymbol{e}_{i_2}) - \Psi_{i_1i_2}(\boldsymbol{x}_1,\boldsymbol{x}_2)\right)\\ &\quad -\sum_{i_1,i_2}\left(\frac{x_{++}+1}{(x_{i_1+}+1)(x_{+i_2}+1)}\Psi^2_{i_1i_2}(\boldsymbol{x}_1+\boldsymbol{e}_{i_1},\boldsymbol{x}_2+\boldsymbol{e}_{i_2})\right) \qquad (5)\end{aligned}$$

is the UMVUE of $\mathrm{R}(\boldsymbol{\lambda},\boldsymbol{\delta}^{ML}) - \mathrm{R}(\boldsymbol{\lambda},\boldsymbol{\delta}^{\psi})$. In what follows we derive a sufficient condition on $\boldsymbol{\Psi}$ to satisfy $\mathrm{Rd}(\boldsymbol{\delta}^{\psi}) \geq 0$ for two classes of $\boldsymbol{\Psi}$.

We begin with the following class of Chou[1]-type estimators,

$$\begin{aligned}\boldsymbol{\delta}^{\phi,\gamma} &= (\delta^{\phi,\gamma}_{11},\ldots,\delta^{\phi,\gamma}_{I_1I_2})',\\ \delta^{\phi,\gamma}_{i_1i_2} &= \delta^{ML}_{i_1i_2}\left(1-\frac{\phi(x_{++})}{(x_{++}+c)^{\gamma+1}}\right), \qquad (6)\end{aligned}$$

where $\phi(\cdot)$ is a nondecreasing function which is not identically zero and $c>0$ and $\gamma \geq 0$ are constants. In the next theorem a sufficient condition on ϕ to improve on $\boldsymbol{\delta}^{ML}$ is presented.

Theorem 2.1 *If $\phi(\cdot)$ is nondecreasing and satisfies*

$$0 \leq \phi(\cdot) \leq \min\left(\ 2(I_1+I_2)-2\gamma-4,\ \ 2c-2\gamma,\ \ (x_{++}+c)^{\gamma+1}\ \right),$$

then $\boldsymbol{\delta}^{\phi,\gamma}$ dominates $\boldsymbol{\delta}^{ML}$ under the loss function (2).

(Proof)

We can prove this theorem by using the procedure in the proof of Theorem 2.1 in Chou[1]. From (5) $\hat{\mathrm{Rd}}(\boldsymbol{\delta}^{\phi,\gamma})$ is

$$\begin{aligned}\hat{\mathrm{Rd}}(\boldsymbol{\delta}^{\phi,\gamma}) &= 2\sum_{i_1,i_2}\left\{\frac{\phi(x_{++}+1)}{B^{\gamma+1}}\cdot\frac{(x_{i_1+}+1)(x_{+i_2}+1)}{x_{++}+1} - \frac{\phi(x_{++})}{(B-1)^{\gamma+1}}\cdot\frac{x_{i_1+}x_{+i_2}}{x_{++}}\right\}\\ &\quad -\sum_{i_1,i_2}\frac{\phi^2(x_{++}+1)}{B^{2\gamma+2}}\cdot\frac{(x_{i_1+}+1)(x_{+i_2}+1)}{x_{++}+1}\\ &= 2\left\{\frac{\phi(x_{++}+1)}{B^{\gamma+1}}\cdot\frac{(x_{++}+I_1)(x_{++}+I_2)}{x_{++}+1} - \frac{\phi(x_{++})x_{++}}{(B-1)^{\gamma+1}}\right\}\\ &\quad -\frac{\phi^2(x_{++}+1)}{B^{2\gamma+2}}\cdot\frac{(x_{++}+I_1)(x_{++}+I_2)}{x_{++}+1}, \qquad (7)\end{aligned}$$

where $B = x_{++}+c+1$. Since $\phi(\cdot)$ is nondecreasing and $B>1$,

$$2\left\{\frac{\phi(x_{++}+1)}{B^{\gamma+1}}\cdot\frac{(x_{++}+I_1)(x_{++}+I_2)}{x_{++}+1} - \frac{\phi(x_{++})x_{++}}{(B-1)^{\gamma+1}}\right\}$$

$$
\begin{aligned}
& -\frac{\phi^2(x_{++}+1)}{B^{2\gamma+2}} \cdot \frac{(x_{++}+I_1)(x_{++}+I_2)}{x_{++}+1} \\
\geq\ & \frac{\phi(x_{++}+1)}{B^{\gamma+2}(B-1)^{\gamma+1}} \Big\{ 2B(B-1)^{\gamma+1} \frac{(x_{++}+I_1)(x_{++}+I_2)}{x_{++}+1} - 2B^{\gamma+2}x_{++} \\
& \qquad - \phi(x_{++}+1)(B-1)^{\gamma+1} \frac{(x_{++}+I_1)(x_{++}+I_2)}{x_{++}+1} \Big\} \\
\geq\ & \frac{\phi(x_{++}+1)}{B^{\gamma+2}(B-1)^{\gamma+1}} \Big\{ 2B^{\gamma+1}(B-(\gamma+1)) \frac{(x_{++}+I_1)(x_{++}+I_2)}{x_{++}+1} - 2B^{\gamma+2}x_{++} \\
& \qquad - \phi(x_{++}+1)B^{\gamma+1} \frac{(x_{++}+I_1)(x_{++}+I_2)}{x_{++}+1} \Big\} \\
=\ & \frac{\phi(x_{++}+1)}{B(B-1)^{\gamma+1}} \Big\{ (2(I_1+I_2)-2\gamma-4-\phi(x_{++}+1))x_{++} \\
& \qquad + (2c-2\gamma-\phi(x_{++}+1))(I_1+I_2-1) \\
& \qquad + (2(c+x_{++})-2\gamma-\phi(x_{++}+1)) \frac{(I_1-1)(I_2-1)}{x_{++}+1} \Big\}
\end{aligned}
$$

The second inequality follows from the fact that $(B-1)^{\gamma+1} \geq B^{\gamma+1} - (\gamma+1)B^{\gamma}$ for $\gamma \geq 0$. The right hand side is always nonnegative if $\phi(\cdot) \leq \min(2(I_1+I_2)-2\gamma-4, 2c-2\gamma)$, which completes the proof. ∎

$\delta^{\phi,0}$ corresponds to the estimators by Tsui and Press[11]. If $c \geq I_1+I_2-1$, $\delta^{\phi,0}$ corresponds to the estimator by Clevenson and Zidek[2].

