A shrinkage method to the multiplicative model when Poisson random variables are observed in the form of two-way contingency tables

*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, introduce the estimators which shrink to the specified values or an order statistic in one dimension and in two dimensions, proposed by Chang and Shinozaki (2022), are considered and are shown to dominate the maximum likelihood estimator (MLE) under normalized squared error loss. Further, Assuming the full model, shrinkage to the multiplicative model is devised to improve upon the unbiased estimator by finding out the patterns where the observed frequency is not smaller than the estimated frequency for each cell.

1 Introduction

We consider 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.

To review the history of the simultaneous estimation of Poisson means briefly, let x_i be independently distributed as $Po(\lambda_i), i = 1, \ldots, p$. Clevenson and Zidek (1975) were the first to propose a class of estimators of the form

$$\hat{\lambda}_i^{CZ}(\boldsymbol{x}) = x_i - \frac{\varphi(W)}{W + p - 1} x_i, \qquad i = 1, \ldots, p,$$

where $W = \sum_i^p x_i$ and $\boldsymbol{x} = (x_1, \ldots, x_p)$. Clevenson and Zidek (1975) have shown that when $p \geq 2$ and if $\varphi(\cdot)$ is non-decreasing and satisfies $0 \leq \varphi(\cdot) \leq 2(p-1)$, then $\hat{\boldsymbol{\lambda}}^{CZ}(\boldsymbol{x}) = (\hat{\lambda}_1^{CZ}(\boldsymbol{x}), \ldots, \hat{\lambda}_p^{CZ}(\boldsymbol{x}))$ dominates $\boldsymbol{x}$ under the normalized squared error loss

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

Since then considerable efforts have been devoted to the problem by many authors. Broad classes of dominating estimators have been given by many authors, including Tsui and Press (1982), Hwang (1982), Ghosh et al. (1983) and Chou (1991). Dominance results have been shown for the other loss functions, including the squared error loss one $\sum_{i=1}^{p}(\hat{\lambda}_i - \lambda_i)^2$. See, for example, Peng (1975), Tsui and Press (1982), Ghosh et al. (1983) and Ghosh and Yang (1988). Improved estimators over the unbiased one have been given for a class of discrete distributions, including those belonging to a one-parameter exponential family. See, Hwang (1982), Ghosh et al. (1983), Tsui (1984, 1986) and Chou (1991). Chang and Shinozaki (2019) have given new types of shrinkage estimators when a covariate is available and also when order restrictions on the means are present. Simultaneous prediction of Poisson random variable has been treated by Komaki (2004), who has given fundamental admissibility results under the Kullback-Leibler loss. A class of generalized Bayes estimators has been introduced in Komaki (2015) and Hamura and Kubokawa (2019) have derived the condition under which the generalized Bayes estimators dominate the standard estimators when sample sizes are different.

One important case to which simultaneous estimation of Poisson means is applied is the one where Poisson random variables are observed in the form of a multi-way contingency table. Hara and Takemura (2006) have studied the simultaneous estimation of the Poisson means in multi-way multiplicative models and have given a class of estimators which improve upon the maximum likelihood estimators (MLE) under the normalized squared error loss by shrinking them toward the origin. Motived by Hara and Takemura, Chang and Shinozaki (2022) have studied the simultaneous estimation of the Poisson means in multi-way multiplicative models and have proposed the 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.

In this paper two problems on improving upon the MLE of the Poisson means is considered when Poisson random variables are observed in the form of a two-way contingency table. In Section 2, we first introduce the some results of Chang and Shinozaki (2022), the 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 Section 3, assuming the full model, shrinkage to the multiplicative model is devised by determining the basic cells so that the observed frequency is not smaller than the estimated frequency for each of the other cells. The detail proofs of the basic results used in the discussion of Section 3 are given in Appendix in Chang and Shinozaki (2022), https://doi.org/10.1007/s42081-022-00169-9.

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

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

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.

2 One-dimensional shrinkage to an order statistic or a specified point

2.1. One-dimensional shrinkage to an order or a specified point statistic.

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

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

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

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 $x_{ijk}, i = 1, \dots, I,\ j = 1, \dots, J,\ k = 1, \dots, K$ 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 $\lambda_{1jk}, \dots, \lambda_{Ijk}$ 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

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

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

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 $N \geq 2$.

