Simultaneous estimation of Poisson means in two-way contingency tables under normalized squared error loss in multiplicative models

*Chang Yuan-Tsung, **Shinozaki Nobuo
*Department of Social Information, Mejiro University
**Faculty of Science and Technology, Keio University

Abstract

Shrinkage estimation of Poisson means is considered when observations are given in the form of a two-way contingency table. Assuming a multiplicative Poisson model, estimators which shrink to the specified values or an order statistic in one dimension and in two dimensions are considered and are shown to dominate the maximum likelihood estimator (MLE) under normalized squared error loss.

1 Introduction

Consider a two-way multiplicative model where x_{ij}, $i = 1, \ldots, I$, $j = 1, \ldots, J$, are independent random Poisson random variables with means

$$\lambda_{ij} = \lambda\alpha_i\beta_j, \quad i = 1, \ldots, I, \ \ j = 1, \ldots, J,$$

where $\alpha_i \geq 0$ and $\beta_j \geq 0$ satisfy $\sum_{i=1}^{I} \alpha_i = 1$ and $\sum_{j=1}^{J} \beta_j = 1$, respectively. We denote the one-dimensional frequencies and the total frequency by

$$x_{i+} = \sum_{j=1}^{J} x_{ij},\ i = 1, \ldots, I,\ x_{+j} = \sum_{i=1}^{I} x_{ij},\ j = 1, \ldots, J,\ x_{++} = \sum_{i=1}^{I}\sum_{j=1}^{J} x_{ij}.$$

As discussed in Hara and Takemura (2006) complete sufficient statistics are $\boldsymbol{x}_1 = (x_{1+}, \ldots, x_{I+})$ and $\boldsymbol{x}_2 = (x_{+1}, \ldots, x_{+J})$. The MLE of λ_{ij} is

$$\hat{\lambda}_{ij}^{ML} = \begin{cases} \dfrac{x_{i+}x_{+j}}{x_{++}} & \text{if} \quad x_{++} \neq 0 \\ 0 & \text{if} \quad x_{++} = 0. \end{cases}$$

They have given a class of improved estimators which shrink the MLE toward the origin under the normalized squared error loss. The simple one is

$$\delta_{ij}^{HT} = \frac{x_{i+}x_{+j}}{x_{++}}\left\{1 - \frac{d}{x_{++} + d}\right\}, \quad i = 1, \ldots, I,\ j = 1, \ldots, J, .$$

where d is a positive constant, $0 < d \leq 2(I + J - 2)$ and $I, J \geq 2$.

The following lemma is a special case of Lemma 2.1 of Hara and Takemura (2006) and is useful to evaluate the risk of the shrinkage estimators when normalized squared error loss is concerned.

Lemma 1.1. If $g(\boldsymbol{x}_1, \boldsymbol{x}_2)$ is a real-valued function satisfying $E|g(\boldsymbol{x}_1, \boldsymbol{x}_2)| < \infty$ and $g(\boldsymbol{x}_1, \boldsymbol{x}_2) = 0$ when $x_{i+} = 0$ or $x_{+j} = 0$, then

$$E\left\{\frac{g(\boldsymbol{x}_1, \boldsymbol{x}_2)}{\lambda_{ij}}\right\} = E\left\{\frac{(x_{++} + 1)}{(x_{i+} + 1)(x_{+j} + 1)} g(\boldsymbol{x}_1 + \boldsymbol{e}_i^I, \boldsymbol{x}_2 + \boldsymbol{e}_j^J)\right\},$$

where $\boldsymbol{e}_i^I$ ($\boldsymbol{e}_j^J$) is $I \times 1$ ($J \times 1$) unit vector with i-th (j-th) component 1.

In Section 2 multiplicative Poisson model is assumed and shrinkage to a specified value or an order statistic is considered in one dimension and in two dimensions. In subsections 2.1 and 2.2 we give improved estimators which shrink x_{i+} in the MLE to an order statistic and a given non-negative constant, respectively. Further, in subsections 2.3 and 2.4 we are concerned with impoved estimators which shrink x_{i+} and x_{+j} simultaneously. A discussion is given in subsection 2.5 and simulation results which show the performance of the proposed estimators are given in subsection 2.6.

2 Shrinkage estimators in the multiplicative Poisson model

First, we consider one-dimensional shrinkage to an order statistic.

