SIMULTANEOUS ESTIMATION OF THE MEANS IN SOME POISSON LOG LINEAR MODELS

Hisayuki Hara* and Akimichi Takemura**

In this article we study the simultaneous estimation of the Poisson 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 (1975), Tsui and Press (1982) and Chou (1991).

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

1. Introduction

Consider a two-way contingency table $x_{i_1i_2}$, $i_1 = 1, \ldots, I_1$, $i_2 = 1, \ldots, I_2$, 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 $\text{Mn}(x_{++}, p_{11}, \ldots, p_{I_1I_2})$ 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

$$\lambda_{i_1i_2} = \lambda \alpha_{i_1i_2},$$

$$\sum_{i=1}^{I_j} \alpha_{i_1i_2} = 1 \quad \text{for} \quad j = 1, 2.$$ 

The model (1.1) is generally known as the multiplicative Poisson model. In this article we first address the simultaneous estimation of $\lambda = (\lambda_{11}, \ldots, \lambda_{I_1I_2})'$ 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 (1975) correspond to the estimation of $\lambda$ in the saturated model in which there are no constraints on $\lambda_{i_1i_2}$. Clevenson and Zidek (1975) proved the inadmissibility of the maximum likelihood estimator


*Department of Geosystem Engineering, Graduate School of Technology, University of Tokyo, Bunkyo-ku, Tokyo 113-8656, Japan. Email: hara@geosys.t.u-tokyo.ac.jp

**Department of Mathematical Informatics, Graduate School of Information Science and Technology, University of Tokyo, Bunkyo-ku, Tokyo 113-8656, Japan. Email: takemura@stat.t.u-tokyo.ac.jp
(MLE) when \( I_1 \cdot I_2 \geq 2 \) and provided a class of estimators which dominate the MLE under the normalized squared loss

\[
L(\lambda, \hat{\lambda}) = \sum_{i_1, i_2} \frac{1}{\lambda_{i_1 i_2}} (\hat{\lambda}_{i_1 i_2} - \lambda_{i_1 i_2})^2.
\]

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

Tsui and Press (1982) enlarged the class of Clevenson and Zidek and also derived improved estimators under \( k \)-normalized squared error loss, \( L_k(\lambda, \hat{\lambda}) = \sum_{i_1, i_2} (\hat{\lambda}_{i_1 i_2} - \lambda_{i_1 i_2})^2 / \lambda_{i_1 i_2}^k \). Hwang (1982) extended their results to the estimation of the mean parameters of a subclass of discrete exponential families. He generalized the identity of Hudson (1978) and derived an unbiased estimator of the difference of two risk functions. Chou (1991) 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 (2004) 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 (1981), Ghosh et al. (1983), Ghosh and Yang (1988), Johnstone (1986), Tsui (1984).

We consider to apply these arguments to the estimation of \( \lambda \) for the multiplicative Poisson model (1.1). In Section 2 we provide two classes of estimators which dominate the MLE under the loss (1.2). The one corresponds exactly to the class of Chou (1991). 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.

These results suggest that we may apply the argument in Sections 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,

\[
x_{i_1 i_2 i_3} \sim \text{Po}(\lambda_{i_1 i_2 i_3}), \quad \lambda_{i_1 i_2 i_3} = \lambda \frac{\alpha_{i_1 i_2} \beta_{i_2 i_3}}{\gamma_{i_2}}, \quad i_j = 1, \ldots, I_j, \quad I_j \geq 2, \quad j = 1, 2, 3,
\]

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

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 \( \lambda = (\lambda_{11}, \ldots, \lambda_{I_1 I_2})' \) in two-way multiplicative model (1.1) under the loss function (1.2). In this model \( \mathbf{x}_1 = (x_{1+}, \ldots, x_{I_1+})' \) and \( \mathbf{x}_2 = (x_{+1}, \ldots, x_{+I_2})' \) are the complete
sufficient statistics. The joint probability function of \( x_1 \) and \( x_2 \) is easily obtained by

\[
\Pr(x_1, x_2) = \exp(-\lambda)\lambda^{x_+} \prod_{i_1} \alpha_{1i_1}^{x_{i_1}+} \prod_{i_2} \alpha_{2i_2}^{x_{i_2}+} \cdot t(x_1, x_2),
\]

where

\[
t(x_1, x_2) = \frac{x_+!}{\prod_{i_1} x_{i_1}! \prod_{i_2} x_{i_2}!}.
\]

From (2.1) the MLE of \( \lambda \) is

\[
\delta^{ML} = \left( \delta^{ML}_{11}, \ldots, \delta^{ML}_{1i_2} \right)', \quad \delta^{ML}_{i_1i_2} = \begin{cases} \frac{x_{i_1} + x_{i_2}}{x_+} & \text{if } x_+ \neq 0, \\ 0 & \text{if } x_+ = 0. \end{cases}
\]

The following lemma corresponds to the identity of Hudson (1978) and Hwang (1982).

**Lemma 2.1.** Let \( x_1 \) and \( x_2 \) have the probability function (2.1). If \( g \) is a real valued function with \( \mathbb{E}[g(x_1, x_2)] < \infty \) and \( g(x_1, x_2) = 0 \) whenever \( x_{i_1} < m \) or \( x_{i_2} < m \), then

\[
\mathbb{E} \left[ \frac{1}{\lambda^m} \cdot g(x_1, x_2) \right] = \mathbb{E} \left[ \frac{t(x_1 + me_{i_1}, x_2 + me_{i_2})}{t(x_1, x_2)} \cdot g(x_1 + me_{i_1}, x_2 + me_{i_2}) \right],
\]

where \( e_{ij} \) 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 (1978) and Hwang (1982) and is omitted here. From this lemma with \( m = -1 \) and \( g(x_1, x_2) \equiv 1 \), \( \delta^{ML} \) is found to be the uniformly minimum variance unbiased estimator (UMVUE) of \( \lambda \).

We address the estimators which dominate \( \delta^{ML} \) under the loss function (1.2) from the following class

\[
\delta^\psi = \delta^{ML} - \Psi(x_1, x_2),
\]

where \( \Psi(x_1, x_2) = (\Psi_{11}(x_1, x_2), \ldots, \Psi_{1i_2}(x_1, x_2))' \). By using (2.2) in Lemma 2.1 with \( m = 1 \), the difference between two risk functions of \( \delta^{ML} \) and \( \delta^\psi \) is expressed as

\[
\mathcal{R}(\lambda, \delta^{ML}) - \mathcal{R}(\lambda, \delta^\psi) \equiv \mathbb{E}[L(\lambda, \delta^{ML}) - L(\lambda, \delta^\psi)]
\]

\[
= \sum_{i_1, i_2} \mathbb{E} \left[ \frac{1}{\lambda_{i_1i_2}} \Psi_{i_1i_2}(x_1, x_2)(2\delta_{i_1i_2}^{ML} - \Psi_{i_1i_2}(x_1, x_2)) - 2\Psi_{i_1i_2}(x_1, x_2) \right]
\]

\[
= \sum_{i_1, i_2} \mathbb{E}[2(\Psi_{i_1i_2}(x_1 + e_{i_1}, x_2 + e_{i_2}) - \Psi_{i_1i_2}(x_1, x_2))]
\]

\[
- \sum_{i_1, i_2} \mathbb{E} \left[ \frac{x_++1}{(x_{i_1}+1)(x_{i_2}+1)} \Psi_{i_1i_2}^2(x_1 + e_{i_1}, x_2 + e_{i_2}) \right].
\]
The last term implies that
\[
(2.3) \quad \hat{R}_d(\delta^\psi) = 2 \sum_{i_1, i_2} (\Psi_{i_1 i_2}(x_1 + e_{i_1}, x_2 + e_{i_2}) - \Psi_{i_1 i_2}(x_1, x_2))
\]
\[
- \sum_{i_1, i_2} \left( \frac{x_{++} + 1}{(x_{i_1 i_2} + 1)(x_{i_2 i_2} + 1)} - \frac{x_{i_1 i_2}^2}{(x_{i_1 i_2} + 1)^2} \right) \Psi_{i_1 i_2}^2(x_1 + e_{i_1}, x_2 + e_{i_2})
\]
is the UMVUE of \( R(\lambda, \delta^{ML}) - R(\lambda, \delta^\psi) \). In what follows we derive a sufficient condition on \( \Psi \) to satisfy \( \hat{R}_d(\delta^\psi) \geq 0 \) for two classes of \( \Psi \).