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 :

$$\delta_{ij}^{(3)} = \frac{1}{x_{++}}\left\{x_{i+} - \varphi_1(W_1)\frac{(x_{i+}-x_{(\ell)+})^+}{W_1+d_1}\right\}\left\{x_{+j} - \varphi_2(W_2)\frac{(x_{+j}-x_{+(m)})^+}{W_2+d_2}\right\},$$
$$i = 1, \ldots, I,\ j = 1, \ldots, J, \qquad (2.4)$$

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

We also consider two-dimensional shrinkage to the order statistics and to the specified two positive values.
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

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

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

$$\delta_{ij}^{(4)} = \frac{1}{x_{++}}\Big\{x_{i+} - \varphi_1(N_1, W_1)\frac{(x_{i+} - b_i)^+}{W_1 + d_1(N_1)}\Big\}\Big\{x_{+j} - \varphi_2(N_2, W_2)\frac{(x_{+j} - c_j)^+}{W_2 + d_2(N_2)}\Big\},$$
$$i = 1, \ldots, I,\ j = 1, \ldots, J, \quad (2.5)$$

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 $\varphi_i(N_i, W_i)$ is a non-decreasing function of W_i and satisfies $0 \leq \varphi_i(N_i, W_i) \leq (N_i - 1)^+$ for any $N_i \geq 0$, and that $d_i(N_i) \geq (N_i - 1)^+ / N_i \sup_{W_i} \varphi_i(N_i, W_i)$, for any $N_i \geq 0$, $i = 1, 2$. 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 $i-$th dimension is made only when $N_i \geq 2$.

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.

3 Shrinkage to the multiplicative Poisson model

Here we consider saturated (full) model and propose a shrinkage method to the multiplicative model to improve upon the unbiased estimator. In 3.1 we deal with the 2×3 table case to explain the idea of the method. In 3.2 general two-way contingency table is treated. A numerical example is given in 3.3 and a discussion is given in 3.4. Although the numbers of rows and columns are denoted by I and J in Section 2, here we denote them by m and n for simplicity.

Now we state a useful result due to Chang and Shinozaki (2019). Let x_i be distributed as $Po(\lambda_i)$, $i = 1, \dots, p$, and suppose that $x_1, \dots, x_p$ are statistically independent. Let $b_i, i = 1, \dots, p$, be specified non-negative values and let $C = \{(x_1, \dots, x_p) | x_i \geq b_i, i = 1, \dots, p\}$. We consider a class of estimators which shrink only when $\boldsymbol{x} = (x_1, \dots, x_p) \in C$. Letting I_C be the indicator function of C, estimators of the following form are considered:

$$\delta_i(\boldsymbol{x}) = x_i - \varphi(W) \frac{(x_i - b_i)}{W + d} I_C, \quad i = 1, \dots, p, \tag{3.1}$$

where $W = \sum_{i=1}^{p} (x_i - b_i)$ and d is a positive constant. For $p \geq 2$, Chang and Shinozaki (2019) have shown the following.

Lemma 3.1. Let $\varphi(\cdot)$ be a non-decreasing function which satisfies $0 \leq \varphi(\cdot) \leq 2(p-1)$ and suppose that $d \geq \sup \varphi(\cdot)/2$. Then $(\delta_1(\boldsymbol{x}), \dots, \delta_p(\boldsymbol{x}))$ dominates $\boldsymbol{x}$ under the normalized squared error loss $\sum_{i=1}^{p} (\hat{\lambda}_i - \lambda_i)^2 / \lambda_i$.

Remark 3.1. Since the two estimators are the same outside C, the averaged loss of $(\delta_1(\boldsymbol{x}), \dots, \delta_p(\boldsymbol{x}))$ over C is smaller than or equal to that of $\boldsymbol{X}$. Further, as stated in Chang and Shinozaki (2019), Lemma 3.1 is true even when the inequality $x_i \geq b_i$ is replaced by $x_i > b_i$ for some of p coordinates in the definition of C. We use this Remark 3.1 in the subsection 3.2.4.

3.1 2×3 table

Consider a 2×3 table whose components $x_{ij}, i = 1, 2, j = 1, 2, 3$ are independent Poisson random variables with respective means λ_{ij}. The multiplicative (independent) model is described as

$$\lambda_{ij} = \lambda p_i q_j, \quad i = 1, 2, \quad j = 1, 2, 3,$$

where $\lambda = \sum_{i=1}^{2} \sum_{j=1}^{3} \lambda_{ij}$ and $p_i \geq 0$ and $q_j \geq 0$ satisfy $p_1 + p_2 = 1$ and $q_1 + q_2 + q_3 = 1$, respectively. When the model is true, the row ratio $x_{i_1 j}/x_{i_2 j}$ $(j = 1, 2, 3)$ is an estimator of p_{i_1}/p_{i_2}, $(i_1, i_2 = 1, 2)$ and the column ratio x_{ij_1}/x_{ij_2} $(i = 1, 2)$ is an estimator of q_{j_1}/q_{j_2}, $(j_1, j_2 = 1, 2, 3)$. If we choose four x_{ij}'s pertinently so that a row ratio and two column ratios are determined, we obtain the estimated frequencies of the remaining two cells under independence. In case when the observed frequency is larger than or equal to the estimated frequency for the two cells, we shrink the two observed frequencies to their respective estimated frequencies. For any 2×3 table there are three ways to choose four