2.1. One-dimensional shrinkage to an order statistic

Let $x_{(\ell)+}$ be the ℓ-th smallest observation among $x_{1+}, \ldots, x_{I+}$. We assume that $I \geq \ell + 2$ and consider the following estimator which shrinks x_{i+} toward $x_{(\ell)+}$ when $x_{i+} \geq x_{(\ell)+}$:

$$\delta_{ij}^{(1)} = \frac{x_{+j}}{x_{++}}\left\{x_{i+} - \varphi(W)\frac{(x_{i+} - x_{(\ell)+})^+}{W + d}\right\}, \quad i = 1, \ldots, I,\ j = 1, \ldots, J,$$

where $W = \sum_{i=1}^{I}(x_{i+} - x_{(\ell)+})^+$, $a^+ = \max(0, a)$ and d is a positive constant. Then we have the following.

Theorem 2.1. Suppose that $\varphi(W)$ is a non-decreasing function satisfying $0 \leq \varphi(W) \leq 2(I - \ell - 1)$ and that $d \geq \sup \varphi(W)/2$. Then $\delta_{ij}^{(1)}, i = 1, \ldots, I$ improves upon the MLE $\hat{\lambda}_{ij}^{ML}, i = 1, \ldots, I$ under the loss function $\sum_{i=1}^{I}(\hat{\lambda}_{ij} - \lambda_{ij})^2/\lambda_{ij}$ for any $j = 1, \ldots, J$.

Remark 2.1. Theorem 2.1 can be generalized directly to the case of Poisson multiplicative model for a multi-way contingency tables by using a lemma (Lemma 3.1 of Hara and Takemura (2006)) which is a generalization of Lemma 2.1. For example, consider the case of a 3-way contingency table $x_{ijk}, i = 1, \ldots, I,\ j = 1, \ldots, J,\ k = 1, \ldots, K$ where x_{ijk} are independent Poisson random variables with means λ_{ijk}. Let $x_{i++}, x_{+j+}, x_{++k}$ and x_{+++}

denote the one-dimensional marginal frequencies and the total frequency. Let j and k be arbitrarily fixed and consider the simultaneous estimation of $\lambda_{1jk}, \ldots, \lambda_{Ijk}$ under the loss function $\sum_{i=1}^{I}(\hat{\lambda}_{ijk} - \lambda_{ijk})^2/\lambda_{ijk}$. Then, by adopting similar notations and conditions on $\varphi(W)$ and d, we see that the estimator

$$\frac{x_{+j+}x_{++k}}{x_{+++}^2}\left\{x_{i++} - \varphi(W)\frac{(x_{i++} - x_{(\ell)++})^+}{W + d}\right\},\ i = 1, \ldots, I$$

improves upon the MLE $x_{i++}x_{+j+}x_{++k}/x_{+++}^2, i = 1, \ldots, I$.

Next we consider the estimators shrink $\hat{\lambda}_{ij}^{ML}$ to a specified non-negative values, b_i.

2.2. One-dimensional shrinkage to a specified point

Let $b_i \geq 0, i = 1, \ldots, I$ be given numbers and we propose the following shrinkage estimator which shrinks x_{i+} to b_i when $x_{i+} \geq b_i$:

$$\delta_{ij}^{(2)} = \frac{x_{+j}}{x_{++}}\left\{x_{i+} - \varphi(N, W)\frac{(x_{i+} - b_i)^+}{W + d(N)}\right\}, \quad i = 1, \ldots, I,\ j = 1, \ldots, J,$$

where $W = \sum_{i=1}^{I}(x_{i+} - b_i)^+$ and $N = \#\{i | x_{i+} \geq b_i\}$. Then we have the following.

Theorem 2.2, Suppose that $\varphi(N, W)$ is a non-decreasing function of W and satisfies $0 \leq \varphi(N, W) \leq 2(N - 1)^+$ for any $0 \leq N \leq I$. Suppose that $d(N) \geq \sup_W \varphi(N, W)/2$. Then $\delta_{ij}^{(2)}, i = 1, \ldots, I$ improves upon the MLE $\hat{\lambda}_{ij}^{ML}, i = 1, \ldots, I$ under the loss function $\sum_{i=1}^{I}(\hat{\lambda}_{ij} - \lambda_{ij})^2/\lambda_{ij}$ for any $j = 1, \ldots, J$.
It may be noticed that the shrinkage is made only when $N \geq 2$.