We begin with the following class of Chou (1991)-type estimators,
\[
\delta^{\phi, \gamma} = (\delta^{\phi, \gamma}_{11}, \ldots, \delta^{\phi, \gamma}_{I_1 I_2})',
\]
\[
(2.4) \quad \delta^{\phi, \gamma}_{i_1 i_2} = \delta^{ML}_{i_1 i_2} \left( 1 - \frac{\phi(x_{++})}{(x_{++} + c)^{\gamma + 1}} \right),
\]
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 \( \phi \) to improve on \( \delta^{ML} \) is presented.

**Theorem 2.1.** Suppose that \( I_1 \geq 2 \) and \( I_2 \geq 2 \). If \( \phi(\cdot) \) is nondecreasing and satisfies
\[
0 \leq \phi(x) \leq \min(2(I_1 + I_2) - 2\gamma - 4, \ 2c - 2\gamma, \ (x + c)^{\gamma + 1})
\]
for all \( x \geq 0 \), then \( \delta^{\phi, \gamma} \) dominates \( \delta^{ML} \) under the loss function (1.2).

**Proof.** We can prove this theorem by using the procedure in the proof of Theorem 2.1 in Chou (1991). From (2.3) \( \hat{R}_d(\delta^{\phi, \gamma}) \) is
\[
(2.5) \quad \hat{R}_d(\delta^{\phi, \gamma}) = 2 \sum_{i_1, i_2} \left\{ \frac{\phi(x_{++} + 1)}{B^{\gamma + 1}} \cdot \frac{(x_{i_1 i_2} + 1)(x_{i_2 i_2} + 1)}{x_{++} + 1} \right. \]
\[
- \frac{\phi(x_{++})}{(B - 1)^{\gamma + 1}} \cdot \frac{x_{i_1 i_2} x_{i_2 i_2}}{x_{++}} \}
\]
\[
- \sum_{i_1, i_2} \frac{\phi^2(x_{++} + 1)}{B^{2\gamma + 2}} \cdot \frac{(x_{i_1 i_2} + 1)(x_{i_2 i_2} + 1)}{x_{++} + 1}
\]
\[
= 2 \left\{ \frac{\phi(x_{++} + 1)}{B^{\gamma + 1}} \cdot \frac{(x_{++} + I_1)(x_{++} + I_2)}{x_{++} + 1} - \frac{\phi(x_{++})x_{++}}{(B - 1)^{\gamma + 1}} \right. \}
\]
\[
- \frac{\phi^2(x_{++} + 1)}{B^{2\gamma + 2}} \cdot \frac{(x_{++} + I_1)(x_{++} + I_2)}{x_{++} + 1}
\]
where \( B = x_{++} + c + 1 \). Since \( \phi(\cdot) \) is nondecreasing and \( B > 1 \),
\[
2 \left\{ \frac{\phi(x_{++} + 1)}{B^{\gamma + 1}} \cdot \frac{(x_{++} + I_1)(x_{++} + I_2)}{x_{++} + 1} - \frac{\phi(x_{++})x_{++}}{(B - 1)^{\gamma + 1}} \right. \}
\]
\[
- \frac{\phi^2(x_{++} + 1)}{B^{2\gamma + 2}} \cdot \frac{(x_{++} + I_1)(x_{++} + I_2)}{x_{++} + 1}
\]
order to prove this theorem we present the following lemma. 

The second inequality follows from the fact that \((1975)\) is a natural extension of the Bayes estimator discussed in Clevenson and Zidek (2.7). 

Next we consider the following class, 

\[
\delta^\nu = (\delta_{i1}^\nu, \ldots, \delta_{i1I_2}^\nu),
\]

\[
\delta_{i1i2}^\nu = \frac{x_{i1} + x_{i2}}{(x_{i1} + I_1 - 1)(x_{i2} + I_2 - 1)} \cdot \frac{x_{i1} + I_1 + I_2 - 2}{x_{i1} + \nu + I_1 + I_2 - 2} \cdot x_{i1},
\]

where \(\nu \geq 0\) is a constant. \(\delta^\nu\) is the Bayes estimator with respect to the prior measure \(\pi(\lambda, \alpha_1, \alpha_2 | \nu)\), 

\[
(2.7) \quad \pi(\lambda, \alpha_1, \alpha_2 | \nu) d\lambda = \prod_{i_1=1}^{I_1-1} d\alpha_{i1} \prod_{i_2=1}^{I_2-1} d\alpha_{2i2} = m(\lambda) d\lambda \prod_{i_1=1}^{I_1-1} d\alpha_{i1} \prod_{i_2=1}^{I_2-1} d\alpha_{2i2},
\]

\[
m(\lambda) = \int_0^\infty (1 + \lambda t)^{-\nu} t^{-(I_1+I_2-1)} \exp(-t^{-1}) dt,
\]

\[
\alpha_1 = (\alpha_{11}, \ldots, \alpha_{1I_1}), \quad \alpha_2 = (\alpha_{21}, \ldots, \alpha_{2I_2}).
\]

For \(\nu > 1\), \(\pi(\lambda, \alpha_1, \alpha_2, \nu)\) is proper and therefore \(\delta^\nu\) is admissible. This estimator is a natural extension of the Bayes estimator discussed in Clevenson and Zidek (1975).

The following theorem gives the condition on \(\nu\) for \(\delta^\nu\) to dominate \(\delta^{ML}\).

**Theorem 2.2.** Suppose that \(I_1 \geq 2\) and \(I_2 \geq 2\). If \(0 \leq \nu \leq I_1 + I_2 - 2\), \(\delta^\nu\) improves on \(\delta^{ML}\) under the loss function \((1.2)\). If \(\nu > 1\), \(\delta^\nu\) is admissible.

This result is consistent with the one of Clevenson and Zidek (1975). In order to prove this theorem we present the following lemma.
Lemma 2.2. Consider the following class of estimators
\[ \delta^h = \left( 1 - \frac{h(x_+)}{x_+} \right) \delta^{ML}. \]

If \( h(\cdot) \) is nonnegative, nondecreasing and satisfies
\[
\frac{2(x + 1)^2}{(x + I_1)(x + I_2)} - \left( 1 - \frac{h(x + 1)}{x + 1} \right) \leq 1 \tag{2.8}
\]
for all \( x \geq 0 \), then \( \delta^h \) dominates \( \delta^{ML} \) under the loss (1.2).

Proof. From (2.3) \( \hat{R}_d(\delta^h) \) satisfies
\[
\hat{R}_d(\delta^h) = 2 \left( \frac{h(x_+ + 1)(x_+ + I_1)(x_+ + I_2)}{(x_+ + 1)^2} - h(x_+) \right) - \frac{h^2(x_+ + 1)(x_+ + I_1)(x_+ + I_2)}{(x_+ + 1)^3}
\geq \frac{h(x_+ + 1)(x_+ + I_1)(x_+ + I_2)}{(x_+ + 1)^2}
\cdot \left\{ 2 \left( 1 - \frac{(x_+ + 1)^2}{(x_+ + I_1)(x_+ + I_2)} \right) - \frac{h(x_+ + 1)}{x_+ + 1} \right\}. 
\]
The inequality follows from the assumption that \( h(\cdot) \) is nondecreasing. The right hand side is always nonnegative when \( h(\cdot) \) satisfies the condition (2.8). 

Lemma 2.3. Define \( h^{\nu,\kappa}(\cdot) \) by
\[ h^{\nu,\kappa}(x) = x \left( 1 - \frac{x^2(x + \kappa + I_1 + I_2 - 2)}{(x + I_1 - 1)(x + I_2 - 1)(x + \nu + \kappa + I_1 + I_2 - 2)} \right). \]

Then \( h^{\nu,\kappa}(x) \) is nondecreasing on \( x \geq 0 \) when \( \nu \geq 0 \) and \( \kappa \geq 0 \).

Proof. \( h^{\nu,\kappa}(x) \) is rewritten by
\[
h^{\nu,\kappa}(x) = \frac{(I_1 - 1)x}{x + I_1 - 1} + \frac{(I_2 - 1 + \nu)x^2}{(x + I_1 - 1)(x + I_2 - 1)(x + \nu + \kappa + I_1 + I_2 - 2)} \left( x + \frac{I_2 - 1}{I_2 - 1 + \nu} (\nu + \kappa + I_1 + I_2 - 2) \right),
\]
which is the sum of nondecreasing functions for \( x \geq 0 \) when \( \nu \geq 0 \) and \( \kappa \geq 0 \). Thus the proof is completed. 