x_{ij}'s if we take notice of the numbers of four x_{ij}'s which belong to respective columns: $\mathrm{col}(2,1,1), \mathrm{col}(1,2,1)$ and $\mathrm{col}(1,1,2)$. By $\mathrm{col}(y_1, y_2, y_3)(y_j \geq 1, j = 1,2,3, y_1+y_2+y_3 = 4)$ we mean the case where y_j elements are chosen from the j-th column so that a row ratio and two column ratios are determined.

Col$(2,1,1)$. We first give a partition of the total set $S = \{X | x_{ij} \geq 0, i = 1,2, j = 1,2,3\}$, where $X = \{x_{ij}, i = 1,2, j = 1,2,3\}$. For that purpose we first choose the two variables x_{11} and x_{21} in the first column and the row ratio x_{11}/x_{21} is determinated. Next we choose one variable each from the second and third columns. There are four cases depending on whether $x_{11}/x_{21} \geq x_{12}/x_{22}$ or not and whether $x_{11}/x_{21} \geq x_{13}/x_{23}$ or not. Let the four sets $S_\ell, \ell = 1,2,3,4$ be defined as follows:

$$\begin{aligned} S_1 &= \{X | x_{11}/x_{21} \geq x_{12}/x_{22},\ x_{11}/x_{21} \geq x_{13}/x_{23}\}, \\ S_2 &= \{X | x_{11}/x_{21} \geq x_{12}/x_{22},\ x_{11}/x_{21} < x_{13}/x_{23}\}, \\ S_3 &= \{X | x_{11}/x_{21} < x_{12}/x_{22},\ x_{11}/x_{21} \geq x_{13}/x_{23}\}, \\ S_4 &= \{X | x_{11}/x_{21} < x_{12}/x_{22},\ x_{11}/x_{21} < x_{13}/x_{23}\}. \end{aligned}$$

Then $S_\ell, \ell = 1,2,3,4$ are disjoint and $\bigcup_{\ell=1}^4 S_\ell$ is the total set S. Thus $S_\ell, \ell = 1,2,3,4$ give a partition of S.

Consider the case where an observation $X \in S_4$. We choose the variables x_{22} and x_{23} from the second and third columns, respectively whenever $X \in S_4$. Then the estimated frequencies of the $(1,2)$ and $(1,3)$ cells based on x_{11}, x_{21}, x_{22} and x_{23} are given as

$$\hat{x}_{12} = x_{22} \times (x_{11}/x_{21}) \text{ and } \hat{x}_{13} = x_{23} \times (x_{11}/x_{21}),$$

respectively and we have $x_{12} > \hat{x}_{12}$ and $x_{13} > \hat{x}_{13}$. Suppose that

$$\begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \end{bmatrix} = \begin{bmatrix} 4 & 6 & 8 \\ 8 & 6 & 4 \end{bmatrix}$$

is observed. Then $X \in S_4$ and, fixing $x_{11} = 4$, $x_{21} = 8$, $x_{22} = 6$, and $x_{23} = 4$, we have $\hat{x}_{12} = 3$ and $\hat{x}_{13} = 2$. Thus we have an observation in the two dimensional set $x_{12} > 3$ and $x_{13} > 2$. We apply the estimator (3.1) with $p = 2$, $x_1 = x_{12}$, $x_2 = x_{13}$, $b_1 = 3$ and $b_2 = 2$ and have the following estimator: When $X \in S_4$

$$\psi_{ij}^{(1)}(X) = \begin{cases} x_{ij}, & (i,j) = (1,1),(2,1),(2,2) \text{ and } (2,3), \\ x_{ij} - \dfrac{a(x_{ij} - \hat{x}_{ij})}{(x_{12} - \hat{x}_{12}) + (x_{13} - \hat{x}_{13}) + d}, & (i,j) = (1,2) \text{ and } (1,3), \end{cases}$$

where $0 < a \leq 2$ and $d \geq a/2$. When $X \in S_\ell, \ell = 1,2,3$, $\psi_{ij}^{(1)}(X)$ is similarly defined and the estimator for the case col$(2,1,1)$ is given as

$$\Psi^{(1)}(X) = \{\psi_{ij}^{(1)}(X), i = 1,2,\ j = 1,2,3\},\ X \in S.$$

We will show that the estimator improves upon the unbiased estimator generally in 3.2. In our numerical example, putting $a = 2 - 1 = 1$ and $d = a/2 = 1/2$, we have

$$\Psi^{(1)}(X) = \begin{bmatrix} 4 & 5.684 & 7.368 \\ 8 & 6 & 4 \end{bmatrix}.$$