Next we consider two-dimensional shrinkage to order statistics or to a specified point.

2.3. Two-dimensional shrinkage to order statistics

Let $x_{(\ell)+}$ and $x_{+(m)}$ be the ℓ-th and m-th smallest observation among $x_{1+}, \ldots, x_{I+}$ and $x_{+1}, \ldots,\ x_{+J}$, respectively. We assume that $I \geq \ell + 2$ and $J \geq m + 2$ and consider the estimator which shrinks x_{i+} toward $x_{(\ell)+}$ when $x_{i+} \geq x_{(\ell)+}$ in the first dimension and shrinks x_{+j} toward $x_{+(m)}$ when $x_{+j} \geq x_{+(m)}$ in the second dimension simultaneously. To improve upon the MLE $\hat{\lambda}_{ij}^{ML}$, we propose the following estimator :

$$\delta_{ij}^{(3)} = \frac{1}{x_{++}}\left\{x_{i+} - \varphi_1(W_1)\frac{(x_{i+} - x_{(\ell)+})^+}{W_1 + d_1}\right\}\left\{x_{+j} - \varphi_2(W_2)\frac{(x_{+j} - x_{+(m)})^+}{W_2 + d_2}\right\},$$
$$i = 1, \ldots, I,\ j = 1, \ldots, J,$$

where $W_1 = \sum_{i=1}^{I}(x_{i+} - x_{(\ell)+})^+$ and $W_2 = \sum_{j=1}^{J}(x_{+j} - x_{+(m)})^+$ and d_1 and d_2 are positive constants. Then we have the following.

Theorem 2.3. Suppose that $\varphi_1(W_1)$ and $\varphi_2(W_2)$ are non-decreasing functions satisfying $0 \leq \varphi_1(W_1) \leq I - \ell - 1$ and $0 \leq \varphi_2(W_2) \leq J - m - 1$, respectively. If $d_1 \geq (I - \ell - 1)/(I - \ell) \sup \varphi_1(W_1)$ and $d_2 \geq (J - m - 1)/(J - m) \sup \varphi_2(W_2)$. Then

$\delta_{ij}^{(3)}, i = 1, \ldots, I, j = 1, \ldots, J$ improves upon the MLE $\hat{\lambda}_{ij}^{ML}$ under the loss function $\sum_{i=1}^{I}\sum_{j=1}^{J}(\hat{\lambda}_{ij} - \lambda_{ij})^2/\lambda_{ij}$.

Remark 2.3. Theorem 2.3 is directly generalized to the case of multi-way contingency tables. Since the notations and conditions are essentially the same, we only give a sketch of the result for the case of 3-way contingency table. We shrink x_{i++} toward $x_{(\ell)++}$ when $x_{i++} \geq x_{(\ell)++}$ in the first dimension and shrink x_{+j+} toward $x_{+(m)+}$ when $x_{+j+} \geq x_{+(m)+}$ in the second dimension. Under the loss function

$$\sum_{i=1}^{I}\sum_{j=1}^{J}(\hat{\lambda}_{ijk} - \lambda_{ijk})^2/\lambda_{ijk},$$

where $k = 1, \cdots, K$ is arbitrarily fixed, the improved estimator is given by

$$\begin{aligned}\delta_{ijk} \;=\; & \frac{x_{++k}}{x_{+++}^2}\left\{x_{i++} - \varphi_1(W_1)\frac{(x_{i++} - x_{(\ell)++})^+}{W_1 + d_1}\right\} \\ & \left\{x_{+j+} - \varphi_2(W_2)\frac{(x_{+j+} - x_{+(m)+})^+}{W_2 + d_2}\right\}, \; i = 1, \ldots, I, \; j = 1, \ldots, J.\end{aligned}$$