Next we consider the following class,

$$
\begin{aligned}
\boldsymbol{\delta}^{\nu} &= (\delta^{\nu}_{11}, \ldots, \delta^{\nu}_{I_1 I_2})', \\
\delta^{\nu}_{i_1 i_2} &= \frac{x_{i_1+}x_{+i_2}}{(x_{++}+I_1-1)(x_{++}+I_2-1)} \cdot \frac{x_{++}+I_1+I_2-2}{x_{++}+\nu+I_1+I_2-2} \cdot x_{++},
\end{aligned} \tag{8}
$$

where $\nu \geq 0$ is a constant. We note that for $\nu > 0$ this class is not included in (6). $\boldsymbol{\delta}^{\nu}$ is the Bayes estimator with respect to the prior measure $\pi(\lambda, \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2 | \nu)$,

$$
\pi(\lambda, \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2|\nu) \mathrm{d}\lambda \prod_{i_1=1}^{I_1-1} \mathrm{d}\alpha_{1i_1} \prod_{i_2=1}^{I_2-1} \mathrm{d}\alpha_{2i_2} = m(\lambda)\mathrm{d}\lambda \prod_{i_1=1}^{I_1-1} \mathrm{d}\alpha_{1i_1} \prod_{i_2=1}^{I_2-1} \mathrm{d}\alpha_{2i_2},
$$

$$
m(\lambda) = \int_0^{\infty} (1+\lambda t)^{-\nu} t^{-(I_1+I_2-1)} \exp(-t^{-1}) \mathrm{d}t,
$$

$$
\boldsymbol{\alpha}_1 = (\alpha_{11}, \ldots, \alpha_{1I_1}), \quad \boldsymbol{\alpha}_2 = (\alpha_{21}, \ldots, \alpha_{2I_2}).
$$

For $\nu > 1$, $\pi(\lambda, \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \nu)$ is proper and therefore $\boldsymbol{\delta}^{\nu}$ is admissible. This estimator is a natural extension of the Bayes estimator discussed in Clevenson and Zidek[2].

The following theorem gives the condition on ν to dominate $\boldsymbol{\delta}^{ML}$.

Theorem 2.2 *Suppose that $I_1 \geq 2$ and $I_2 \geq 2$. If $0 \leq \nu < I_1+I_2-2$, $\boldsymbol{\delta}^{\nu}$ improves on $\boldsymbol{\delta}^{ML}$ under the loss function(2). If $\nu > 1$, $\boldsymbol{\delta}^{\nu}$ is admissible.*

This result is consistent to the one of Clevenson and Zidek[2]. The proof is complicated. See the Appendix for the details.

3 Extension to the Poisson multiplicative model for multi-way contingency tables

We can extend the results of Theorem 2.1 to multi-way multiplicative models. Extension of Theorem 2.2 seems to be technically difficult. J-way multiplicative model is expressed as

$$x_{i_1 i_2 \cdots i_J} \sim \mathrm{Po}(\lambda_{i_1 i_2 \cdots i_J}), \quad \lambda_{i_1 i_2 \cdots i_J} = \lambda \alpha_{1 i_1} \alpha_{2 i_2} \cdots \alpha_{J i_J}, \tag{9}$$
$$i_j = 1, \ldots, I_j, \quad I_j \geq 2, \quad j = 1, \ldots, J,$$
$$\sum_{i_j} \alpha_{j i_j} = 1, \ j = 1, \ldots, J,$$

where $x_{i_1 i_2 \cdots i_J}$ are assumed to be independently distributed. The problem is to estimate $\boldsymbol{\lambda} = \{\lambda_{i_1 i_2 \cdots i_J}\}$ simultaneously under the loss function

$$\mathrm{L}(\boldsymbol{\lambda}, \hat{\boldsymbol{\lambda}}) = \sum_{j=1}^{J} \sum_{i_j=1}^{I_j} \frac{1}{\lambda_{i_1 i_2 \cdots i_J}} (\hat{\lambda}_{i_1 i_2 \cdots i_J} - \lambda_{i_1 i_2 \cdots i_J})^2. \tag{10}$$

Denote the j-th one-dimensional marginal frequencies and the total frequency by

$$x^+_{j,i_j} = \sum_{i_s : s \neq j} x_{i_1 i_2 \cdots i_J}, \quad x^+ = \sum_{i_j} x^+_{j,i_j}.$$

Then the set of one-dimensional marginal frequencies $\boldsymbol{x}^+_j = (x^+_{j,1}, \ldots, x^+_{j,I_j})'$, $j = 1, \ldots, J$, is the complete sufficient statistic. Its joint probability function is given by

$$\Pr(\boldsymbol{x}^+_1, \ldots, \boldsymbol{x}^+_J) = \exp(-\lambda) \lambda^{x^+} \prod_j \prod_{i_j} \alpha_{j i_j}^{x^+_{j,i_j}} \cdot t(\boldsymbol{x}^+_1, \ldots, \boldsymbol{x}^+_J), \tag{11}$$

where

$$t(\boldsymbol{x}^+_1, \ldots, \boldsymbol{x}^+_J) = \frac{(x^+!)^{J-1}}{\prod_j \prod_{i_j} x^+_{j,i_j}!}.$$

From (11) the MLE is

$$\boldsymbol{\delta}^{ML} = \{\delta^{ML}_{i_1 i_2 \cdots i_J}\}, \quad \delta^{ML}_{i_1 i_2 \cdots i_J} = \begin{cases} \dfrac{\prod_j x^+_{j,i_j}}{(x^+)^{J-1}} & \text{if } x^+ \neq 0, \\ 0 & \text{if } x^+ = 0. \end{cases}$$

The following lemma is the identity which corresponds to (4).

Lemma 3.1 *Let $(\boldsymbol{x}^+_1, \ldots, \boldsymbol{x}^+_J)$ have the probability function (11). If g is a real valued function with $E|g(\boldsymbol{x}^+_1, \ldots, \boldsymbol{x}^+_J)| < \infty$ and $g(\boldsymbol{x}^+_1, \ldots, \boldsymbol{x}^+_J) = 0$ whenever $x^+_{j,i_j} < m$ for some (j, i_j), then*

$$E\left[\frac{1}{\lambda^m_{i_1 i_2 \cdots i_J}} g(\boldsymbol{x}^+_1, \ldots, \boldsymbol{x}^+_J)\right]$$
$$= E\left[\frac{t(\boldsymbol{x}^+_1 + m\boldsymbol{e}_{i_1}, \ldots, \boldsymbol{x}^+_J + m\boldsymbol{e}_{i_J})}{t(\boldsymbol{x}^+_1, \ldots, \boldsymbol{x}^+_J)} g(\boldsymbol{x}^+_1 + m\boldsymbol{e}_{i_1}, \ldots, \boldsymbol{x}^+_J + m\boldsymbol{e}_{i_J})\right], \tag{12}$$

where $\boldsymbol{e}_{i_j}$ are given as in Lemma 2.1.

The proof is omitted here. In the same way as the previous section $\boldsymbol{\delta}^{ML}$ is found to be the UMVUE of $\boldsymbol{\lambda}$ also for J-way multiplicative model.

We consider the following class of Chou-type estimators,

$$\begin{aligned}\boldsymbol{\delta}^{\phi,\gamma} &= \{\delta^{\phi,\gamma}_{i_1 i_2 \cdots i_J}\},\\ \delta^{\phi,\gamma}_{i_1 i_2 \cdots i_J} &= \delta^{ML}_{i_1 i_2 \cdots i_J}\left(1 - \frac{\phi(x^+)}{(x^+ + c)^{\gamma+1}}\right),\end{aligned}$$

where $\phi(\cdot)$, c, and γ are given as in the previous section. The following theorem generalizes Theorem 2.1 to J-way layouts.

Theorem 3.1 *Suppose that $I_j \geq 2$, $j = 1, \ldots, J$. If $\phi(\cdot)$ is nondecreasing and satisfies*

$$0 \leq \phi(\cdot) \leq \min\left(\ 2\textstyle\sum_{j=1}^J I_j - 2\gamma - 2J,\ \ 2c - 2\gamma,\ \ (x^+ + c)^{\gamma+1}\ \right),$$

then $\boldsymbol{\delta}^{\phi,\gamma}$ dominates $\boldsymbol{\delta}^{ML}$ under the loss function (10).

We use the following lemma to prove this theorem.