Case (1,2,1). We choose $x_{12} = 6$ and $x_{22} = 6$ in the second column and the row ratio $x_{12}/x_{22} = 6/6 = 1$ is determined. Choosing $x_{11} = 4$ and $x_{23} = 4$ further, we have $x_{21} = 8 > 4 = \hat{x}_{21}$ and $x_{13} = 8 > 4 = \hat{x}_{13}$. Thus we shrink (x_{21}, x_{13}) to $(\hat{x}_{21}, \hat{x}_{13})$ in our example. Generally we obtain the estimator in this way and denote it by $\psi_{ij}^{(2)}(X)$, as

$$\psi^{(2)}(X) = \begin{bmatrix} 4 & 6 & 7.529 \\ 7.529 & 6 & 4 \end{bmatrix}.$$

Case (1,1,2). We fix $x_{13} = 8$ and $x_{23} = 4$ in the third column and $x_{11} = 4$ and $x_{12} = 6$ further. Then we have $x_{21} = 8 > 2 = \hat{x}_{21}$ and $x_{22} = 6 > 3 = \hat{x}_{22}$. We shrink (x_{21}, x_{22}) to $(\hat{x}_{21}, \hat{x}_{22})$ in our example. Generally we denote the estimator by $\psi_{ij}^{(3)}(X)$, as

$$\psi^{(3)}(X) = \begin{bmatrix} 4 & 6 & 8 \\ 7.368 & 5.684 & 4 \end{bmatrix}.$$

By averaging the three estimators we have

$$\psi_{ij}(X) = \frac{1}{3}\left\{\psi_{ij}^{(1)}(X) + \psi_{ij}^{(2)}(X) + \psi_{ij}^{(3)}(X)\right\}, \quad i = 1, 2, \; j = 1, 2, 3,$$

which is expected to show more stable performance than $\psi_{ij}^{(k)}(X)$ $(k = 1, 2, 3)$ alone. It is easily seen that $\Psi(X) = \{\psi_{ij}(X), i = 1, 2, j = 1, 2, 3\}$ gives an improvement upon X since each $\Psi^{(k)}(X) = \{\psi_{ij}^{(k)}(X), i = 1, 2, j = 1, 2, 3\}$ $(k = 1, 2, 3)$ improves upon X and the randomized estimator

$$\hat{\lambda}_{ij}(X) = \psi_{ij}^{(k)}(X), \quad \text{with probability } \frac{1}{3}, \quad k = 1, 2, 3, \quad i = 1, 2, \; j = 1, 2, 3,$$

is improved upon by $\Psi(X)$ because the loss function is convex. In our numerical example, by putting $a = 2 - 1 = 1$ and $d = a/2 = 1/2$, we have

$$\Psi(X) = \begin{bmatrix} 4 & 5.895 & 7.633 \\ 7.633 & 5.895 & 4 \end{bmatrix}.$$

3.2 $m \times n$ table

Consider an $m \times n$ table whose (i, j)-th element is $x_{ij}, i = 1, \ldots, m, \; j = 1, \ldots, n$, where x_{ij}'s are independent Poisson random variables with respective means λ_{ij}. We denote the table by $X = \{x_{ij}, i = 1, \ldots, m, j = 1, \ldots, n\}$. In the independent (multiplicative) model

$$\lambda_{ij} = \lambda p_i q_j,$$

where $\lambda = \sum_{i=1}^m \sum_{j=1}^n \lambda_{ij}$ and $p_j \geq 0$ and $q_j \geq 0$ satisfy $\sum_{i=1}^m p_i = 1$ and $\sum_{j=1}^n q_j = 1$, respectively. We consider shrinking the observed frequencies to the estimated frequencies under independence.

3.2.1 Connectedness

The row ratio x_{i_1j}/x_{i_2j} $(j = 1, \ldots, n)$ is an estimator of p_{i_1}/p_{i_2} and the column ratio x_{ij_1}/x_{ij_2} $(i = 1, \ldots, m)$ is the one of q_{j_1}/q_{j_2}. If $(m + n - 1)$ x_{ij}'s are fixed pertinently so that row ratios and column ratios are determined, we obtain the estimated frequencies of the remaining $(m - 1)(n - 1)$ cells which are suited to the pattern under independence. We will obtain several sets of $(m+n-1)$ x_{ij}'s for which the observed frequency is larger than or equal to the estimated frequency for the remaining $(m - 1)(n - 1)$ cells.