2.4. Two-dimensional shrinkage to a specified point

Let $b_i \geq 0, i = 1, \ldots, I$ and $c_j \geq 0, j = 1, \ldots, J$ be given numbers. Assuming that $I, J \geq 2$, we shrink x_{i+} to b_i when $x_{i+} \geq b_i$ and x_{+j} to c_j when $x_{+j} \geq c_j$. To improve upon the MLE $\hat{\lambda}_{ij}^{ML}$, we propose the following estimator

$$\delta_{ij}^{(4)} = \frac{1}{x_{++}}\left\{x_{i+} - \varphi_1(N_1, W_1)\frac{(x_{i+} - b_i)^+}{W_1 + d_1(N_1)}\right\}\left\{x_{+j} - \varphi_2(N_2, W_2)\frac{(x_{+j} - c_j)^+}{W_2 + d_2(N_2)}\right\}, \quad i = 1, \ldots, I, \; j = 1, \ldots, J,$$

where $W_1 = \sum_{i=1}^{I}(x_{i+} - b_i)^+, W_2 = \sum_{j=1}^{J}(x_{+j} - c_j)^+,\ N_1 = \#\{i | x_{i+} \geq b_i, i = 1, \ldots, I\}$ and $N_2 = \#\{j | x_{+j} \geq c_i, j = 1, \ldots, J\}$. Although it may be natural to put the condition $\sum_{i=1}^{I} b_i = \sum_{j=1}^{J} c_j$, we do not need it in the following.

Theorem 2.4. Suppose that $\varphi_\ell(N_\ell, W_\ell)$ is a non-decreasing function of W_ℓ and satisfies $0 \leq \varphi_\ell(N_\ell, W_\ell) \leq (N_\ell - 1)^+$ for any $N_\ell \geq 0$, and that $d_\ell(N_\ell) \geq (N_\ell - 1)^+/N_\ell \sup_{W_\ell} \varphi_\ell(N_\ell, W_\ell)$, for any $N_\ell \geq 0$, $\ell = 1, 2$. Then $\delta_{ij}^{(4)}$ improves upon the MLE $\hat{\lambda}_{ij}^{ML}$ under the loss function $\sum_{i=1}^{I}\sum_{j=1}^{J}(\hat{\lambda}_{ij} - \lambda_{ij})^2/\lambda_{ij}$.

It may be noticed that the shrinkage in the ℓ-th dimension is made only when $N_\ell \geq 2$.

2.5. A discussion

Here we mention the possibility of the two-dimensional shrinkage estimators other than $\delta_{ij}^{(3)}$ and $\delta_{ij}^{(4)}$ given in subsections 2.3 and 2.4, respectively. We only give two alternative estimators for $\delta_{ij}^{(4)}$. The following estimator is the simple average of the one-dimensional shrinkage estimator $\delta_{ij}^{(2)}$ and its counterpart which makes shrinkage in the second dimension:

$$\frac{x_{i+}x_{+j}}{x_{++}} - \frac{\varphi_1(N_1, W_1)}{2} \frac{x_{+j}}{x_{++}} \frac{(x_{i+} - b_i)^+}{W_1 + d_1(N_1)} - \frac{\varphi_2(N_2, W_2)}{2} \frac{x_{i+}}{x_{++}} \frac{(x_{+j} - c_j)^+}{W_2 + d_2(N_2)},$$

where W_i and $N_i, i = 1, 2$, are defined in 2.4. It is easily shown that this estimator improves upon the MLE when $\varphi(N_i, W_i)$ and $d_i(N_i)$, $i = 1, 2$, satisfy the similar conditions as given in Theorem 2.2.

We may pool W_1 and W_2 and consider the following estimator

$$\frac{x_{i+}x_{+j}}{x_{++}} - \frac{\varphi(N, W)}{2} \frac{x_{+j}(x_{i+} - b_i)^+ + x_{i+}(x_{+j} - c_j)^+}{x_{++}\{W + d(N)\}},$$

where $W = (W_1 + W_2)/2$ and $N = N_1 + N_2$. Although this estimator will dominate the MLE under suitable conditions on $\varphi(N, W)$ and $d(N)$, we do not pursue it here further.

Unfortunately, these two estimators do not give the estimates which belong to the parameter space of the multiplicative Poisson models, whereas the estimators $\delta_{ij}^{(3)}$ and $\delta_{ij}^{(4)}$ do.