Lemma 3.2

$$\frac{\prod_{j=1}^J (x^+ + I_j)}{(x^+ + 1)^{J-1}} = x^+ + \sum_{j=1}^J I_j - (J-1) + \xi_J(x^+), \tag{13}$$

where $\xi_J(\cdot)$ is a positive function.

(Proof)
We prove this lemma by induction. When $J = 2$, $\xi_2(x^+)$ is

$$\xi_2(x^+) = \frac{(I_1 - 1)(I_2 - 1)}{x^+ + 1} > 0.$$

Suppose that (13) holds for $J = J_0$. Then

$$\begin{aligned}\frac{\prod_{j=1}^{J_0+1}(x^+ + I_j)}{(x^+ + 1)^{J_0}} &= \left(x^+ + \sum_{j=1}^{J_0} I_j - (J_0 - 1) + \xi_{J_0}(x^+)\right)\left(1 + \frac{I_{J_0+1} - 1}{x^+ + 1}\right)\\ &= x^+ + \sum_{j=1}^{J_0+1} I_j - J_0 + \frac{(I_{J_0+1} - 1)(\sum_{j=1}^{J_0} I_j - J_0)}{x^+ + 1} + \xi_{J_0}(x^+)\frac{I_{J_0+1} + x^+}{x^+ + 1}.\end{aligned}$$

Since $I_j \geq 2$ for $j = 1, \ldots, J$,

$$\xi_{J_0+1}(x^+) \equiv \frac{(I_{J_0+1} - 1)(\sum_{j=1}^{J_0} I_j - J_0)}{x^+ + 1} + \xi_{J_0}(x^+)\frac{I_{J_0+1} + x^+}{x^+ + 1} > 0,$$

which completes the proof. ∎

(Proof of Theorem 3.1)

By using Lemma 3.1 with $m = 1$, the UMVUE of the difference between the risk functions of $\boldsymbol{\delta}^{ML}$ and $\boldsymbol{\delta}^{\phi,\gamma}$ is expressed by

$$\begin{aligned}
\hat{\text{Rd}}(\boldsymbol{\delta}^{\phi,\gamma}) &= 2\sum_{j=1}^{J}\sum_{i_j=1}^{I_j}\left\{\frac{\phi(x^+ + 1)}{B^{\gamma+1}}\cdot\frac{\prod_j(x^+_{j,i_j}+1)}{(x^+ + 1)^{J-1}} - \frac{\phi(x^+)}{(B-1)^{\gamma+1}}\cdot\frac{\prod_j x^+_{j,i_j}}{(x^+)^{J-1}}\right\} \\
&\quad - \sum_{j=1}^{J}\sum_{i_j=1}^{I_j}\frac{\phi^2(x^+ + 1)}{B^{2\gamma+2}}\cdot\frac{\prod_j(x^+_{j,i_j}+1)}{(x^+ + 1)^{J-1}} \\
&= 2\left\{\frac{\phi(x^+ + 1)}{B^{\gamma+1}}\cdot\frac{\prod_j(x^+ + I_j)}{(x^+ + 1)^{J-1}} - \frac{\phi(x^+)x^+}{(B-1)^{\gamma+1}}\right\} \\
&\quad - \frac{\phi^2(x^+ + 1)}{B^{2\gamma+2}}\cdot\frac{\prod_j(x^+ + I_j)}{(x^+ + 1)^{J-1}},
\end{aligned}$$

where $B = x^+ + c + 1$. By using (13), we can show the following in the same way as (8),

$$\begin{aligned}
&2\left\{\frac{\phi(x^+ + 1)}{B^{\gamma+1}}\cdot\frac{\prod_j(x^+ + I_j)}{(x^+ + 1)^{J-1}} - \frac{\phi(x^+)x^+}{(B-1)^{\gamma+1}}\right\} \\
&\qquad - \frac{\phi^2(x^+ + 1)}{B^{2\gamma+2}}\cdot\frac{\prod_j(x^+ + I_j)}{(x^+ + 1)^{J-1}} \\
\geq\ & \frac{\phi(x^+ + 1)}{B^{\gamma+2}(B-1)^{\gamma+1}}\left\{2B^{\gamma+1}(B-(\gamma+1))\frac{\prod_j(x^+ + I_j)}{(x^+ + 1)^{J-1}} - 2B^{\gamma+2}x^+\right. \\
&\qquad\qquad \left. - \phi(x^+ + 1)B^{\gamma+1}\frac{\prod_j(x^+ + I_j)}{(x^+ + 1)^{J-1}}\right\} \\
=\ & \frac{\phi(x^+ + 1)}{B(B-1)^{\gamma+1}}\left\{\Big(2\sum_{j=1}^{J} I_j - 2\gamma - 2J - \phi(x^+ + 1)\Big)x^+\right. \\
&\qquad + (2c - 2\gamma - \phi(x^+ + 1))(I_1 + I_2 - 1) \\
&\qquad \left. + (2(c + x^+) - 2\gamma - \phi(x^+ + 1))\xi_J(x^+)\right\},
\end{aligned}$$

which completes the proof. ∎

4 Estimation in Poisson decomposable model for three-way contingency tables

The results in the previous sections suggest that we may be able to apply the arguments to more general Poisson log linear models. In this section we take up the following decomposable model for three-way layouts,

$$x_{i_1 i_2 i_3} \sim \text{Po}\left(\lambda_{i_1 i_2 i_3}\right), \quad \lambda_{i_1 i_2 i_3} = \lambda\frac{\alpha_{i_1 i_2}\beta_{i_2 i_3}}{\gamma_{i_2}}, \tag{14}$$

$$i_j = 1, \ldots, I_j, \quad I_j \geq 2, \quad j = 1, 2, 3,$$

$$\sum_{i_1}\alpha_{i_1 i_2} = \sum_{i_3}\beta_{i_2 i_3} = \gamma_{i_2}, \quad \sum_{i_2}\gamma_{i_2} = 1.$$

The problem is to estimate $\boldsymbol{\lambda} = (\lambda_{111}, \ldots, \lambda_{I_1 I_2 I_3})'$ under the loss function

$$\mathrm{L}(\boldsymbol{\lambda}, \hat{\boldsymbol{\lambda}}) = \sum_{i_1, i_2, i_3} \frac{1}{\lambda_{i_1 i_2 i_3}} (\hat{\lambda}_{i_1 i_2 i_3} - \lambda_{i_1 i_2 i_3})^2. \tag{15}$$

Denote the relevant two-dimensional marginal frequencies by

$$\boldsymbol{x}_{1 i_2 +} = (x_{1 i_2 +}, \ldots, x_{I_1 i_2 +})', \quad \boldsymbol{x}_{+ i_2 3} = (x_{+ i_2 1}, \ldots, x_{+ i_2 I_3})',$$

$$\boldsymbol{x}_{12} = (\boldsymbol{x}'_{11+}, \ldots, \boldsymbol{x}'_{1 I_2 +})', \quad \boldsymbol{x}_{23} = (\boldsymbol{x}'_{+13}, \ldots, \boldsymbol{x}'_{+ I_2 3})'.$$