We also use the following inequality. For \( y > 0, a_k > 0, k = 1, 2, \ldots, K \geq 2, \)
\[
\prod_{k=1}^{K} (y + a_k) = y^K + y^{K-1} \sum_{k=1}^{K} a_k + \cdots + \prod_{k=1}^{K} a_k > y^K + y^{K-1} \sum_{k=1}^{K} a_k,
\]

hence
\[
\frac{y^{K-1}(y + \sum_{k=1}^{K} a_k)}{\prod_{k=1}^{K} (y + a_k)} < 1. \tag{2.9}
\]
**Proof of Theorem 2.2.** Let \( h(\cdot) \) for the class (2.6) be denoted by \( h^\nu(\cdot) \). Then \( h^\nu(\cdot) \) is expressed by

\[
h^\nu(x) = x \left( 1 - \frac{x^2(x + I_1 + I_2 - 2)}{(x + I_1 - 1)(x + I_2 - 1)(x + \nu + I_1 + I_2 - 2)} \right).
\]

From Lemma 2.3 with \( \kappa = 0 \), \( h^\nu(x) \) is nondecreasing for \( x \geq 0 \) and \( \nu \geq 0 \).

Next we have for \( 0 \leq \nu \leq I_1 + I_2 - 2 \)

\[
(2.10) \quad \frac{2(x + 1)^2}{(x + I_1)(x + I_2)} - \left( 1 - \frac{h^\nu(x + 1)}{x + 1} \right)
\]

\[
= \frac{2(x + 1)^2}{(x + I_1)(x + I_2)} - \frac{(x + I_1 + I_2 - 1)(x + 1)^2}{(x + I_1)(x + I_2)(x + \nu + I_1 + I_2 - 1)}
\]

\[
= \frac{(x + I_1)(x + I_2)(x + 2\nu + I_1 + I_2 - 1)}{(x + I_1)(x + I_2)(x + \nu + I_1 + I_2 - 1)}
\]

\[
\leq \frac{(x + 1)^2(x + 2I_1 + 2I_2 - 5)}{(x + I_1)(x + I_2)(x + 2I_1 + 2I_2 - 3)} \leq 1.
\]

The first inequality follows from the fact that the left hand side is nondecreasing on \( \nu \). The second inequality follows from (2.9) with \( K = 3, y = x + 1, a_1 = I_1 - 1, a_2 = I_2 - 1 \) and \( a_3 = 2I_1 + 2I_2 - 4 \). Thus \( h^\nu(\cdot) \) satisfies (2.8). From Lemma 2.2 the proof is completed. \( \Box \)

We conjecture that the class in Theorem 2.2 is included in the class of Theorem 2.1 with \( \gamma = 0 \) and \( c = (1/2)(\nu + I_1 + I_2 - 2) \). However its proof seems to be difficult.

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

We can extend the results of Theorems 2.1 and 2.2 to multi-way multiplicative models. \( J \)-way multiplicative model is expressed as

\[
(3.1) \quad x_{i_1i_2\ldots i_J} \sim \text{Po}(\lambda_{i_1i_2\ldots i_J}), \quad \lambda_{i_1i_2\ldots i_J} = \lambda \alpha_{i_1j_1} \alpha_{i_2j_2} \ldots \alpha_{i_Jj_J},
\]

\[
\sum_{i_j} \alpha_{i_j} = 1, \quad j = 1, \ldots, J,
\]

where \( x_{i_1i_2\ldots i_J} \) are assumed to be independently distributed. The problem is to estimate \( \lambda = \{ \lambda_{i_1i_2\ldots i_J} \} \) simultaneously under the loss function

\[
(3.2) \quad L(\lambda, \hat{\lambda}) = \sum_{j=1}^J \sum_{i_j=1}^{I_j} \frac{1}{\lambda_{i_1i_2\ldots i_J}} (\hat{\lambda}_{i_1i_2\ldots i_J} - \lambda_{i_1i_2\ldots i_J})^2.
\]

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

\[
x_{j,i_j}^+ = \sum_{i_s : s \neq j} x_{i_1i_2\ldots i_J}, \quad x^+ = \sum_{i_j} x_{j,i_j}^+.
\]
Then the set of one-dimensional marginal frequencies $x_j^+ = (x_{j,1}^+, \ldots, x_{j,I_j}^+)'$, $j = 1, \ldots, J$, is the complete sufficient statistic. Its joint probability function is given by

$$
Pr(x_1^+, \ldots, x_J^+) = \exp(-\lambda)\lambda^{x^+} \prod_j \prod_{ij} x_{j,ij}^{+} \cdot t(x_1^+, \ldots, x_J^+),
$$

where

$$
t(x_1^+, \ldots, x_J^+) = \frac{(x^+)^{J-1}}{\prod_j \prod_{ij} x_{j,ij}^+}.
$$

From (3.3) the MLE is

$$
\delta_{\text{ML}} = \left\{ \delta_{\text{ML}_{i_1i_2\ldots i_J}} \right\}, \delta_{\text{ML}_{i_1i_2\ldots i_J}} = \begin{cases} \frac{\prod_j x_{j,ij}^+}{(x^+)^{J-1}} & \text{if } x^+ \neq 0, \\ 0 & \text{if } x^+ = 0. \end{cases}
$$

The following lemma is the identity which corresponds to (2.2).

**Lemma 3.1.** Let $(x_1^+, \ldots, x_J^+)$ have the probability function (3.3). If $g$ is a real valued function with $E[g(x_1^+, \ldots, x_J^+)] < \infty$ and $g(x_1^+, \ldots, x_J^+) = 0$ whenever $x_{j,ij}^+ < m$ for some $(j, i_j)$, then

$$
E \left[ \frac{1}{\lambda_{i_1i_2\ldots i_J}^m} g(x_1^+, \ldots, x_J^+) \right] = E \left[ \frac{t(x_1^+ + me_{i_1}, \ldots, x_J^+ + me_{i_J})}{t(x_1^+, \ldots, x_J^+)} g(x_1^+ + me_{i_1}, \ldots, x_J^+ + me_{i_J}) \right],
$$

where $e_{ij}$ are given as in Lemma 2.1.

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

We consider the following class of Chou-type estimators,

$$
\delta_{\phi,\gamma} = \left\{ \delta_{\phi_{i_1i_2\ldots i_J}} \right\},
$$

$$
\delta_{\phi_{i_1i_2\ldots i_J}} = \delta_{\text{ML}_{i_1i_2\ldots i_J}} \left( 1 - \frac{\phi(x^+)}{(x^+ + c)^{\gamma+1}} \right),
$$

where $\phi(\cdot)$, $c$, and $\gamma$ 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

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

The proof is omitted here.
for all $x \geq 0$, then $\delta^{\phi, \gamma}$ dominates $\delta^{ML}$ under the loss function (3.2).

We use the following lemma to prove this theorem.

**Lemma 3.2.**

\[
\prod_{j=1}^{J} (x^+ + I_j) = x^+ + 1 + \sum_{j=1}^{J} I_j - J + \xi_J(x^+),
\]

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

**Proof.** By (2.9) with $y = x^+ + 1$, $K = J$, and $a_j = I_j - 1$, $j = 1, \ldots, J$, we see that $(x^+ + 1)^{J-1} \xi_J(\cdot)$ is positive. Hence $\xi_J(x^+)$ is positive for all $x^+ \geq 0$.