Let Z be a subset of X satisfying $|Z| = m + n - 1$ and let v_i (y_j) denote the number of elements of Z which belong to the i-th row (j-th column) of X. Thus we have

$$\sum_{i=1}^{m} v_i = \sum_{j=1}^{n} y_j = m + n - 1.$$

To define all row ratios based on Z, we first define the row ratio of the i_1-th and i_2-th rows by $s_{i_1i_2} = x_{i_1j}/x_{i_2j}$ if $x_{i_1j},\ x_{i_2j} \in Z$. Then we extend the definition based on the defined row ratios. For example, if $m = n = 3$ and $Z = \{x_{11}, x_{13}, x_{21}, x_{22}, x_{32}\}$, s_{12} and s_{23} are defined first. Then we set $s_{13} = s_{12}s_{23}$. Column ratios are similarly defined. We notice that it is necessary that

$$v_i > 0,\ \ i = 1, \ldots, m \ \text{ and } \ y_j > 0, j = 1,\ \ldots, n \tag{3.2}$$

for all row ratios and column ratios to be defined. However, even if the condition (3.2) is satisfied, all row and column ratios are not necessarily well defined. Consider, for example, the case where $m = n = 3$ and $Z = \{x_{11}, x_{12}, x_{21}, x_{22}, x_{33}\}$. Third row (column) is isolated and row (column) ratios including it are not defined. We need a further condition on Z.

Remark 3.2. Some $x_{ij}'s \in Z$ may be 0 in some cases. We set $0/0 = 1$ so that $s_{ij}s_{ji} = s_{ii} = 1$ always holds. Thus for any $0 < a < b$ we assume that $0/b < 0/a$ and $a/0 < b/0$.

We first introduce the following definitions of connectedness to define row and column ratios definitely.

Definition 3.1. (Connectedness of two elements of Z). Let Z be a subset of X.
1. $x_{ab} \in Z$ and $x_{cd} \in Z$ are connected if $a = c$ or $b = d$.
Further,
2. $x_{ab} \in Z$ and $x_{ef} \in Z$ are connected if x_{ab} and x_{cd} are connected and x_{cd} and x_{ef} are connected for some $x_{cd} \in Z$.

Thus two elements of Z are connected if one is reachable from the other by way of two elements of Z which are on the same row or column.

Definition 3.2. (Connectedness of Z). Let Z be a subset of X. Z is connected if any two elements of Z are connected.

Let

$$M = \{1, 2, \ldots, m\}, \ \text{ and } \ N = \{1, 2, \ldots, n\}.$$

Then we have the following.

Proposition 3.1. Z is not connected if and only if there exist $\emptyset \neq Z_1 \subset Z$, $\emptyset \neq M_1 \subset M$ and $\emptyset \neq N_1 \subset N$ such that

$$Z_1 \subset \{x_{ij}, i \in M_1, j \in N_1\} \ \text{ and } \ Z_1^c \subset \{x_{ij}, i \in M_1^c, j \in N_1^c\},$$

where $Z_1^c = Z \setminus Z_1 \neq \emptyset$, $M_1^c = M \setminus M_1 \neq \emptyset$, and $N_1^c = N \setminus N_1 \neq \emptyset$.

3.2.2 Basis and protrusive basis

Now we introduce the following.

Definition 3.3. (Basis). $Z \subset X$ is a basis of X if $|Z| = m + n - 1$, $v_i > 0, i = 1, \ldots, m$, $y_j > 0, j = 1, \ldots, n$ and Z is connected.

As we show in Proposition 3.2 below, if Z is a basis of X, all row and column ratios are determined. The following lemma is useful to show Proposition 3.2 as well.

Lemma 3.2. Let Z be a basis of an $m \times n$ table X.

1) If x_{i_0,j_0} is the only element of the i_0-th row which belongs to Z, then $Z \setminus \{x_{i_0,j_0}\}$ is a basis of the $(m-1) \times n$ table which we obtain by deleting the i_0-th row from X.

2) If x_{i_0,j_0} is the only element of the j_0-th column which belongs to Z, then $Z \setminus \{x_{i_0,j_0}\}$ is a basis of the $m \times (n-1)$ table which we obtain by deleting the j_0-th column from X.

Proposition 3.2. If Z is a basis of X, all row and column ratios are uniquely determined by Z.

Note. From the argument where $m = 2$ in the following proof, we see that for the cases where $m = 2$ or $n = 2$ if Z satisfies the condition

$$|Z| = m + n - 1, \; v_i > 0, i = 1, \ldots, m, \; y_j > 0, j = 1, \ldots, n, \tag{3.3}$$

then Z is connected and thus is a basis of X.