In this model $(\boldsymbol{x}_{12}, \boldsymbol{x}_{23})$ is the complete sufficient statistic. The joint probability function of the sufficient statistic is

$$\Pr(\boldsymbol{x}_{12}, \boldsymbol{x}_{23}) = \exp(-\lambda)\ \lambda^{x_{+++}} \frac{\prod_{i_1, i_2} \alpha_{i_1 i_2}^{x_{i_1 i_2 +}} \prod_{i_2, i_3} \beta_{i_2 i_3}^{x_{+ i_2 i_3}}}{\prod_{i_2} \gamma_{i_2}^{x_{+ i_2 +}}}\ t^+(\boldsymbol{x}_{12}, \boldsymbol{x}_{23}), \tag{16}$$

$$t^+(\boldsymbol{x}_{12}, \boldsymbol{x}_{23}) = \prod_{i_2} t(\boldsymbol{x}_{1 i_2 +}, \boldsymbol{x}_{+ i_2 3}), \quad t(\boldsymbol{x}_{1 i_2 +}, \boldsymbol{x}_{+ i_2 3}) = \frac{x_{+ i_2 +}!}{\prod_{i_1} x_{i_1 i_2 +}! \prod_{i_3} x_{+ i_2 i_3}!}.$$

The MLE of $\lambda_{i_1 i_2 i_3}$ is

$$\boldsymbol{\delta}^{ML} = (\delta_{111}^{ML}, \ldots, \delta_{I_1 I_2 I_3}^{ML})', \quad \delta_{i_1 i_2 i_3}^{ML} = \begin{cases} \dfrac{x_{i_1 i_2 +} x_{+ i_2 i_3}}{x_{+ i_2 +}} & \text{if } x_{+ i_2 +} \neq 0, \\ 0 & \text{if } x_{+ i_2 +} = 0. \end{cases}$$

The identity for this model is as follows.

Lemma 4.1 *Let $\boldsymbol{x}_{12}$ and $\boldsymbol{x}_{23}$ have the probability function (16). Let $g(\cdot)$, $\mathbf{e}_{i_1}$ and $\mathbf{e}_{i_3}$ be given as in Lemma 2.1. Define $\mathbf{e}_{i_1 i_2}$ by the $I_1 I_2 \times 1$ vectors with 1 as the $(i_1 + I_2(i_2 - 1))$-th component and 0 for others. $\mathbf{e}_{i_2 i_3}$ is defined in the same way as $\mathbf{e}_{i_1 i_2}$. Then*

$$\begin{aligned} &E\left[\frac{1}{\lambda_{i_1 i_2 i_3}^m} g(\boldsymbol{x}_{12}, \boldsymbol{x}_{23})\right] \\ &\quad = E\left[\frac{t(\boldsymbol{x}_{1 i_2 +} + m\boldsymbol{e}_{i_1}, \boldsymbol{x}_{+ i_2 3} + m\boldsymbol{e}_{i_3})}{t(\boldsymbol{x}_{1 i_2 +}, \boldsymbol{x}_{+ i_2 3})} \cdot g(\boldsymbol{x}_{12} + m\boldsymbol{e}_{i_1 i_2}, \boldsymbol{x}_{23} + m\boldsymbol{e}_{i_2 i_3})\right]. \end{aligned} \tag{17}$$

This lemma can be also proved in the same way as Hudson[6] and Hwang[7]. Specializing (17) to g which depends only on $\boldsymbol{x}_{1 i_2 +}, \boldsymbol{x}_{+ i_2 3}$ we obtain the following identity.

$$\begin{aligned} &\mathrm{E}\left[\frac{1}{\lambda_{i_1 i_2 i_3}^m} g(\boldsymbol{x}_{1 i_2 +}, \boldsymbol{x}_{+ i_2 3})\right] \\ &\quad = \mathrm{E}\left[\frac{t(\boldsymbol{x}_{1 i_2 +} + m\boldsymbol{e}_{i_1}, \boldsymbol{x}_{+ i_2 3} + m\boldsymbol{e}_{i_3})}{t(\boldsymbol{x}_{1 i_2 +}, \boldsymbol{x}_{+ i_2 3})} \cdot g(\boldsymbol{x}_{1 i_2 +} + m\boldsymbol{e}_{i_1}, \boldsymbol{x}_{+ i_2 3} + m\boldsymbol{e}_{i_3})\right]. \end{aligned} \tag{18}$$

From (16) if i_2 is fixed, the model (14) is reduced to two-way multiplicative model (1). Thus we first consider the following class,

$$\boldsymbol{\delta}^{\psi} = \boldsymbol{\delta}^{ML} - \boldsymbol{\Psi}(\boldsymbol{x}_{12}, \boldsymbol{x}_{23}) = (\delta_{111}^{\psi}, \ldots, \delta_{I_1 I_2 I_3}^{\psi})' \tag{19}$$

$$\boldsymbol{\Psi}(\boldsymbol{x}_{12}, \boldsymbol{x}_{23}) = \Big\{ \Psi_{i_1 i_2 i_3}(\boldsymbol{x}_{1 i_2 +}, \boldsymbol{x}_{+ i_2 3}) \Big\}.$$

By using (18) in Lemma 4.1 $\hat{\mathrm{Rd}}(\boldsymbol{\delta}^{\Psi})$ is expressed by

$$\begin{aligned}
\hat{\mathrm{Rd}}(\boldsymbol{\delta}^{\psi}) &= \sum_{i_2} \hat{\mathrm{Rd}}(\boldsymbol{\delta}^{\psi}_{i_2}), \\
\hat{\mathrm{Rd}}(\boldsymbol{\delta}^{\psi}_{i_2}) &= 2 \sum_{i_1, i_3} \Big(\Psi_{i_1 i_2 i_3}(\boldsymbol{x}_{1 i_2 +} + \boldsymbol{e}_{i_1}, \boldsymbol{x}_{+ i_2 3} + \boldsymbol{e}_{i_3}) - \Psi_{i_1 i_2 i_3}(\boldsymbol{x}_{1 i_2 +}, \boldsymbol{x}_{+ i_2 3}) \Big) \\
&\quad - \sum_{i_1, i_3} \left(\frac{x_{+ i_2 +} + 1}{(x_{i_1 i_2 +} + 1)(x_{+ i_2 i_3} + 1)} \Psi^2_{i_1 i_2 i_3}(\boldsymbol{x}_{1 i_2 +} + \boldsymbol{e}_{i_1}, \boldsymbol{x}_{+ i_2 3} + \boldsymbol{e}_{i_3}) \right), \qquad (20)
\end{aligned}$$

where $\boldsymbol{\delta}^{\psi}_{i_2} = (\delta^{\psi}_{1 i_2 1}, \dots, \delta^{\psi}_{I_1 i_2 I_3})'$. (20) corresponds to (5). Since $\hat{\mathrm{Rd}}(\boldsymbol{\delta}^{\psi}_{i_2}) \geq 0$ for each i_2 implies $\hat{\mathrm{Rd}}(\boldsymbol{\delta}^{\psi}) \geq 0$, the following results on two classes of estimators can be obtained in the same way as Theorem 2.1 and 2.2.