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

\[
\hat{R}_d(\delta^{\phi, \gamma}) = 2 \sum_{j=1}^{J} \sum_{i,j=1}^{I_j} \left\{ \frac{\phi(x^+ + 1)}{B^{\gamma + 1}} \cdot \frac{\prod_{j=1}^{J} (x^+ + I_j)}{(x^+ + 1)^{J-1}} - \frac{\phi(x^+)}{(B - 1)^{\gamma + 1}} \cdot \frac{\prod_{j=1}^{J} (x^+ + I_j)}{(x^+ + 1)^{J-1}} \right\}
\]

\[
- \sum_{j=1}^{J} \sum_{i,j=1}^{I_j} \frac{\phi^2(x^+ + 1)}{B^{2\gamma + 2}} \cdot \frac{\prod_{j=1}^{J} (x^+ + I_j)}{(x^+ + 1)^{J-1}}
\]

\[
= 2 \left\{ \frac{\phi(x^+ + 1)}{B^{\gamma + 1}} \cdot \frac{\prod_{j=1}^{J} (x^+ + I_j)}{(x^+ + 1)^{J-1}} - \frac{\phi(x^+)}{(B - 1)^{\gamma + 1}} \cdot \frac{\prod_{j=1}^{J} (x^+ + I_j)}{(x^+ + 1)^{J-1}} \right\}
\]

\[
- \frac{\phi^2(x^+ + 1)}{B^{2\gamma + 2}} \cdot \frac{\prod_{j=1}^{J} (x^+ + I_j)}{(x^+ + 1)^{J-1}},
\]

where $B = x^+ + c + 1$. By using (3.5), we can show the following in the same way as (2.6),

\[
2 \left\{ \frac{\phi(x^+ + 1)}{B^{\gamma + 1}} \cdot \frac{\prod_{j=1}^{J} (x^+ + I_j)}{(x^+ + 1)^{J-1}} - \frac{\phi(x^+)}{(B - 1)^{\gamma + 1}} \cdot \frac{\prod_{j=1}^{J} (x^+ + I_j)}{(x^+ + 1)^{J-1}} \right\}
\]

\[
\geq \frac{\phi(x^+ + 1)}{B^{\gamma + 2}(B - 1)^{\gamma + 1}} \left\{ 2B^{\gamma + 1}(B - (\gamma + 1)) \prod_{j=1}^{J} (x^+ + I_j) \right\} - 2B^{\gamma + 2}x^+
\]

\[
- \phi(x^+ + 1)B^{\gamma + 1} \prod_{j=1}^{J} (x^+ + I_j)
\]

\[
= \frac{\phi(x^+ + 1)}{B(B - 1)^{\gamma + 1}} \left\{ \left( 2 \sum_{j=1}^{J} I_j - 2\gamma - 2J - \phi(x^+ + 1) \right) x^+
\right.
\]

\[
+ (2c - 2\gamma - \phi(x^+ + 1)) \left( \sum_{j=1}^{J} I_j - (J - 1) \right)
\]

\[
+ (2(c + x^+) - 2\gamma - \phi(x^+ + 1))\xi_J(x^+),
\]
which completes the proof. □

Next we consider to generalize the Bayes estimator (2.6) to J-way layouts. The prior measure for the J-way model (3.1) which corresponds to (2.7) is expressed by

$$\pi(\lambda, \alpha_1, \ldots, \alpha_J \mid \nu) = m(\lambda) \prod_j \prod_{i_j} d\alpha_{j,i_j},$$

$$m(\lambda) = \int_0^\infty (1 + \lambda t)^{-\nu} t^{-(\sum_j I_j - J + 1)} \exp(-t^{-1}) dt,$$

$$\alpha_j = (\alpha_{j,1}, \ldots, \alpha_{j,I_j}), \quad j = 1, \ldots, J.$$

The Bayes estimator with respect to $$\pi(\lambda, \alpha_1, \ldots, \alpha_J \mid \nu)$$ is written by

$$\delta^\nu = \{ \delta^\nu_{i_1 i_2 \ldots i_J} \}.$$  

(3.6)

$$\delta^\nu_{i_1 i_2 \ldots i_J} = \frac{\prod_j x_{j,i_j}^{+} \cdot x^{+} + \sum_j I_j - J \cdot x^{+}}{\prod_j (x^{+} + I_j - 1) \cdot x^{+} + \nu + \sum_j I_j - J \cdot x^{+}},$$

where $$\nu \geq 0$$ is a constant. The following theorem is the generalization of Theorem 2.2.

**Theorem 3.2.** Suppose that $$I_j \geq 2, \ j = 1, \ldots, J.$$ If $$0 \leq \nu \leq \sum_{j=1}^J I_j - J,$$ $$\delta^\nu$$ improves on $$\delta^{ML}$$ under the loss function (3.2). If $$\nu > 1,$$ $$\delta^\nu$$ is admissible.

In order to prove this theorem we present the following lemmas which corresponds to Lemmas 2.2 and 2.3.

**Lemma 3.3.** Consider the following class of estimators

$$\delta^h = \left(1 - \frac{h(x^{+})}{x^{+}}\right) \delta^{ML}.$$  

If $$h(\cdot)$$ is nonnegative, nondecreasing and satisfies

(3.7) \[ \frac{2(x + 1)^J}{\prod_{j=1}^J (x + I_j)} - \left(1 - \frac{h(x + 1)}{x + 1}\right) \leq 1 \]

for all $$x \geq 0,$$ then $$\delta^h$$ dominates $$\delta^{ML}$$ under the loss (1.2).

This lemma can be proved in the same way as the proof of Lemma 2.2 by using Lemma 3.1 with $$m = 1.$$

**Lemma 3.4.** Define $$h^{\nu,\kappa}(\cdot)$$ by

$$h^{\nu,\kappa}(x) = x \left(1 - \frac{x^{J}(x + \kappa + \sum_{j=1}^J I_j - J)}{\prod_{j=1}^J (x + I_j - 1)(x + \nu + \kappa + \sum_{j=1}^J I_j - J)}\right).$$

Then $$h^{\nu,\kappa}(x)$$ is nondecreasing on $$x \geq 0$$ when $$\nu \geq 0$$ and $$\kappa \geq 0.$$
Proof. The proof is by induction on $J$. The lemma for $J = 2$ has been already proved in Lemma 2.3. Suppose $J > 2$ and assume that lemma holds up to $J - 1$. $h^{\nu,\kappa}(\cdot)$ can be rewritten by

$$h^{\nu,\kappa}(x) = x \left( 1 - \frac{I_J - 1}{x + I_J - 1} \right) \left( x^{J-1} + \frac{x^{J-1}(x + \kappa + \sum_{j=1}^J I_j - J)}{\prod_{j=1}^J(x + I_j - 1)(x + \nu + \kappa + \sum_{j=1}^J I_j - J)} \right)$$

$$= x \left( 1 - \frac{(I_J - 1)x^J}{\prod_{j=1}^J(x + I_j - 1)(x + \nu + \kappa + \sum_{j=1}^J I_j - J)} \right) \cdot \frac{(x + \nu + \kappa + \sum_{j=1}^J I_j - J)}{(x + \nu + \kappa + \sum_{j=1}^J I_j - J)},$$

where $\tilde{\kappa} = \kappa + I_J - 1$. The first term is nondecreasing from the inductive assumption. The second term is also nondecreasing. Thus we complete the proof.

$\square$

Proof of Theorem 3.2. Let $h(\cdot)$ for the class (3.6) be denoted by $h^{\nu}(\cdot)$. Then $h^{\nu}(\cdot)$ is written by

$$h^{\nu}(x) = x \left( 1 - \frac{x^J(x + \sum_{j=1}^J I_j - J)}{\prod_{j=1}^J(x + I_j - 1)(x + \nu + \sum_{j=1}^J I_j - J)} \right).$$

From Lemma 3.4 with $\kappa = 0$, $h^{\nu}(x)$ is nondecreasing for all $\nu \geq 0$.