Now we give a canonical form for a basis Z of an $m \times n$ table X when we focus on a specific row (or column) and apply only interchanges of rows and columns. We obtain an expression of a row (columum) ratio by using the canonical form. The canonical form is shown in Table 3.1, where “O” means that the corresponding x_{ij} belongs to Z. The columns $j_1, \ldots, j_a$ have at least two elements of Z, including the one in the first (originally i_1-th) row. $R_\beta, 1 \leq \beta \leq a$, denotes the set of numbers of rows which have at least one element of Z which is reachable from $x_{i_1 j_\beta}$ without passing through $x_{i_1 j_\eta}, \eta \neq \beta$. A detailed proof is given in Appendix 1. By applying this expression to the transposed X and taking the transpose again, we also obtain a canonical form for a basis Z of X when we focus on a specific column.

Based on the canonical form, we obtain an expression of the row ratio $s_{i_k i_1}$ in terms of Z. We first see that $s_{i_k i_1}$ has the factor $1/x_{i_1 j_\gamma}$ if i_k belongs to $R_\gamma, 1 \leq \gamma \leq a$. See Table 3.2 which essentially shows the canonical form for a basis Z when we focus on the j_γ-*th* column. We set $i_1 = i(1)$ and $j_\gamma = j(1)$. We also notice that there exists unique $x_{i(2)j(1)} \in Z, i(2) \in R_\gamma$ from which each element of Z in the i_k-th row is reachable without passing through the other $x_{ij(1)} \in Z, i \neq i(2)$ (Table 3.3). Thus we see that $s_{i(2)i(1)} = x_{i(2)j(1)}/x_{i(1)j(1)}$ and our problem has reduced to the one of obtaining an expression of $s_{i_k i(2)}$ based on a smaller table $\widetilde{X}$ in Table 3.3. We note that the set of elements of $\widetilde{X}$ which belong to Z form a basis of $\widetilde{X}$ as is shown in Appendix 1 for the canonical form

Table 3.1 A canonical form for a basis Z based on the i_1-th row

	j_1	j_2	$\cdots$	j_β	$\cdots$	j_a	
i_1	O	O		O		O	O $\cdots$ O
R_1	O O O O O O						
$\vdots$		O O O O O					
R_2							
$\vdots$			$\ddots$				
R_β				O O O O O			
$\vdots$					$\ddots$		
R_a						O O O O O	

when we focus on a row. We repeat the similar procedure and have the expression

$$s_{i_k i_1} = \Pi_\eta \frac{x_{i(\eta+1)j(\eta)}}{x_{i(\eta)j(\eta)}}, \tag{3.4}$$

where η denotes the step number. We may notice that $s_{i_k i_1}$ is expressed in terms of all different x_{ij}'s since $j(\eta), \eta = 1, 2, \cdots$, are all different. For the columun ratio $t_{j_\ell j_1}$, a similar expression is obtained. When a basis Z of an $m \times n$ table X is given, row ratios

Table 3.2 A canonical form for a basis Z based on the j_γ-th column

			$j_\gamma = j(1)$				
	$i_1 = i(1)$		O				
			O	O O			
R_γ	i_k		O		O O O O		
			O			O O O O	

$s_{\alpha\beta}$, $\alpha, \beta = 1, \ldots, m$ and column ratios $t_{\gamma\delta}$, $\gamma, \delta = 1, \ldots, n$ are uniquely determined. We may notice that $s_{\alpha\beta} s_{\beta\alpha} = 1$ for any $\alpha, \beta = 1, \ldots, m$ and $t_{\gamma\delta} t_{\delta\gamma} = 1$ for any $\gamma, \delta = 1, \ldots, n$.

Table 3.3 A canonical form for a basis Z based on the $j(1)$-th column

		$j(1)$					
$i(1)$		O					
		O	O				
			O				
$i(2)$		O		O			
i_k				O	O	$\leftarrow \widetilde{X}$	
					O		
		O				O	
						O	O
							O

Then we define the following.

Definition 3.4. (Estimated frequency). Using any $x_{ab} \in Z$, the estimated frequency of the (i, j)-th cell of X based on a basis Z is defined as

$$\hat{x}_{ij} = x_{ab} s_{ia} t_{jb}, \quad i = 1, \ldots, m, \ j = 1, \ldots, n.$$

We note that $\hat{x}_{ij} = x_{ij}$ if $x_{ij} \in Z$. Further, we see that the estimated frequency is independent of the choice of $x_{ab} \in Z$ since $s_{\alpha\beta} s_{\beta\gamma} = s_{\alpha\gamma}$ for any $\alpha, \beta, \gamma = 1, \ldots, m$ and $t_{\delta\epsilon} t_{\epsilon\phi} = t_{\delta\phi}$ for any $\delta, \epsilon, \phi = 1, \ldots, n$. Now we give the following.