Theorem 4.1 *Suppose that nondecreasing functions $\phi_{i_2}(\cdot)$, $i_2 = 1, \dots, I_2$, satisfy*

$$0 \leq \phi_{i_2}(\cdot) \leq \min\Big(\; 2(I_1 + I_3) - 2\gamma - 4, \quad 2c - 2\gamma, \quad (x_{+ i_2 +} + c)^{\gamma+1} \; \Big), \quad c > 0, \quad \gamma \geq 0.$$

Then $\boldsymbol{\delta}^{\phi_{i_2}, \gamma} = (\delta^{\phi_1, \gamma}_{111}, \dots, \delta^{\phi_{I_2}, \gamma}_{I_1 I_2 I_3})'$,

$$\delta^{\phi_{i_2}, \gamma}_{i_1 i_2 i_3} = \delta^{ML}_{i_1 i_2 i_3} \left(1 - \frac{\phi_{i_2}(x_{+ i_2 +})}{(x_{+ i_2 +} + c)^{\gamma+1}} \right) = \frac{x_{i_1 i_2 +} x_{+ i_2 i_3}}{n x_{+ i_2 +}} \left(1 - \frac{\phi_{i_2}(x_{+ i_2 +})}{(x_{+ i_2 +} + c)^{\gamma+1}} \right) \qquad (21)$$

dominates $\boldsymbol{\delta}^{ML}$ under the loss function (15).

Theorem 4.2 $\boldsymbol{\delta}^{\nu_{i_2}} = (\delta^{\nu_1}_{111}, \dots, \delta^{\nu_{I_2}}_{I_1 I_2 I_3})'$,

$$\delta^{\nu_{i_2}}_{i_1 i_2 i_3} = \frac{x_{i_1 i_2 +} x_{+ i_2 i_3}}{(x_{+ i_2 +} + I_1 - 1)(x_{+ i_2 +} + I_3 - 1)} \cdot \frac{x_{+ i_2 +} + I_1 + I_3 - 2}{x_{+ i_2 +} + \nu_{i_2} + I_1 + I_3 - 2} \cdot x_{+ i_2 +} \qquad (22)$$

dominates $\boldsymbol{\delta}^{ML}$ under the loss function (15) if $\nu_{i_2} < I_1 + I_3 - 2$, $i_2 = 1, \dots, I_2$. If $\nu_{i_2} > 1$, $i_2 = 1, \dots, I_2$, then $\boldsymbol{\delta}^{\nu_{i_2}}$ is admissible.

Let $\lambda_{i_2} = \lambda \gamma_{i_2}$, $\theta_{i_1 i_2} = \alpha_{i_1 i_2} / \gamma_{i_2}$ and $\eta_{i_2 i_3} = \beta_{i_2 i_3} / \gamma_{i_2}$. As we mentioned in Section 2, $\boldsymbol{\delta}^{\nu_{i_2}}$ is the Bayes estimator of $\boldsymbol{\lambda}$ with respect to the prior measure

$$\prod_{i_2=1}^{I_2} \pi(\lambda_{i_2}, \boldsymbol{\theta}_{i_2}, \boldsymbol{\eta}_{i_2} | \nu_{i_2}) \mathrm{d}\lambda_{i_2} \prod_{i_1} \mathrm{d}\theta_{i_1 i_2} \prod_{i_3} \mathrm{d}\eta_{i_2 i_3} = \prod_{i_2=1}^{I_2} m(\lambda_{i_2} | \nu_{i_2}) \mathrm{d}\lambda_{i_2} \prod_{i_1} \mathrm{d}\theta_{i_1 i_2} \prod_{i_3} \mathrm{d}\eta_{i_2 i_3},$$

$$m(\lambda_{i_2} | \nu_{i_2}) = \int_0^{\infty} (1 + \lambda_{i_2} t)^{-\nu_{i_2}} t^{-(I_1 + I_3 - 1)} \exp(-t^{-1}) \mathrm{d}t,$$

$$\boldsymbol{\theta}_{i_2} = (\theta_{1 i_2}, \dots, \theta_{I_1 i_2})', \quad \boldsymbol{\eta}_{i_2} = (\eta_{i_2 1}, \dots, \eta_{i_2 I_3})'.$$

So far we considered shrinkage separately for each i_2 slice. Next we consider the third class of estimators

$$\begin{aligned}
\boldsymbol{\delta}^{\phi, \gamma} &= (\delta^{\phi, \gamma}_{111}, \dots, \delta^{\phi, \gamma}_{I_1 I_2 I_3})', \\
\delta^{\phi, \gamma}_{i_1 i_2 i_3} &= \delta^{ML}_{i_1 i_2 i_3} \left(1 - \frac{\phi(x_{+++})}{(x_{+++} + c)^{\gamma+1}} \right), \qquad (23)
\end{aligned}$$

where x_{+++} is the total sample size. Note that in this class the amount of shrinkage depends only on the total sample size x_{+++} and hence this class does not belong to (19). We can find improved estimators also in this class.

Theorem 4.3 *If $\phi(\cdot)$ is nondecreasing and satisfies*

$$0 \le \phi(\cdot) \le \min\Big(\ 2I_2(I_1+I_3-1)-2\gamma-2,\quad 2c,\quad (x_{+++}+c)^{\gamma+1}\ \Big), \tag{24}$$

then $\boldsymbol{\delta}^{\phi,\gamma}$ dominates $\boldsymbol{\delta}^{ML}$ under the loss function (15).

(Proof)
By using (17) in Lemma 4.1, $\hat{\mathrm{Rd}}(\boldsymbol{\delta}^{\phi,\gamma})$ is

$$\begin{aligned}
\hat{\mathrm{Rd}}(\boldsymbol{\delta}^{\phi,\gamma}) &= 2\sum_{i_2}\sum_{i_1,i_3}\left\{\frac{\phi(x_{+++})}{B^{\gamma+1}}\frac{(x_{i_1i_2+}+1)(x_{+i_2i_3}+1)}{x_{+i_2+}+1}-\frac{\phi(x_{+++})}{(B-1)^{\gamma+1}}\frac{x_{i_1i_2+}x_{+i_2i_3}}{x_{+i_2+}}\right\}\\
&\quad -\sum_{i_2}\sum_{i_1,i_3}\frac{\phi^2(x_{+++})}{B^{2\gamma+2}}\frac{(x_{i_1i_2+}+1)(x_{+i_2i_3}+1)}{x_{+i_2+}+1}\\
&= 2\sum_{i_2}\left\{\frac{\phi(x_{+++}+1)}{B^{\gamma+1}}\cdot\frac{(x_{+i_2+}+I_1)(x_{+i_2+}+I_3)}{x_{+i_2+}+1}-\frac{\phi(x_{+++})x_{+i_2+}}{(B-1)^{\gamma+1}}\right\}\\
&\quad -\sum_{i_2}\frac{\phi^2(x_{+++}+1)}{B^{2\gamma+2}}\cdot\frac{(x_{+i_2+}+I_1)(x_{+i_2+}+I_3)}{x_{+i_2+}+1},
\end{aligned}$$

where where $B = x_{+++}+c+1$. Note that

$$\begin{aligned}
\sum_{i_2}\frac{(x_{+i_2+}+I_1)(x_{+i_2+}+I_3)}{x_{+i_2+}+1} &= x_{+++}+I_2(I_1+I_3-1)+\sum_{i_2}\frac{(I_1-1)(I_3-1)}{x_{+i_2+}+1}\\
&\ge x_{+++}+I_2(I_1+I_3-1)+\frac{I_2^2(I_1-1)(I_3-1)}{x_{+++}+I_2}\\
&= \frac{(x_{+++}+I_1I_2)(x_{+++}+I_2I_3)}{x_{+++}+I_2}.
\end{aligned}\tag{25}$$