In the similar way to (2.10), for $0 \leq \nu \leq \sum_{j=1}^J I_j - J$, we have

$$\frac{2(x + 1)^J}{\prod_{j=1}^J(x + I_j)} - \left( 1 - \frac{h^{\nu}(x + 1)}{x + 1} \right) = \frac{(x + 1)^J(x + 2\nu + \sum_{j=1}^J I_j - J + 1)}{\prod_{j=1}^J(x + I_j)(x + \nu + \sum_{j=1}^J I_j - J + 1)}$$

$$\leq \frac{(x + 1)^J}{\prod_{j=1}^J(x + I_j)} \cdot \frac{(x + 3 \sum_{j=1}^J I_j - 3J + 1)}{(x + 2 \sum_{j=1}^J I_j - 2J + 1)} \leq 1.$$ 

Hence $h^{\nu}(\cdot)$ satisfies (3.7). From Lemma 3.3 the proof is completed. $\square$

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,

$$x_{i_1i_2i_3} \sim \text{Po}(\lambda_{i_1i_2i_3}), \quad \lambda_{i_1i_2i_3} = \lambda^{\alpha_{i_1i_2} \beta_{i_2i_3}},$$

$$i_j = 1, \ldots, I_j, \quad I_j \geq 2, \quad j = 1, 2, 3.$$
The identity for this model is as follows.

\[ \sum_{i_1} \alpha_{i_1i_2} = \sum_{i_3} \beta_{i_2i_3} = \gamma_{i_2}, \quad \sum_{i_2} \gamma_{i_2} = 1. \]

The problem is to estimate \( \lambda = (\lambda_{111}, \ldots, \lambda_{1I_2I_3})' \) under the loss function

\[ L(\lambda, \hat{\lambda}) = \sum_{i_1,i_2,i_3} \frac{1}{\lambda_{i_1i_2i_3}} (\hat{\lambda}_{i_1i_2i_3} - \lambda_{i_1i_2i_3})^2. \]

Denote the relevant two-dimensional marginal frequencies by

\[ x_{1i_2+} = (x_{1i_21}, \ldots, x_{1i_2I_2})', \quad x_{+i_23} = (x_{+i_21}, \ldots, x_{+i_2I_3})', \quad x_{12} = (x_{1i_21}', \ldots, x_{1i_2I_2}'), \quad x_{23} = (x_{+i_21}', \ldots, x_{+i_2I_3}'). \]

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

\[ \Pr(x_{12}, x_{23}) = \exp(-\lambda) \lambda_{x_{1i_2+}+} \prod_{i_1,i_2} \alpha_{i_1i_2} \prod_{i_2} \beta_{i_2i_3}^{x_{+i_2i_3}} - t^+(x_{12}, x_{23}), \]

\[ t^+(x_{12}, x_{23}) = \prod_{i_2} t(x_{1i_2+}, x_{+i_23}), \]

\[ t(x_{1i_2+}, x_{+i_23}) = \frac{x_{+i_2+}!}{\prod_{i_1} x_{i_1i_2+}! \prod_{i_3} x_{+i_2i_3}!}. \]

The MLE of \( \lambda_{i_1i_2i_3} \) is

\[ \delta^{ML} = (\delta^{ML}_{111}, \ldots, \delta^{ML}_{I_1I_2I_3}), \quad \delta^{ML}_{i_1i_2i_3} = \begin{cases} \frac{x_{i_1i_2+} + x_{+i_2i_3}}{x_{+i_2+}} & \text{if } x_{+i_2+} \neq 0, \\ 0 & \text{if } x_{+i_2+} = 0. \end{cases} \]

The identity for this model is as follows.

**Lemma 4.1.** Let \( x_{12} \) and \( x_{23} \) have the probability function (4.3). Let \( g(\cdot), e_{i_1} \) and \( e_{i_3} \) be given as in Lemma 2.1. Define \( e_{i_1i_2} \) by the \( 1 \times 1 \) vectors with 1 as the \((i_1 + I_2(i_2 - 1))\)-th component and 0 for others. \( e_{i_2i_3} \) is defined in the same way as \( e_{i_1i_2} \). Then

\[ \mathbb{E} \left[ \frac{1}{\lambda^{m}_{i_1i_2i_3}} g(x_{12}, x_{23}) \right] = \mathbb{E} \left[ \frac{t(x_{1i_2+} + me_{i_1}, x_{+i_23} + me_{i_3})}{t(x_{1i_2+}, x_{+i_23})} \right] \cdot g(x_{12} + me_{i_1i_2}, x_{23} + me_{i_2i_3}). \]

This lemma can be also proved in the same way as Hudson (1978) and Hwang (1982). Specializing (4.4) to \( g \) which depends only on \( x_{1i_2+}, x_{+i_23} \) we obtain the
following identity.

\[
\begin{align*}
(4.5) \quad & \quad E \left[ \frac{1}{\lambda_{1123}^m} g(x_{1i2+}, x_{+i23}) \right] \\
& \quad = E \left[ \frac{t(x_{1i2+} + me_{i1}, x_{+i23} + me_{i3})}{t(x_{1i2+}, x_{+i23})} \cdot g(x_{1i2+} + me_{i1}, x_{+i23} + me_{i3}) \right].
\end{align*}
\]

From (4.3) if \( i_2 \) is fixed, the model (4.1) is reduced to two-way multiplicative model (1.1). Thus we first consider the following class,

\[
(4.6) \quad \delta^\psi = \delta^{ML} - \Psi(x_{12}, x_{23}) = (\delta^\psi_{111}, \ldots, \delta^\psi_{I_1I_2I_3})',
\]

\[
\Psi(x_{12}, x_{23}) = \{\Psi_{i_1i_2i_3}(x_{1i2+}, x_{+i23})\}.
\]

By using (4.5) in Lemma 4.1, \( \hat{R}_d(\delta^\psi) \) is expressed by

\[
\begin{align*}
\hat{R}_d(\delta^\psi) & = \sum_{i_2} \hat{R}_d(\delta^\psi_{i_2}), \\
(4.7) \quad \hat{R}_d(\delta^\psi_{i_2}) & = 2 \sum_{i_1,i_3} (\Psi_{i_1i_2i_3}(x_{1i2+} + e_{i1}, x_{+i23} + e_{i3}) - \Psi_{i_1i_2i_3}(x_{1i2+}, x_{+i23})) \\
& \quad - \sum_{i_1,i_3} \left( \frac{x_{i2} + 1}{(x_{i2} + 1)(x_{i2} + 1)} \right) \Psi_{i_1i_2i_3}(x_{1i2+} + e_{i1}, x_{+i23} + e_{i3}),
\end{align*}
\]

where \( \delta^\psi_{i_2} = (\delta^\psi_{i_21}, \ldots, \delta^\psi_{i_2I_1I_3})' \). (4.7) corresponds to (2.3). Since \( \hat{R}_d(\delta^\psi_{i_2}) \geq 0 \) for each \( i_2 \) implies \( \hat{R}_d(\delta^\psi) \geq 0 \), the following results on two classes of estimators can be obtained in the same way as Theorems 2.1 and 2.2.

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

\[
0 \leq \phi_{i_2}(x) \leq \min(2(I_1 + I_3) - 2\gamma - 4, 2c - 2\gamma, (x + c)^{\gamma+1}), \quad c > 0, \quad \gamma \geq 0,
\]

for all \( x \geq 0 \). Then \( \delta^{\phi_{i_2}} = (\delta^{\phi_{i_21}}, \ldots, \delta^{\phi_{I_1I_2I_3}})' \),

\[
(4.8) \quad \delta^{\phi_{i_2}}_{i_1i_2i_3} = \delta^{ML}_{i_1i_2i_3} \left( 1 - \frac{\phi_{i_2}(x_{i2} + 1)}{(x_{i2} + 1 + c)^{\gamma+1}} \right) \\
= \frac{x_{i1i2+} x_{i2i3}}{x_{i2+}} \left( 1 - \frac{\phi_{i_2}(x_{i2} + 1)}{(x_{i2} + 1 + c)^{\gamma+1}} \right)
\]
dominates \( \delta^{ML} \) under the loss function (4.2).

**Theorem 4.2.** \( \delta^{\nu_{i_2}} = (\delta^{\nu_{i_21}}, \ldots, \delta^{\nu_{i_2I_1I_3}})' \),

\[
\delta^{\nu_{i_2}}_{i_1i_2i_3} = \frac{x_{i1i2+} x_{i2i3} + 1}{(x_{i2+} + 1)(x_{i2+} + I_1 + I_3 - 1)} \cdot \frac{x_{i2+} + I_1 + I_3 - 2}{x_{i2+} + \nu_{i_2} + I_1 + I_3 - 2} \cdot x_{i2+}
\]
dominates $\mathbf{\delta}^{ML}$ under the loss function (4.2) if $0 \leq \nu_{i_2} \leq I_1 + I_3 - 2$, $i_2 = 1, \ldots, I_2$. If $\nu_{i_2} > 1$, $i_2 = 1, \ldots, I_2$, then $\mathbf{\delta}^{\nu_{i_2}}$ is admissible.