2.6. Simulation study

Through simulation study we evaluate the risk performance of the following estimators:

$$\delta_{ij}^{0} = \frac{x_{i+}x_{+j}}{x_{++}} \left\{ 1 - \frac{d}{x_{++} + d} \right\}, \quad i = 1, \ldots, I, \ j = 1, \ldots, J,$$

$$\delta_{ij}^{min} = \frac{x_{+j}}{x_{++}} \left\{ x_{i+} - d \frac{(x_{i+} - x_{(1)+})^+}{\sum_{k=1}^{I}(x_{k+} - x_{(1)+}) + d} \right\}, \quad i = 1, \ldots, I, \ j = 1, \ldots, J,$$

$$\delta_{ij}^{b} = \frac{x_{+j}}{x_{++}} \left\{ x_{i+} - d(N) \frac{(x_{i+} - b_i)^+}{\sum_{k=1}^{I}(x_{k+} - b_k)^+ + d(N)} \right\}, \quad i = 1, \ldots, I, \ j = 1, \ldots, J.$$

For an estimator δ_{ij} of λ_{ij}, $i = 1, \ldots, I, j = 1, \ldots J$, let $\delta_j = (\delta_{1j}, \ldots, \delta_{Ij}),\ j = 1, \ldots, J$ and $\delta = (\delta_1, \ldots, \delta_J)$. When we estimate λ_{ij} by δ_{ij}, the expected loss is given by

$$E\left\{ \frac{1}{\lambda_{ij}} (\delta_{ij} - \lambda_{ij})^2 \right\} \equiv R(\boldsymbol{\lambda}, \delta_{ij}),$$

where $\boldsymbol{\lambda} = \{\lambda_{ij}, i = 1, \ldots, I, j = 1, \ldots, J\}$. The total risk is

$$R(\boldsymbol{\lambda}, \delta) = \sum_{i=1}^{I} \sum_{j=1}^{J} R(\boldsymbol{\lambda}, \delta_{ij}).$$

We are also concerned with the columnwise risks

$$R(\boldsymbol{\lambda}, \delta_j) = \sum_{i=1}^{I} R(\boldsymbol{\lambda}, \delta_{ij}), \qquad j = 1, \ldots, J.$$

Before describing the details of our simulation study, we state some fundamental properties of the total risk function. We consider a class of estimators δ_{ij} of the form $\psi_i(\boldsymbol{x}_1)x_{+j}/x_{++}, i = 1, \ldots, I, j = 1, \ldots, J$, where $\boldsymbol{x}_1 = (x_{1+}, \ldots, x_{I+})$. We notice that all the estimators considered here are of this form. Applying Lemma 1.1, we have

$$R(\boldsymbol{\lambda}, \delta_{ij}) = E\left\{\psi_i^2(\boldsymbol{x}_1 + \boldsymbol{e}_i^I)\frac{x_{+j}+1}{(x_{i+}+1)(x_{++}+1)} - 2\psi_i(\boldsymbol{x}_1)\frac{x_{+j}}{x_{++}} + \lambda_{ij}\right\},$$

where $\boldsymbol{e}_i^I$ is the $I \times 1$ unit vector with i-th component 1. Thus we have

$$\sum_{j=1}^{J} R(\boldsymbol{\lambda}, \delta_{ij}) = E\left\{\psi_i^2(\boldsymbol{x}_1 + \mathbf{e}_i^I)\frac{x_{++}+J}{(x_{i+}+1)(x_{++}+1)} - 2\psi_i(\boldsymbol{x}_1) + \lambda_{i+}\right\},$$

which is a function of $(\lambda_{1+}, \ldots, \lambda_{I+})$ and is independent of $(\beta_1, \ldots, \beta_J)$. Therefore we see that the total risk $R(\boldsymbol{\lambda}, \delta)$ is independent of $(\beta_1, \ldots, \beta_J)$. Further, if $\psi_i(\boldsymbol{x}_1)$ is of the form $\phi(x_{++})x_{i+}$, we have

$$\sum_{i=1}^{I}\sum_{j=1}^{J} R(\boldsymbol{\lambda}, \delta_{ij}) = E\left\{\phi^2(x_{++}+1)\frac{(x_{++}+I)(x_{++}+J)}{x_{++}+1} - 2\phi(x_{++})x_{++} + \lambda\right\}.$$

Therefore the total risk depends only on $\lambda = \lambda_{++}$. We may notice here that the total risk of $\hat{\lambda}^{ML}$ ($\phi(x_{++} = 1)$ is expressed as

$$R(\boldsymbol{\lambda}, \hat{\lambda}^{ML}) = (I + J - 1) + (I-1)(J-1)(1 - e^{-\lambda})/\lambda,$$

which is a decreasing function of λ.