Since $\phi(x_{+++}+1)/B^{\gamma+1} \le 1$ under the condition (24), $\hat{\mathrm{Rd}}(\delta^{\phi,\gamma})$ is bounded below by

$$\begin{aligned}
\tilde{\mathrm{Rd}}(\delta^{\phi,\gamma}) &= 2\left(\frac{\phi(x_{+++}+1)}{B^{\gamma+1}}\cdot\frac{(x_{+++}+I_1I_2)(x_{+++}+I_2I_3)}{x_{+++}+I_2}-\frac{\phi(x_{+++})}{(B-1)^{\gamma+1}}\cdot x_{+++}\right)\\
&\quad -\frac{\phi^2(x_{+++}+1)}{B^{2\gamma+2}}\cdot\frac{(x_{+++}+I_1I_2)(x_{+++}+I_2I_3)}{x_{+++}+I_2}.
\end{aligned}\tag{26}$$

(26) corresponds to (7) in the proof of Theorem 2.1. By using (25), $\tilde{\mathrm{Rd}}(\delta^{\phi,\gamma}) \ge 0$ under (24) can be proved in the same way as Theorem 2.1. ∎

5 Monte Carlo Study

We study the risk performance of the proposed estimators through Monte Carlo studies. Table 1 and Table 2 present the risks of the MLE and the proposed estimators for some two-way and three-way multiplicative models obtained from 100,000 replications. $\boldsymbol{\delta}^c$ in Table 1 and Table 2 are

$$\boldsymbol{\delta}^c = \boldsymbol{\delta}^{ML}\left(1 - \frac{c}{x_{++} + c}\right) \quad \text{and} \quad \boldsymbol{\delta}^c = \boldsymbol{\delta}^{ML}\left(1 - \frac{c}{x_{+++} + c}\right),$$

respectively. $\boldsymbol{\delta}^c$ is $\boldsymbol{\delta}^{\phi,\gamma}$ in Theorem 3.1 with $\phi = c$ and $\gamma = 0$. We set α_{ji_j} all equal to $1/I_j$.

Table 3 and 4 present the risks of the MLE and the proposed estimators for three-way decomposable models obtained from 100,000 replications. $\boldsymbol{\delta}^{c_2}$ and $\boldsymbol{\delta}^c$ in Table 3 and Table 4 are

$$\boldsymbol{\delta}^{c_{i_2}} = \boldsymbol{\delta}^{ML}\left(1 - \frac{c_{i_2}}{x_{+i_2+} + c_{i_2}}\right) \quad \text{and} \quad \boldsymbol{\delta}^c = \boldsymbol{\delta}^{ML}\left(1 - \frac{c}{x_{+++} + c}\right),$$

respectively. $\boldsymbol{\delta}^{c_{i_2}}$ is $\boldsymbol{\delta}^{\phi_{i_2},\gamma}$ in Theorem 4.1 with $\phi_{i_2} = c_2$ and $\gamma = 0$. $\boldsymbol{\delta}^c$ is $\boldsymbol{\delta}^{\phi,\gamma}$ in Theorem 4.3 with $\phi = c$ and $\gamma = 0$. $\theta_{i_1 i_2}$ and $\eta_{i_2 i_3}$ are all set to $1/(I_1 I_2)$ and $1/(I_2 I_3)$, respectively. In Table 3 all of λ_{i_2} are set to λ/I_2 and Table 4 is the case that not all of λ_{i_2} are equal.

The summary of these experiments is as follows.

- We can confirm the dominance of the proposed estimators against the MLE. As can be expected from the fact that the proposed estimators shrink the MLE towards zero, we see considerable amount of risk reduction when λ is small.
- The improvement is in the inverse proportion to λ.
- We can see from Table 1 and 2 that $\boldsymbol{\delta}^\nu$ shows larger risk reduction than $\boldsymbol{\delta}^c$ when λ is small.
- We can see from Table 3 and 4 that $\boldsymbol{\delta}^c$ shows larger risk reduction when λ_{i_2} are close together. On the contrary $\boldsymbol{\delta}^{c_{i_2}}$ and $\boldsymbol{\delta}^{\nu_{i_2}}$ seem to be better than $\boldsymbol{\delta}^c$ when λ_{i_2} varies widely and λ is large.

6 Some concluding remarks

In this paper we proposed improved estimators in some log linear models for Poisson random variables. We expect that improvements by Chou-type estimators can be extended to the case of general decomposable graphical models for J-way layouts. Since the notation and the proofs for general decomposable models become complicated, we will present our results for general decomposable models in our subsequent paper.

We also think that similar results as in Theorem 2.2 hold for more general models. However our present proof of Theorem 2.2 for the two-way case is very complicated and at present it seems difficult to extend it to more general cases.

In this paper we have considered shrinkage toward the origin assuming that the model is true. From practical viewpoint it might be more attractive to consider the saturated model and establish improvements by shrinkage toward some log linear model. Partial results in this direction have been obtained by the first author but at present sufficient conditions for improvements by shrinkage toward log linear models are rather complicated.

Appendix : Proof of Theorem 2.2

In this section we denote $x = x_{++}$, $a = I_1 + I_2$, $b = I_1 \cdot I_2$. From Lemma 2.1 and (8) $\hat{\mathrm{Rd}}(\delta^\nu)$ can be expressed by

$$
\begin{aligned}
&\hat{\mathrm{Rd}}(\delta^\nu, x) \\
&\quad = \frac{2x^3(x+I_1+I_2-2)}{(x+I_1-1)(x+I_2-1)(x+\nu+I_1+I_2-2)} + \frac{(x+I_1)(x+I_2)}{x+1} - 2x \\
&\qquad - \frac{(x+1)^3(x+I_1+I_2-1)^2}{(x+I_1)(x+I_2)(x+\nu+I_1+I_2-1)^2} \\
&\quad = \frac{g(x)}{(x+I_1-1)(x+I_2-1)(x+I_1)(x+I_2)(x+\nu+I_1+I_2-2)(x+\nu+I_1+I_2-1)^2},
\end{aligned}
$$

where $g(x)$ is a polynomial of degree 7. We obtained the explicit form of $g(x)$ using Mathematica and after some factorization for making the proof easier to see, we have the following expression of $g(x)$:

$$
\begin{aligned}
g(x) &= \sum_{i=0}^{7} c_i(\nu) x^i, \\
c_0(\nu) &= \left(b^2(\nu+a-1) - (a-1)^2\right)(b-a+1)(\nu+a-2) \\
c_1(\nu) &= \left\{2ab(b-a) + b^2(a-4) + 2b(2a-1)\right\}\nu^3 \\
&\quad + \left\{2b(a(3a-10)+11)(b-a) + ab(3a-7) + 3b^2(b-1) + 8b\right\}\nu^2 \\
&\quad + \Big\{2ab((3a-2)(a-4)+18+2b)(b-a) + a^2b^2(3a-10) + 2b^3(a-4) \\
&\qquad + 3a^2(a-2) + b^2(9a-28) + 6b(7a-2) + 3a\Big\}\nu \\
&\quad + \Big\{(2a^3b(a-4) + a^2b^2)(b-a) + a^2b^2(a-5)^2 + 2ab^3(a-4) + a^3(3a-11) \\
&\qquad + a^2b(4a-9)(a-4) + ab^2(10a-39) + 3ab(3a-10) + 3a(4a-1) + 13b^2 + 7b - 1\Big\}, \\
c_2(\nu) &= \left\{a^2(b-a) + 2ab(a-4) + a(3a-2) + b(b+4)\right\}\nu^3 \\
&\quad + \Big\{a^2\left((3a-1)(a-4)+14\right)(b-a) + 4b^2(3a-4) + ab(6a-5)(a-4) \\
&\qquad + 2b(9a-11) + 8a\Big\}\nu^2 \\
&\quad + \Big\{\left(a^2(3a-5)(a-4) + 9a^2b + 13a^2 + 3b^2\right)(b-a) + 2a^2b(3a-5)(a-4) \\
&\qquad + ab(85a-112) + 4b^2(3a-2)(a-4) + a(34a-19) + 8b^2 + 37b\Big\}\nu \\
&\quad + \Big\{\left(a^3(a-4)(a-3) + 3a^3b + 2a^2b + 5\right)(b-a) + 2a^3b(a-4)^2 + a^2b^2(7a-13)(a-4) \\
&\qquad + a^2b(41a-98) + b^3(a-4) + a^2(27a-37) + b^2(7a-26) + b(71a-22) \\
&\qquad + 29a - 8\Big\},
\end{aligned}
$$