Definition 3.5. (Protrusive basis). Let Z be a basis of an $m \times n$ table $X = \{x_{ij}, i = 1, \ldots, m, \ j = 1, \ldots, n\}$ and let $\hat{x}_{ij}$ be the estimated frequency of the (i, j)-th cell of X based on Z. If $x_{ij} \geq \hat{x}_{ij}, i = 1, \ldots, m, \ j = 1, \ldots, n$, then we say that Z gives a protrusive pattern of X. We also say that Z is a p-basis of X for simplicity.

We may notice here that a p-basis depends on the observed value of X, but a basis does not.

3.2.3 Total number of protrusive bases

We begin by the following remark whose proof is clear and is omitted.

Remark 3.3. It can be easily verified that Lemma 3.2 is true even if the word "basis" is replaced by "basis which gives a protrusive pattern of X". Conversely, to look for a basis Z which gives a protrusive pattern of X and has only one element in a row (k-th row, say) (a column (ℓ-th column, say)) of X, we need to find a p-basis Z' of X' which we obtain by deleting the k-th row (ℓ-th column) from X. Once a p-basis Z' of X' is obtained, we choose an element in the k-th row (ℓ-th column) as a member of Z if the resulting $s_{ki}, i \neq k$ ($t_{\ell j}, j \neq \ell$) is minimized. The element is uniquely determined except for a tie. We give a method to treat a tie and will show that it enables us to resolve a tie.

To resolve a tie, we need to make a rule to include the case $A = B$ in $A \geq B$ or $A \leq B$, where A and B are functions of $X = \{x_{ij}, i = 1, \ldots, m, j = 1, \ldots, n\}$ such that A/B is a product of ratios of two x_{ij}'s . We propose the following two methods.

Method 1(based on row-first ordering)

x_{ij}'s are lined up as $x_{11}, x_{12}, \ldots, x_{1n}, x_{21}, x_{22}, \ldots, x_{2n}, \ldots, x_{m1}, x_{m2}, \ldots, x_{mn}$. Let $S_u(x_{i_0 j_0})$ be the set of x_{ij} which succeeds $x_{i_0 j_0}$. If the inequality $A > B$ is rewritten as $x_{i_0 j_0} > f(S_u(x_{i_0 j_0}))$ for some $x_{i_0 j_0}$, then the case $A = B$ is included in $A \geq B$, where $f(S_u(x_{i_0 j_0}))$ is a function of $x_{ij} \in S_u(x_{i_0 j_0})$.

Method 2(based on column-first ordering)

This is the same as Method 1 except that x_{ij}'s are line up as $x_{11}, x_{21}, \ldots, x_{m1}$, $x_{12}, x_{22}, \ldots, x_{m2}, \ldots, x_{1n}, x_{2n}, \ldots, x_{mn}$.

Suppose that a basis Z of X is given and row ratios are determined. As an example let us consider the inequality $x_{ad}/x_{cd} > s_{ac}$. Since s_{ac} is expressed as

$$s_{ac} = \Pi_\eta \frac{x_{i(\eta+1)j(\eta)}}{x_{i(\eta)j(\eta)}},$$

as shown in (3.4), we can easily see that $x_{ad}/x_{cd} > s_{ac}$ is rewritten as $x_{i_0 j_0} > f(S_u(x_{i_0 j_0}))$ or $x_{i_0 j_0} < f(S_u(x_{i_0 j_0}))$ for some $x_{i_0 j_0}$. Thus the case $x_{ad}/x_{cd} = s_{ac}$ is included in $x_{ad}/x_{cd} \geq s_{ac}$ or $x_{ad}/x_{cd} \leq s_{ac}$.

It will not seem that these methods are able to resolve all the ties, especially when the tie occurs among three or more quantities. However, we will show specifically that these methods work well in our case. In this paper we use Method 1 or 2 to determine whether the equality $A = B$ is included in $A \geq B$ or $A \leq B$. We use the notation $A \succ B$ when $A \geq B$ with the equality in the sense of Method 1 or 2. Now we show the following.

Proposition 3.3

(i) Let X be an $m \times (\alpha + 1)$ table and let $\widetilde{X}$ be the $m \times \alpha$ table obtained by deleting the j_0-th ($j_0 = 1, \ldots, \alpha + 1$) column from X, where $\alpha \geq 1$. Suppose that a p-basis $\widetilde{Z}$ is given for $\widetilde{X}$. Then an element $x_{i_a j_0}$ of the j_0-th column of X is uniquely determined by Method 1 (or 2) so that $\widetilde{Z} \cup \{x_{i_a j_0}\}$ is a p-basis of X.