Now we describe the details of the estimators used in our simulation study. We only consider the case $I = 5$ and $J = 4$. The constant d in δ^{min} is chosen as $2(5-2) = 6$, which is an upper bound of $\varphi(W)$ given in Theorem 2.1. Although simulation experiments are performed also for the case $d = 3$, the performance for the case $d = 6$ is generally better. Similarly, $d(N)$ in δ^b is chosen as $2(N-1)^+$. The constant d in δ^0 is set $d = 2(5+4-2) = 14$, since it is of interest and does not show largely different performance compared with the choice of $d = 7$. We have tried two choices of b in δ^b:

$$b_i^{(1)} = \lambda_{i+} \text{ and } b_i^{(2)} = \max(0, \lambda_{i+} - \sqrt{\lambda_{i+}}), \quad i = 1, \ldots, 5.$$

Although these choices use the true values of λ_{i+}'s, it will make clear how much improvement is possible by choosing δ^b pertinently.

To evaluate the risk improvement we use the (total) relative improvement rate

$$\{R(\boldsymbol{\lambda}, \hat{\lambda}^{ML}) - R(\boldsymbol{\lambda}, \delta)\}/R(\boldsymbol{\lambda}, \hat{\lambda}^{ML}),$$

and also the columnwise relative improvement rate

$$\{R(\boldsymbol{\lambda}, \hat{\lambda}_j^{ML}) - R(\boldsymbol{\lambda}, \delta_j)\}/R(\boldsymbol{\lambda}, \hat{\lambda}_j^{ML}), \quad j = 1, 2, 3, 4.$$

In Tables 2.1 and 2.2, based on 100,000 replications, the total relative improvement rates and the columnwise relative improvement rates are given respectively for the four estimators.

Putting $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5)$, $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3, \beta_4)$ and

$$\boldsymbol{\alpha}^{(1)} = (0.2, 0.2, 0.2, 0.2, 0.2),\ \boldsymbol{\beta}^{(1)} = (0.25, 0.25, 0.25, 0.25),$$

$$\boldsymbol{\alpha}^{(2)} = (0.3, 0.2, 0.2, 0.15, 0.15),\ \boldsymbol{\beta}^{(2)} = (0.4, 0.3, 0.15, 0.15),$$

$$\boldsymbol{\alpha}^{(3)} = (0.4, 0.3, 0.1, 0.1, 0.1),\ \boldsymbol{\beta}^{(3)} = (0.6, 0.2, 0.1, 0.1),$$

9 multiplicative models are considered in Table 2.1, where $\lambda = 10, 50, 100$ and $\boldsymbol{\alpha}=\boldsymbol{\alpha}^{(1)}, \boldsymbol{\alpha}^{(2)}, \boldsymbol{\alpha}^{(3)} (\boldsymbol{\beta} = \boldsymbol{\beta}^{(1)}$ is fixed for convenience). In Table 2.2, 9 multiplicative models are considered, where $\lambda = 10, 50, 100$ and

$$(\boldsymbol{\alpha}, \boldsymbol{\beta}) = (\boldsymbol{\alpha}^{(1)}, \boldsymbol{\beta}^{(2)}), (\boldsymbol{\alpha}^{(1)}, \boldsymbol{\beta}^{(3)}), (\boldsymbol{\alpha}^{(3)}, \boldsymbol{\beta}^{(2)}).$$

We summarize the simulation results as follows.