$$
\begin{aligned}
c_3(\nu) &= \Big\{a^2(a-4)+2b(a-2)+6a-2\Big\}\nu^3 \\
&\quad +\Big\{(5a^2+3b)(b-a)+a^2(3a-2)(a-4)+ab(10a-39)+a(35a-36)+8\Big\}\nu^2 \\
&\quad +\Big\{10ab(b-a)+a^3(3a-8)(a-4)+6ab(3a-4)(a-4)+a^2(47a-98) \\
&\qquad +5b^2(a-4)+2b(15a-23)+53a-10\Big\}\nu \\
&\quad +\Big\{(a^4+5a^2b+13a^2+b^2+6b)(b-a)+a^4(a-5)^2+a^3(17a-65) \\
&\qquad +a^2b(10a-23)(a-4)+ab(22a-86)+b^2(7a-4)(a-4)+a(47a-23)+8\Big\}, \\
c_4(\nu) &= (a^2-b-3a+3)\nu^3+\Big\{a^2(6a-23)+5b(a-2)+35a-18\Big\}\nu^2 \\
&\quad +\Big\{3a^2(3a-5)(a-4)+ab(21a-66)+2a(23a-44)+3b^2+36b+25\Big\}\nu \\
&\quad +\Big\{4ab(b-a)+4a^3(a-3)(a-4)+2ab^2(a-4)+3a^2(13a-32) \\
&\qquad +4(3a-5)(a-4)+7b(a-4)+38a-3\}, \\
c_5(\nu) &= -a\nu^3+(2a^3-5a-b+5)\nu^2+\Big\{9a^2(a-4)+2b(3a-7)+48a-22\Big\}\nu \\
&\quad +\Big\{a^2(6a-13)(a-4)+ab(9a-34)+2a(14a-31)+(b^2+22b+13)\Big\}, \\
c_6(\nu) &= -\nu^3+(2-2a)\nu^2+(a^2+2a(a-4)+4)\nu+2b(a-3)+4(a-4)(a-1)+4a-2, \\
c_7(\nu) &= -\nu^2+(a-2)^2.
\end{aligned}
$$

It suffices to show that $c_i(\nu)\geq 0$, $i=1,\ldots,7$, for $\nu\leq a-2$ and $I_1\geq 2$ and $I_2\geq 2$.

$c_0(\nu)$ and $c_7(\nu)$ obviously satisfy this condition. $c_i(\nu)$, $i=1,\ldots,6$, are polynomials of degree 3. Denote $c_i(\nu)=\sum_{j=0}^{3}\tilde{c}_{ij}\nu^j$. Note that a and b satisfies $a\geq 4$, $b\geq 4$ and $b-a\geq 0$. By using this fact we can confirm that

1. $c_i(\nu)$, $i=1,\ldots,4$, satisfy that $\tilde{c}_{ij}\geq 0$, $j=0,\ldots,3$, for integers a, $b\geq 4$.

2. $c_5(\nu)$ and $c_6(\nu)$ satisfy that $\tilde{c}_{i3}\leq 0$, $\tilde{c}_{i0}\geq 0$, $c_i'(0)\geq 0$ and $c_i(a-2)\geq 0$ for integers a, $b\geq 4$. ∎

Table 1. Risks of the MLE and the proposed estimators for two-way multiplicative models.

(i) 2×2 contingency tables

λ	δ^{ML}	δ^{c}			δ^{ν}		
		$c=4$	$c=2$	$c=1$	$\nu=2$	$\nu=1$	$\nu=0$
0.1	3.956	0.254	0.546	1.101	0.190	0.245	0.363
0.5	3.794	0.586	0.892	1.431	0.545	0.612	0.752
1.0	3.640	0.932	1.234	1.740	0.913	0.985	1.135
2.0	3.432	1.452	1.708	2.135	1.462	1.522	1.664
5.0	3.196	2.261	2.355	2.601	2.292	2.290	2.360
10.0	3.102	2.706	2.675	2.801	2.728	2.673	2.684
20.0	3.044	2.914	2.837	2.894	2.923	2.856	2.840
50.0	3.023	2.997	2.942	2.963	2.999	2.956	2.942
100.0	2.996	2.989	2.955	2.966	2.990	2.964	2.956

(ii) 3×3 contingency tables

λ	δ^{ML}	δ^{c}			δ^{ν}		
		$c=8$	$c=4$	$c=2$	$\nu=4$	$\nu=2$	$\nu=0$
0.1	8.824	0.210	0.475	1.136	0.137	0.164	0.229
0.5	8.169	0.575	0.881	1.599	0.518	0.558	0.655
1.0	7.525	0.979	1.308	2.058	0.944	0.991	1.114
2.0	6.732	1.653	1.981	2.731	1.656	1.702	1.845
5.0	5.785	2.942	3.121	3.710	3.004	2.985	3.088
10.0	5.388	3.933	3.895	4.271	4.000	3.884	3.901
20.0	5.214	4.614	4.428	4.633	4.653	4.477	4.436
50.0	5.085	4.944	4.764	4.847	4.955	4.811	4.766
100.0	5.038	4.997	4.877	4.918	5.000	4.908	4.878

(iii) 5×5 contingency tables

λ	δ^{ML}	δ^{c}			δ^{ν}		
		$c=16$	$c=8$	$c=4$	$\nu=8$	$\nu=4$	$\nu=0$
0.1	24.234	0.188	0.428	1.161	0.112	0.123	0.150
0.5	21.619	0.577	0.873	1.743	0.506	0.524	0.575
1.0	19.098	1.029	1.373	2.364	0.972	0.996	1.069
2.0	15.934	1.841	2.236	3.379	1.820	1.846	1.951
5.0	12.173	3.685	4.017	5.210	3.770	3.743	3.855
10.0	10.609	5.550	5.610	6.594	5.700	5.532	5.562
20.0	9.802	7.278	6.970	7.618	7.410	7.062	6.972
50.0	9.316	8.590	8.095	8.394	8.644	8.240	8.100
100.0	9.118	8.889	8.494	8.649	8.908	8.602	8.496

Table 2. Risks of the MLE and the proposed estimators for three-way multiplicative models.