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

$$
\prod_{i_2=1}^{I_2} \pi(\lambda_{i_2}, \theta_{i_2}, \eta_{i_2} | \nu_{i_2}) d\lambda_{i_2} \prod_{i_1} d\theta_{i_1 i_2} \prod_{i_3} d\eta_{i_2 i_3}
$$

$$
= \prod_{i_2=1}^{I_2} m(\lambda_{i_2} | \nu_{i_2}) d\lambda_{i_2} \prod_{i_1} d\theta_{i_1 i_2} \prod_{i_3} d\eta_{i_2 i_3},
$$

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

$$
\theta_{i_2} = (\theta_{i_1}, \ldots, \theta_{I_1 i_2})', \quad \eta_{i_2} = (\eta_{i_2 1}, \ldots, \eta_{i_2 I_3})'.
$$

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

$$
\mathbf{\delta}^{\phi, \gamma} = (\delta_{111}^{\phi, \gamma}, \ldots, \delta_{I_1 I_2 I_3}^{\phi, \gamma})',
$$

$$
\delta_{i_1 i_2 i_3}^{\phi, \gamma} = \delta_{i_1 i_2 i_3}^{ML} \left( 1 - \frac{\phi(x_{i_1 i_2 i_3} + c)}{(x_{i_1 i_2 i_3} + c)^{\gamma + 1}} \right),
$$

where $x_{i_1 i_2 i_3}$ is the total sample size. Note that in this class the amount of shrinkage depends only on the total sample size $x_{i_1 i_2 i_3}$ and hence this class does not belong to (4.6). We can find improved estimators also in this class.

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

$$
0 \leq \phi(x) \leq \min(2I_2(I_1 + I_3 - 1) - 2\gamma - 2, 2c - 2\gamma, (x + c)^{\gamma + 1})
$$

for all $x \geq 0$, then $\mathbf{\delta}^{\phi, \gamma}$ dominates $\mathbf{\delta}^{ML}$ under the loss function (4.2).

**Proof.** By using (4.4) in Lemma 4.1, $\hat{R}(\mathbf{\delta}^{\phi, \gamma})$ is

$$
\hat{R}(\mathbf{\delta}^{\phi, \gamma}) = 2 \sum_{i_2} \sum_{i_1, i_3} \left\{ \frac{\phi(x_{i_1 i_2 + i_3} + 1)}{B^{\gamma + 1}} \frac{(x_{i_1 i_2 + i_3} + 1)(x_{i_1 i_2 + i_3} + 1)}{x_{i_2} + 1} \right\}
$$

$$
- \phi(x_{i_1 i_2 + i_3} + 1) \frac{x_{i_1 i_2 + i_3}}{(B - 1)^{\gamma + 1}} \frac{x_{i_1 i_2 + i_3}}{x_{i_2} + 1}
$$

$$
- \sum_{i_2} \sum_{i_1, i_3} \phi^2(x_{i_1 i_2 + i_3} + 1) \frac{(x_{i_1 i_2 + i_3} + 1)(x_{i_1 i_2 + i_3} + 1)}{x_{i_2} + 1}
$$

$$
= 2 \sum_{i_2} \left\{ \frac{\phi(x_{i_1 i_2 + 1} + 1)}{B^{\gamma + 1}} \frac{(x_{i_1 i_2 + 1} + 1)(x_{i_1 i_2 + 1} + 1)}{x_{i_2} + 1} \right\}
$$

$$
- \phi(x_{i_1 i_2 + 1} + 1) \frac{x_{i_1 i_2 + 1}}{(B - 1)^{\gamma + 1}}
$$

$$
- \sum_{i_2} \phi^2(x_{i_1 i_2 + 1} + 1) \frac{(x_{i_1 i_2 + 1} + 1)(x_{i_1 i_2 + 1} + 1)}{x_{i_2} + 1},
$$
where \( B = x_{+++} + c + 1 \). Note that
\[
(4.10) \quad \sum_{i_2} \frac{(x_{++} + I_1)(x_{++} + I_3)}{x_{++} + 1} = x_{+++} + I_2(I_1 + I_3 - 1) + \sum_{i_2} \frac{(I_1 - 1)(I_3 - 1)}{x_{++} + 1} \geq x_{+++} + I_2(I_1 + I_3 - 1) + \frac{I_3^2(I_1 - 1)(I_3 - 1)}{x_{+++} + I_2} = \frac{(x_{+++} + I_1 I_2)(x_{+++} + I_2 I_3)}{x_{+++} + I_2}.
\]

Since \( \phi(x_{+++} + 1)/B^{\gamma+1} \leq 1 \) under the condition (4.9), \( \tilde{R}_d(\delta^{\phi,\gamma}) \) is bounded below by
\[
(4.11) \quad \tilde{R}_d(\delta^{\phi,\gamma}) = 2 \left( \frac{\phi(x_{+++} + 1)}{B^{\gamma+1}} \right) \cdot \frac{(x_{+++} + I_1 I_2)(x_{+++} + I_2 I_3)}{x_{+++} + I_2} - \frac{\phi(x_{+++})}{(B - 1)^{\gamma+1}} \cdot (x_{+++}) - \frac{\phi^2(x_{+++} + 1)}{B^{2\gamma+2}} \cdot \frac{(x_{+++} + I_1 I_2)(x_{+++} + I_2 I_3)}{x_{+++} + I_2}.
\]

(4.11) corresponds to (2.5) in the proof of Theorem 2.1. By using (4.10), \( \tilde{R}_d(\delta^{\phi,\gamma}) \geq 0 \) under (4.9) can be proved in the same way as Theorem 2.1. \( \Box \)

5. Monte Carlo study

We study the risk performance of the proposed estimators through Monte Carlo studies. Tables 1 and 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. \( \delta^c \) in Tables 1 and 2 are
\[
\delta^c = \delta^{ML} \left( 1 - \frac{c}{x_{+++} + c} \right) \quad \text{and} \quad \delta^c = \delta^{ML} \left( 1 - \frac{c}{x_{+++} + c} \right),
\]
respectively. \( \delta^c \) is \( \delta^{\phi,\gamma} \) in Theorem 3.1 with \( \phi = c \) and \( \gamma = 0 \). We set \( \alpha_{ji} \) all equal to \( 1/I_j \).

Tables 3 and 4 present the risks of the MLE and the proposed estimators for three-way decomposable models obtained from 100,000 replications. \( \delta^{c_2} \) and \( \delta^c \) in Tables 3 and 4 are
\[
\delta^{c_2} = \delta^{ML} \left( 1 - \frac{c_{i_2}}{x_{++} + c_{i_2}} \right) \quad \text{and} \quad \delta^c = \delta^{ML} \left( 1 - \frac{c}{x_{+++} + c} \right),
\]
respectively. \( \delta^{c_2} \) is \( \delta^{\phi_{i_2},\gamma} \) in Theorem 4.1 with \( \phi_{i_2} = c_2 \) and \( \gamma = 0 \). \( \delta^c \) is \( \delta^{\phi,\gamma} \) 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 \( \lambda_{i_2} \) are set to \( \lambda/I_2 \) and Table 4 is the case that not all of \( \lambda_{i_2} \) are equal.

The summary of these experiments is as follows.
Table 1. Risks of the MLE and the proposed estimators for two-way multiplicative models.

(i) $2 \times 2$ contingency tables

<table>
<thead>
<tr>
<th>$\lambda$</th>
<th>$\delta^{ML}$</th>
<th>$\delta^c$</th>
<th>$\delta^\nu$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$c = 4$</td>
<td>$c = 2$</td>
<td>$c = 1$</td>
<td>$\nu = 2$</td>
</tr>
<tr>
<td>0.1</td>
<td>3.956</td>
<td>0.254</td>
<td>0.546</td>
</tr>
<tr>
<td>0.5</td>
<td>3.794</td>
<td>0.586</td>
<td>0.892</td>
</tr>
<tr>
<td>1.0</td>
<td>3.640</td>
<td>0.932</td>
<td>1.234</td>
</tr>
<tr>
<td>2.0</td>
<td>3.432</td>
<td>1.452</td>
<td>1.708</td>
</tr>
<tr>
<td>5.0</td>
<td>3.196</td>
<td>2.261</td>
<td>2.355</td>
</tr>
<tr>
<td>10.0</td>
<td>3.102</td>
<td>2.706</td>
<td>2.675</td>
</tr>
<tr>
<td>20.0</td>
<td>3.044</td>
<td>2.914</td>
<td>2.837</td>
</tr>
<tr>
<td>50.0</td>
<td>3.023</td>
<td>2.898</td>
<td>2.955</td>
</tr>
</tbody>
</table>