1. When we are concerned with the total risk, the relative improvement rate of δ^{min} is larger than that of δ^0 for the case where $\alpha_i = 0.2, i = 1, 2, \ldots, 5$ which seems to be the most favorable case for δ^{min}. Although the relative improvement rate of δ^{min} gets smaller when the values of α_i's are more imbalanced, especially for larger values of λ, it is still nearly identical to that of δ^0.
2. The relative improvement rate of $\delta^{b^{(2)}}$ is the largest among those of all the estimators for all cases. This implies that if we choose $b_i, i = 1, \ldots, I$ pertinently based on the prior information on $\lambda_{i+}, i = 1, \ldots, I$, then we are able to get large improvement. However, the choice $b_i^{(1)} = \lambda_{i+}, i = 1, \ldots, I$ is not pertinent and choosing smaller values such as $\max(0, \lambda_{i+} - \sqrt{\lambda_{i+}}), i = 1, \ldots, I$ is preferable.
3. The value of λ largely affects the performance of not only δ^0 but also δ^{min} and δ^b. When the value of λ gets larger, the improvement rate of the estimators gets smaller.
4. When $\beta_1 = \cdots = \beta_J$, the columnwise risks are all the same and the columnwise relative improvement rates are equal to the total relative improvement rate for any estimator. The imbalance among the values of β_j's affects the columnwise risk performance of δ^0 the most as we see from Tables 2.2. For the columns with relatively larger values of β_j, the columnwise relative improvement rates of δ^0 get smaller and are sometimes negative. This implies that δ^0 will not dominate the MLE in terms of the columnwise risks for large values of d.

Table 2.1 Total risks and relative improvement rates

$\boldsymbol{\alpha}$	Estimators	$d/d(N)$	Total risk			$1-R(\boldsymbol{\lambda},\delta)/R(\boldsymbol{\lambda},\hat{\lambda}^{ML})$		
			$\lambda=10$	$\lambda=50$	$\lambda=100$	$\lambda=10$	$\lambda=50$	$\lambda=100$
	$\hat{\lambda}^{ML}$		9.181	8.237	8.123			
	δ^0	14	5.219	7.721	7.966	0.432	0.063	0.019
$\boldsymbol{\alpha}^{(1)}$	δ^{min}	6	5.055	6.370	6.752	0.449	0.227	0.169
	$\delta^{b^{(1)}}$	$2(N-1)^+$	5.816	6.630	6.934	0.367	0.195	0.146
	$\delta^{b^{(2)}}$	$2(N-1)^+$	4.898	6.168	6.620	0.466	0.251	0.185
	$\hat{\lambda}^{ML}$		9.174	8.238	8.121			
	δ^0	14	5.217	7.720	7.963	0.431	0.063	0.019
$\boldsymbol{\alpha}^{(2)}$	δ^{min}	6	5.161	6.747	7.223	0.437	0.181	0.111
	$\delta^{b^{(1)}}$	$2(N-1)^+$	5.988	6.688	6.922	0.347	0.188	0.148
	$\delta^{b^{(2)}}$	$2(N-1)^+$	4.970	6.180	6.626	0.458	0.250	0.184
	$\hat{\lambda}^{ML}$		9.180	8.258	8.107			
	δ^0	14	5.222	7.735	7.951	0.431	0.063	0.019
$\boldsymbol{\alpha}^{(3)}$	δ^{min}	6	5.429	7.249	7.577	0.409	0.122	0.065
	$\delta^{b^{(1)}}$	$2(N-1)^+$	5.675	6.571	6.853	0.382	0.204	0.155
	$\delta^{b^{(2)}}$	$2(N-1)^+$	4.598	6.165	6.540	0.499	0.253	0.193