(i) $2 \times 2 \times 2$ contingency tables

λ	δ^{ML}	δ^c $c = 6$	$c = 3$	$c = 1.5$
0.1	7.830	0.262	0.612	1.402
0.5	7.056	0.614	0.987	1.772
1.0	6.386	0.998	1.380	2.145
2.0	5.549	1.608	1.959	2.647
5.0	4.640	2.681	2.851	3.306
10.0	4.305	3.402	3.377	3.632
20.0	4.158	3.819	3.693	3.820
50.0	4.059	3.987	3.876	3.924
100.0	4.028	4.008	3.937	3.960

(ii) $3 \times 3 \times 3$ contingency tables

λ	δ^{ML}	δ^c $c = 12$	$c = 6$	$c = 3$
0.1	25.726	0.261	0.668	1.826
0.5	21.948	0.643	1.094	2.309
1.0	18.498	1.081	1.568	2.826
2.0	14.227	1.843	2.350	3.618
5.0	9.795	3.461	3.826	4.897
10.0	8.275	4.925	4.989	5.737
20.0	7.592	6.095	5.858	6.289
50.0	7.238	6.848	6.522	6.705
100.0	7.113	6.991	6.751	6.843

(iii) $5 \times 5 \times 5$ contingency tables

λ	δ^{ML}	δ^c $c = 24$	$c = 12$	$c = 6$
0.1	117.471	0.303	0.858	2.698
0.5	94.579	0.703	1.309	3.241
1.0	74.247	1.183	1.848	3.895
2.0	49.779	2.067	2.798	4.971
5.0	25.797	4.238	4.943	7.125
10.0	18.496	6.723	7.129	9.007
20.0	15.544	9.376	9.221	10.565
50.0	13.967	11.936	11.216	11.880
100.0	13.499	12.787	12.084	12.435

Table 3. Risks of the MLE and the proposed estimators for the decomposable models for three-way contingency tables when all of λ_{i_2} are equal to λ/I_2

(i) $2 \times 2 \times 2$ contingency tables

		$\delta^{c_{i_2}}$		δ^{c}		δ^{ν}	
λ	δ^{ML}	$c_{i_2} = 4$	$c_{i_2} = 2$	$c = 10$	$c = 5$	$\nu = 2$	$\nu = 1$
0.1	7.981	0.416	0.996	0.165	0.337	0.282	0.389
0.5	7.772	0.768	1.366	0.535	0.752	0.659	0.780
1.0	7.575	1.171	1.781	0.954	1.206	1.089	1.223
2.0	7.265	1.863	2.463	1.670	1.941	1.825	1.967
5.0	6.732	3.299	3.754	3.131	3.291	3.334	3.435
10.0	6.397	4.525	4.716	4.373	4.308	4.587	4.584
20.0	6.209	5.419	5.358	5.307	5.056	5.463	5.354
50.0	6.070	5.893	5.740	5.850	5.586	5.906	5.778
100.0	6.040	5.990	5.877	5.975	5.793	5.994	5.906

(ii) $3 \times 3 \times 3$ contingency tables

		$\delta^{c_{i_2}}$		δ^{c}		δ^{ν}	
λ	δ^{ML}	$c_{i_2} = 8$	$c_{i_2} = 4$	$c = 28$	$c = 14$	$\nu = 4$	$\nu = 2$
0.1	26.787	0.433	1.196	0.134	0.237	0.206	0.278
0.5	26.003	0.817	1.628	0.529	0.681	0.605	0.692
1.0	25.161	1.277	2.142	1.003	1.207	1.086	1.190
2.0	23.773	2.146	3.091	1.892	2.176	1.995	2.125
5.0	20.835	4.339	5.343	4.131	4.468	4.312	4.458
10.0	18.459	6.996	7.815	6.810	6.949	7.113	7.180
20.0	16.790	10.136	10.418	9.935	9.567	10.344	10.175
50.0	15.722	13.399	12.918	13.228	12.286	13.539	13.037
100.0	15.371	14.540	13.933	14.445	13.530	14.600	14.095

(iii) $5 \times 5 \times 5$ contingency tables

		$\delta^{c_{i_2}}$		δ^{c}		δ^{ν}	
λ	δ^{ML}	$c_{i_2} = 16$	$c_{i_2} = 8$	$c = 88$	$c = 44$	$\nu = 8$	$\nu = 4$
0.1	123.509	0.533	1.657	0.118	0.173	0.158	0.200
0.5	119.929	0.932	2.109	0.522	0.615	0.560	0.610
1.0	116.596	1.432	2.695	1.023	1.165	1.061	1.123
2.0	110.322	2.404	3.805	2.006	2.250	2.044	2.125
5.0	95.103	5.146	6.849	4.800	5.298	4.860	4.980
10.0	79.457	9.208	11.159	8.938	9.651	9.109	9.236
20.0	64.717	15.744	17.615	15.529	16.056	16.022	16.007
50.0	52.984	27.695	27.992	27.428	26.285	28.445	27.604
100.0	49.033	36.386	34.864	36.059	33.320	37.049	35.316

Table 4. Risks of the MLE and the proposed estimators for the decomposable models for three-way contingency tables when not all of λ_{i_2} are equal.
(iv) $2 \times 2 \times 2$ contingency tables

		$\delta^{c_{i_2}}$		δ^{c}		δ^{ν}	
$(\lambda_1,\ \lambda_2)$	δ^{ML}	$c_{i_2}=4$	$c_{i_2}=2$	$c=10$	$c=5$	$\nu=2$	$\nu=1$
(0.1,0.9)	7.580	1.116	1.705	0.955	1.209	1.030	1.155
(0.2,0.8)	7.585	1.141	1.741	0.955	1.209	1.057	1.186
(0.3,0.7)	7.572	1.157	1.762	0.954	1.207	1.074	1.206
(0.4,0.6)	7.570	1.166	1.774	0.954	1.206	1.084	1.217
(1.0,9.0)	6.736	3.586	3.867	4.462	4.463	3.591	3.612
(2.0,8.0)	6.548	4.044	4.299	4.413	4.377	4.080	4.098
(3.0,7.0)	6.461	4.323	4.547	4.384	4.333	4.375	4.383
(4.0,6.0)	6.417	4.477	4.679	4.374	4.314	4.537	4.538

(v) $3 \times 3 \times 3$ contingency tables

		$\delta^{c_{i_2}}$		δ^{c}		δ^{ν}	
$(\lambda_1,\ \lambda_2,\lambda_3)$	δ^{ML}	$c_{i_2}=8$	$c_{i_2}=4$	$c=28$	$c=14$	$\nu=4$	$\nu=2$
(0.1,0.1,0.8)	25.427	1.247	2.101	1.007	1.219	1.054	1.155
(0.1,0.2,0.7)	25.339	1.258	2.118	1.005	1.215	1.066	1.168
(0.1,0.3,0,6)	25.314	1.266	2.128	1.004	1.212	1.074	1.177
(0.2,0.3,0.5)	25.250	1.272	2.135	1.004	1.210	1.081	1.184
(0.3,0.3,0.4)	25.177	1.276	2.139	1.003	1.207	1.085	1.189
(1.0,1.0,8.0)	20.559	5.601	6.291	6.982	7.357	5.600	5.611
(1.0,2.0,7.0)	19.827	6.076	6.809	6.919	7.211	6.113	6.141
(1.0,3.0,6.0)	19.456	6.380	7.120	6.893	7.144	6.444	6.478
(2.0,3.0,5.0)	18.791	6.780	7.578	6.839	7.014	6.877	6.934
(3.0,3.0,4.0)	18.492	6.966	7.779	6.814	6.955	7.080	7.145

References

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

[2] Clevenson, M. L. and Zidek, J. V. (1975). Simultaneous Estimation of the Means of Independent Poisson Laws, *J. Amer. Statist. Assoc.*, **70**, 698-705.

[3] 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.

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

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

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

[7] 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.

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

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

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

[11] 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.