(ii) $3 \times 3$ contingency tables

<table>
<thead>
<tr>
<th>$\lambda$</th>
<th>$\delta^{ML}$</th>
<th>$\delta^c$</th>
<th>$\delta^\nu$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$c = 8$</td>
<td>$c = 4$</td>
<td>$c = 2$</td>
<td>$\nu = 4$</td>
</tr>
<tr>
<td>0.1</td>
<td>8.824</td>
<td>0.210</td>
<td>0.475</td>
</tr>
<tr>
<td>0.5</td>
<td>8.169</td>
<td>0.575</td>
<td>0.881</td>
</tr>
<tr>
<td>1.0</td>
<td>7.525</td>
<td>0.979</td>
<td>1.308</td>
</tr>
<tr>
<td>2.0</td>
<td>6.732</td>
<td>1.653</td>
<td>1.981</td>
</tr>
<tr>
<td>5.0</td>
<td>5.785</td>
<td>2.942</td>
<td>3.121</td>
</tr>
<tr>
<td>10.0</td>
<td>5.388</td>
<td>3.933</td>
<td>3.895</td>
</tr>
<tr>
<td>50.0</td>
<td>5.085</td>
<td>4.944</td>
<td>4.764</td>
</tr>
<tr>
<td>100.0</td>
<td>5.038</td>
<td>4.997</td>
<td>4.877</td>
</tr>
</tbody>
</table>

(iii) $5 \times 5$ contingency tables

<table>
<thead>
<tr>
<th>$\lambda$</th>
<th>$\delta^{ML}$</th>
<th>$\delta^c$</th>
<th>$\delta^\nu$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$c = 16$</td>
<td>$c = 8$</td>
<td>$c = 4$</td>
<td>$\nu = 8$</td>
</tr>
<tr>
<td>0.1</td>
<td>24.234</td>
<td>0.188</td>
<td>0.428</td>
</tr>
<tr>
<td>0.5</td>
<td>21.619</td>
<td>0.577</td>
<td>0.873</td>
</tr>
<tr>
<td>1.0</td>
<td>19.098</td>
<td>1.029</td>
<td>1.373</td>
</tr>
<tr>
<td>2.0</td>
<td>15.934</td>
<td>1.841</td>
<td>2.236</td>
</tr>
<tr>
<td>5.0</td>
<td>12.173</td>
<td>3.685</td>
<td>4.017</td>
</tr>
<tr>
<td>10.0</td>
<td>10.609</td>
<td>5.550</td>
<td>5.610</td>
</tr>
<tr>
<td>50.0</td>
<td>9.316</td>
<td>8.590</td>
<td>8.095</td>
</tr>
<tr>
<td>100.0</td>
<td>9.118</td>
<td>8.889</td>
<td>8.494</td>
</tr>
</tbody>
</table>

- 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 $\lambda$ is small.
Table 2. Risks of the MLE and the proposed estimators for three-way multiplicative models.

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

<table>
<thead>
<tr>
<th>$\lambda$</th>
<th>$\delta^{ML}$</th>
<th>$\delta^c$</th>
<th>$\delta^\nu$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$c = 6$</td>
<td>$c = 3$</td>
<td>$c = 1.5$</td>
</tr>
<tr>
<td>0.1</td>
<td>7.830</td>
<td>0.262</td>
<td>0.612</td>
</tr>
<tr>
<td>0.5</td>
<td>7.056</td>
<td>0.614</td>
<td>0.987</td>
</tr>
<tr>
<td>1.0</td>
<td>6.386</td>
<td>0.998</td>
<td>1.380</td>
</tr>
<tr>
<td>2.0</td>
<td>5.549</td>
<td>1.608</td>
<td>1.959</td>
</tr>
<tr>
<td>5.0</td>
<td>4.640</td>
<td>2.681</td>
<td>2.851</td>
</tr>
<tr>
<td>50.0</td>
<td>4.059</td>
<td>3.987</td>
<td>3.876</td>
</tr>
<tr>
<td>100.0</td>
<td>4.028</td>
<td>4.008</td>
<td>3.937</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>$\lambda$</th>
<th>$\delta^{ML}$</th>
<th>$\delta^c$</th>
<th>$\delta^\nu$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$c = 12$</td>
<td>$c = 6$</td>
<td>$c = 3$</td>
</tr>
<tr>
<td>0.1</td>
<td>25.726</td>
<td>0.261</td>
<td>0.668</td>
</tr>
<tr>
<td>0.5</td>
<td>21.948</td>
<td>0.643</td>
<td>1.094</td>
</tr>
<tr>
<td>1.0</td>
<td>18.498</td>
<td>1.081</td>
<td>1.568</td>
</tr>
<tr>
<td>2.0</td>
<td>14.227</td>
<td>1.843</td>
<td>2.350</td>
</tr>
<tr>
<td>10.0</td>
<td>8.275</td>
<td>4.925</td>
<td>4.989</td>
</tr>
<tr>
<td>20.0</td>
<td>7.592</td>
<td>6.095</td>
<td>5.858</td>
</tr>
<tr>
<td>100.0</td>
<td>7.113</td>
<td>6.991</td>
<td>6.751</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>$\lambda$</th>
<th>$\delta^{ML}$</th>
<th>$\delta^c$</th>
<th>$\delta^\nu$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$c = 24$</td>
<td>$c = 12$</td>
<td>$c = 6$</td>
</tr>
<tr>
<td>0.1</td>
<td>117.471</td>
<td>0.303</td>
<td>0.858</td>
</tr>
<tr>
<td>0.5</td>
<td>94.579</td>
<td>0.703</td>
<td>1.309</td>
</tr>
<tr>
<td>1.0</td>
<td>74.247</td>
<td>1.183</td>
<td>1.848</td>
</tr>
<tr>
<td>2.0</td>
<td>49.779</td>
<td>2.067</td>
<td>2.798</td>
</tr>
<tr>
<td>5.0</td>
<td>25.797</td>
<td>4.238</td>
<td>4.943</td>
</tr>
<tr>
<td>50.0</td>
<td>13.967</td>
<td>11.936</td>
<td>11.216</td>
</tr>
</tbody>
</table>

- The improvement is in the inverse proportion to $\lambda$.
- We can see from Tables 1 and 2 that $\delta^\nu$ shows larger risk reduction than $\delta^c$ when $\lambda$ is small.
- We can see from Tables 3 and 4 that $\delta^c$ shows larger risk reduction when
Table 3. Risks of the MLE and the proposed estimators for the decomposable models for three-way contingency tables when all of $\lambda_i$ are equal to $\lambda/I_2$.