Table 2.2 Columnwise relative improvements rates

$\boldsymbol{\alpha}$, $\boldsymbol{\beta}$	λ	Estimator	1st column	2nd column	3rd column	4th column	total risk
$\boldsymbol{\alpha}=\boldsymbol{\alpha}^{(1)}$ $\boldsymbol{\beta}=\boldsymbol{\beta}^{(2)}$	10	δ^0	0.316	0.388	0.541	0.544	0.431
		δ^{min}	0.415	0.436	0.482	0.483	0.449
		$\delta^{b^{(1)}}$	0.387	0.375	0.347	0.348	0.367
		$\delta^{b^{(2)}}$	0.431	0.452	0.500	0.502	0.466
	50	δ^0	-0.020	0.031	0.147	0.150	0.063
		δ^{min}	0.237	0.230	0.216	0.216	0.226
		$\delta^{b^{(1)}}$	0.220	0.205	0.168	0.169	0.195
		$\delta^{b^{(2)}}$	0.264	0.257	0.237	0.237	0.251
	100	δ^0	-0.033	-0.001	0.073	0.075	0.019
		δ^{min}	0.185	0.175	0.152	0.152	0.169
		$\delta^{b^{(1)}}$	0.169	0.155	0.122	0.122	0.146
		$\delta^{b^{(2)}}$	0.204	0.193	0.165	0.165	0.185
$\boldsymbol{\alpha}=\boldsymbol{\alpha}^{(1)}$ $\boldsymbol{\beta}=\boldsymbol{\beta}^{(3)}$	10	δ^0	0.218	0.480	0.615	0.616	0.431
		δ^{min}	0.386	0.463	0.503	0.504	0.449
		$\delta^{b_i^{(1)}}$	0.403	0.358	0.336	0.336	0.367
		$\delta^{b^{(2)}}$	0.400	0.480	0.523	0.524	0.466
	50	δ^0	-0.083	0.100	0.205	0.208	0.062
		δ^{min}	0.246	0.221	0.207	0.208	0.226
		$\delta^{b^{(1)}}$	0.241	0.183	0.149	0.150	0.195
		$\delta^{b^{(2)}}$	0.276	0.244	0.226	0.227	0.251
	100	δ^0	-0.069	0.044	0.111	0.111	0.020
		δ^{min}	0.199	0.161	0.139	0.140	0.169
		$\delta^{b^{(1)}}$	0.187	0.135	0.104	0.105	0.146
		$\delta^{b^{(2)}}$	0.220	0.176	0.150	0.151	0.185
$\boldsymbol{\alpha}=\boldsymbol{\alpha}^{(3)}$ $\boldsymbol{\beta}=\boldsymbol{\beta}^{(2)}$	10	δ^0	0.316	0.386	0.543	0.545	0.431
		δ^{min}	0.359	0.389	0.457	0.459	0.409
		$\delta^{b^{(1)}}$	0.404	0.390	0.360	0.360	0.382
		$\delta^{b^{(2)}}$	0.466	0.486	0.531	0.533	0.499
	50	δ^0	-0.020	0.030	0.149	0.150	0.062
		δ^{min}	0.103	0.114	0.142	0.142	0.122
		$\delta^{b^{(1)}}$	0.231	0.216	0.175	0.176	0.204
		$\delta^{b^{(2)}}$	0.267	0.259	0.238	0.238	0.253
	100	δ^0	-0.030	-0.001	0.072	0.074	0.020
		δ^{min}	0.056	0.061	0.075	0.076	0.065
		$\delta^{b^{(1)}}$	0.179	0.164	0.128	0.128	0.154
		$\delta^{b^{(2)}}$	0.214	0.200	0.170	0.171	0.193

References

[1] Chang, Y.-T. and Shinozaki, N. (2006). Estimation of ordered means of two Poisson distributions. *Comm. Statist.-Theory and Methods-*, 35, 1993-2003.

[2] Chang, Y.-T. and Shinozaki, N. (2019). New types of shrinkage estimators of Poisson means under the normalized squared error loss. *Comm. Statist.-Theory and Methods-*, 48, 1108-1122.

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

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

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

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

[7] Hamura, Y. and Kubokawa, T. (2019). Simultaneous estimation of parameters of Poisson distributions with unbalanced sample sizes. *Japanese J. Statist. and Data Analysis*, https://doi.org/10.1007/s42081-019-00039-x.

[8] Hara, H. and Takemura, A. (2006). Simultaneous estimation of the means in some Poisson log linear models. *J. Japan Statist. Soc.*, 36, 17-36.

[9] 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.

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

[11] Komaki, F. (2015). Simultaneous prediction for independent Poisson process with different durations. *J. Multivariate Anal.*, 141, 35-48.

[12] Peng, J. C. M. (1975). Simultaneous estimation of the parameters of independent Poisson distributions. Technical Reports 78. Department of Statistics, Stanford University.

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

[14] Tsui, K. W. (1986). Further developments on the robustness of Clevenson-Zidek-type means estimators. *J. Amer. Statist. Assoc.*, 81, 176-180.

[15] 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.

Acknowledgement: This work was supported by the Research Institute for Mathematical Sciences, an International Joint Usage/Research Center located in Kyoto University. This work is salso upported by Grant-in-Aid for Scientific Research (C) No. 18K11196. Two authors also thank editor in chief Professor Koike Ken-ichi, Nihon University for his many supports.

*Chang Yuan-Tsung
Mejiro University
Tokyo 161-8539
JAPAN
E-mail address: chogenso@gmail.com

**Shinozaki Nobuo
Keio University
Kanagawa 223-8522
JAPAN