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

<table>
<thead>
<tr>
<th>$\lambda$</th>
<th>$\delta^{ML}$</th>
<th>$\delta^c_{i_2} = 4$</th>
<th>$\delta^c_{i_2} = 2$</th>
<th>$\delta^c_{c = 10}$</th>
<th>$\delta^c_{c = 5}$</th>
<th>$\delta^v = 2$</th>
<th>$\delta^v = 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>7.981</td>
<td>0.416</td>
<td>0.996</td>
<td>0.165</td>
<td>0.337</td>
<td>0.282</td>
<td>0.389</td>
</tr>
<tr>
<td>0.5</td>
<td>7.772</td>
<td>0.768</td>
<td>1.366</td>
<td>0.535</td>
<td>0.752</td>
<td>0.659</td>
<td>0.780</td>
</tr>
<tr>
<td>1.0</td>
<td>7.575</td>
<td>1.171</td>
<td>1.781</td>
<td>0.954</td>
<td>1.206</td>
<td>1.089</td>
<td>1.223</td>
</tr>
<tr>
<td>2.0</td>
<td>7.265</td>
<td>1.863</td>
<td>2.463</td>
<td>1.670</td>
<td>1.941</td>
<td>1.825</td>
<td>1.967</td>
</tr>
<tr>
<td>10.0</td>
<td>6.397</td>
<td>4.525</td>
<td>4.716</td>
<td>4.373</td>
<td>4.308</td>
<td>4.587</td>
<td>4.584</td>
</tr>
<tr>
<td>20.0</td>
<td>6.209</td>
<td>5.419</td>
<td>5.358</td>
<td>5.307</td>
<td>5.056</td>
<td>5.463</td>
<td>5.354</td>
</tr>
<tr>
<td>50.0</td>
<td>6.070</td>
<td>5.893</td>
<td>5.740</td>
<td>5.850</td>
<td>5.586</td>
<td>5.906</td>
<td>5.778</td>
</tr>
<tr>
<td>100.0</td>
<td>6.040</td>
<td>5.990</td>
<td>5.877</td>
<td>5.975</td>
<td>5.793</td>
<td>5.994</td>
<td>5.906</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>$\lambda$</th>
<th>$\delta^{ML}$</th>
<th>$\delta^c_{i_2} = 8$</th>
<th>$\delta^c_{i_2} = 4$</th>
<th>$\delta^c_{c = 28}$</th>
<th>$\delta^c_{c = 14}$</th>
<th>$\delta^v = 4$</th>
<th>$\delta^v = 2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>26.787</td>
<td>0.433</td>
<td>1.196</td>
<td>0.134</td>
<td>0.237</td>
<td>0.206</td>
<td>0.278</td>
</tr>
<tr>
<td>0.5</td>
<td>26.003</td>
<td>0.817</td>
<td>1.628</td>
<td>0.529</td>
<td>0.681</td>
<td>0.605</td>
<td>0.692</td>
</tr>
<tr>
<td>1.0</td>
<td>25.161</td>
<td>1.277</td>
<td>2.142</td>
<td>1.003</td>
<td>1.207</td>
<td>1.086</td>
<td>1.190</td>
</tr>
<tr>
<td>2.0</td>
<td>23.773</td>
<td>2.146</td>
<td>3.091</td>
<td>1.892</td>
<td>2.176</td>
<td>1.995</td>
<td>2.125</td>
</tr>
<tr>
<td>5.0</td>
<td>20.835</td>
<td>4.339</td>
<td>5.343</td>
<td>4.131</td>
<td>4.468</td>
<td>4.312</td>
<td>4.458</td>
</tr>
<tr>
<td>50.0</td>
<td>15.722</td>
<td>13.399</td>
<td>12.918</td>
<td>13.228</td>
<td>12.286</td>
<td>13.539</td>
<td>13.037</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>$\lambda$</th>
<th>$\delta^{ML}$</th>
<th>$\delta^c_{i_2} = 16$</th>
<th>$\delta^c_{i_2} = 8$</th>
<th>$\delta^c_{c = 88}$</th>
<th>$\delta^c_{c = 44}$</th>
<th>$\delta^v = 8$</th>
<th>$\delta^v = 4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>123.509</td>
<td>0.533</td>
<td>1.657</td>
<td>0.118</td>
<td>0.173</td>
<td>0.158</td>
<td>0.200</td>
</tr>
<tr>
<td>0.5</td>
<td>119.929</td>
<td>0.932</td>
<td>2.109</td>
<td>0.522</td>
<td>0.615</td>
<td>0.560</td>
<td>0.610</td>
</tr>
<tr>
<td>1.0</td>
<td>116.596</td>
<td>1.432</td>
<td>2.695</td>
<td>1.023</td>
<td>1.165</td>
<td>1.061</td>
<td>1.123</td>
</tr>
<tr>
<td>2.0</td>
<td>110.322</td>
<td>2.404</td>
<td>3.805</td>
<td>2.006</td>
<td>2.250</td>
<td>2.044</td>
<td>2.125</td>
</tr>
<tr>
<td>5.0</td>
<td>95.103</td>
<td>5.146</td>
<td>6.849</td>
<td>4.800</td>
<td>5.298</td>
<td>4.860</td>
<td>4.980</td>
</tr>
<tr>
<td>20.0</td>
<td>64.717</td>
<td>15.744</td>
<td>17.615</td>
<td>15.529</td>
<td>16.056</td>
<td>16.022</td>
<td>16.007</td>
</tr>
<tr>
<td>50.0</td>
<td>52.984</td>
<td>27.695</td>
<td>27.992</td>
<td>27.428</td>
<td>26.285</td>
<td>28.445</td>
<td>27.604</td>
</tr>
<tr>
<td>100.0</td>
<td>49.033</td>
<td>36.386</td>
<td>34.864</td>
<td>36.059</td>
<td>33.320</td>
<td>37.049</td>
<td>35.316</td>
</tr>
</tbody>
</table>
Table 4. Risks of the MLE and the proposed estimators for the decomposable models for three-way contingency tables when not all of $\lambda_{i2}$ are equal.

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

<table>
<thead>
<tr>
<th>$(\lambda_1, \lambda_2)$</th>
<th>$\delta^{ML}$</th>
<th>$\delta^{c_{i2}}$</th>
<th>$\delta^c$</th>
<th>$\delta^\nu$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$c_{i2} = 4$</td>
<td>$c_{i2} = 2$</td>
<td>$c = 10$</td>
</tr>
<tr>
<td>(0.1, 0.9)</td>
<td>7.580</td>
<td>1.116</td>
<td>1.705</td>
<td>0.955</td>
</tr>
<tr>
<td>(0.2, 0.8)</td>
<td>7.585</td>
<td>1.141</td>
<td>1.741</td>
<td>0.955</td>
</tr>
<tr>
<td>(0.3, 0.7)</td>
<td>7.572</td>
<td>1.157</td>
<td>1.762</td>
<td>0.954</td>
</tr>
<tr>
<td>(0.4, 0.6)</td>
<td>7.570</td>
<td>1.166</td>
<td>1.774</td>
<td>0.954</td>
</tr>
<tr>
<td>(1.0, 0.9)</td>
<td>6.736</td>
<td>3.586</td>
<td>3.867</td>
<td>4.462</td>
</tr>
<tr>
<td>(2.0, 0.8)</td>
<td>6.548</td>
<td>4.044</td>
<td>4.299</td>
<td>4.413</td>
</tr>
<tr>
<td>(3.0, 0.7)</td>
<td>6.461</td>
<td>4.323</td>
<td>4.547</td>
<td>4.384</td>
</tr>
<tr>
<td>(4.0, 0.6)</td>
<td>6.417</td>
<td>4.477</td>
<td>4.679</td>
<td>4.374</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>$(\lambda_1, \lambda_2, \lambda_3)$</th>
<th>$\delta^{ML}$</th>
<th>$\delta^{c_{i2}}$</th>
<th>$\delta^c$</th>
<th>$\delta^\nu$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$c_{i2} = 8$</td>
<td>$c_{i2} = 4$</td>
<td>$c = 28$</td>
</tr>
<tr>
<td>(0.1, 0.1, 0.8)</td>
<td>25.427</td>
<td>1.247</td>
<td>2.101</td>
<td>1.007</td>
</tr>
<tr>
<td>(0.1, 0.2, 0.7)</td>
<td>25.339</td>
<td>1.258</td>
<td>2.118</td>
<td>1.005</td>
</tr>
<tr>
<td>(0.1, 0.3, 0.6)</td>
<td>25.314</td>
<td>1.266</td>
<td>2.128</td>
<td>1.004</td>
</tr>
<tr>
<td>(0.2, 0.3, 0.5)</td>
<td>25.250</td>
<td>1.272</td>
<td>2.135</td>
<td>1.004</td>
</tr>
<tr>
<td>(0.3, 0.3, 0.4)</td>
<td>25.177</td>
<td>1.276</td>
<td>2.139</td>
<td>1.003</td>
</tr>
<tr>
<td>(1.0, 1.0, 0.8)</td>
<td>20.559</td>
<td>5.601</td>
<td>6.291</td>
<td>6.982</td>
</tr>
<tr>
<td>(1.0, 2.0, 7.0)</td>
<td>19.827</td>
<td>6.076</td>
<td>6.809</td>
<td>6.919</td>
</tr>
<tr>
<td>(1.0, 3.0, 6.0)</td>
<td>19.456</td>
<td>6.380</td>
<td>7.120</td>
<td>6.893</td>
</tr>
<tr>
<td>(2.0, 3.0, 5.0)</td>
<td>18.791</td>
<td>6.780</td>
<td>7.578</td>
<td>6.839</td>
</tr>
<tr>
<td>(3.0, 3.0, 4.0)</td>
<td>18.492</td>
<td>6.966</td>
<td>7.779</td>
<td>6.814</td>
</tr>
</tbody>
</table>

$\lambda_{i2}$ are close together. On the contrary $\delta^{c_{i2}}$ and $\delta^{\nu_{i2}}$ seem to be better than $\delta^c$ when $\lambda_{i2}$ varies widely and $\lambda$ 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 the estimators proposed in this article 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.

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.
Acknowledgements

The authors are grateful to two anonymous referees for helpful comments and suggestions which have led to essential improvements to the original version of the manuscript.

References