(ii) Let X be an $(\alpha + 1) \times n$ table and let $\widetilde{X}$ be the $\alpha \times n$ table obtained by deleting the i_0-th ($i_0 = 1, \ldots, \alpha + 1$) row from X. Suppose that a p-basis $\widetilde{Z}$ is given for $\widetilde{X}$. Then an element $x_{i_0 j_a}$ of the i_0-th row of X is uniquely determined by Method 1 (or 2) so that $\widetilde{Z} \cup \{x_{i_0 j_a}\}$ is a p-basis of X.

A proof is given in Appendix 2.

Remark 3.4 As for (i) of Proposition 3.3, if there exist columns of $\widetilde{X}$ each of which has only element belonging to $\widetilde{Z}$, the columns do not contribute to determine the row ratios. Therefore, deleting the columns, we may assume that all columns of $\widetilde{X}$ have two or more elements which belong to $\widetilde{Z}$. A similar remark also applies to (ii) of Proposition 3.3.

Let $T_{m\times n}$ be the total number of bases which give protrusive patterns of an observed $m \times n$ table X. To discuss $T_{m\times n}$, we first need the following.

Definition 3.6. Let Z be a p-basis of an $m \times n$ table X.

(i) If the i-th row of X has v_i elements belonging to Z, we say Z is a row$(v_1, v_2, \ldots, v_m)$ p-basis, where $v_i \geq 1, i = 1, \ldots, m$ and $\sum_{i=1}^m v_i = m + n - 1$.

(ii) If the j-th column of X has y_j elements belonging to Z, we say Z is a col$(y_1, y_2, \ldots, y_n)$ p-basis, where $y_j \geq 1, j = 1, \ldots, n$ and $\sum_{j=1}^{n} y_j = m + n - 1$.

The existence of a row$(v_1, v_2, \ldots, v_m)$ (col$(y_1, y_2, \ldots, y_n)$) p-basis is established by the following proposition whose proof is given in Appendix 3.

Proposition 3.4. Let an $m \times n$ table X be observed.

(i) For any $(v_1, v_2, \ldots, v_m)$ which satisfies $v_i \geq 1, i = 1, \ldots, m$ and $\sum_{i=1}^{m} v_i = m + n - 1$, there exists a row$(v_1, v_2, \ldots, v_m)$ p-basis of X.

(ii) For any $(y_1, y_2, \ldots, y_n)$ which satisfies $y_j \geq 1, j = 1, \ldots, n$ and $\sum_{j=1}^{n} y_j = m + n - 1$, there exists a col$(y_1, y_2, \ldots, y_n)$ p-basis of X.

Now we have the following proposition whose proof is given in Appendix 4.

Proposition 3.5. $T_{m \times n} = {}_{m+n-2}C_{m-1}$ for any $m, n \geq 2$.

To show the uniqueness of a row$(v_1, v_2, \ldots, v_m)$ (col$(y_1, y_2, \ldots, y_n)$) p-basis, we need the following lemma whose proof is given in Appendix 5.

Lemma 3.3. For any $m, n \geq 1$, let

$$\mathcal{T}_{m,n} = \{(y_1, \ldots, y_n) | y_j \geq 1, j = 1, \ldots, n, \sum_{j=1}^{n} y_j = m + n - 1\}.$$

Then $|\mathcal{T}_{m,n}| = {}_{m+n-2}C_{m-1}$.

From Proposition 3.5 and Lemma 3.3, we see that $T_{m \times n} = |\mathcal{T}_{m,n}|$. Thus using Proposition 3.4 we have the following.

Corollary 3.1 Let an $m \times n$ table X be observed.

(i) For any $(v_1, v_2, \ldots, v_m)$ which satisfy $v_i \geq 1, i = 1, \ldots, m$ and $\sum_{i=1}^{m} v_i = m + n - 1$, there exists a unique row$(v_1, v_2, \ldots, v_m)$ p-basis of X.

(ii) For any $(y_1, y_2, \ldots, y_n)$ which satisfy $y_j \geq 1, j = 1, \ldots, n$ and $\sum_{j=1}^{n} y_j = m + n - 1$, there exists a unique col$(y_1, y_2, \ldots, y_n)$ p-basis of X.

3.2.4 A numerical algorithm and a shrinkage estimator

Here we describe a numerical algorithm for all protrusive bases and propose a shrinkage estimator which dominates the unbiased estimator. It may be noticed that Method 1 or Method 2 is applied to resolve a tie.

Numerical algorithm. For a protrusive basis Z of an $m \times n$ table X, let $Y = \{y_1, \ldots, y_n\}$ be the set of numbers of elements of Z in each column of X. Thus

$$y_j > 0, \quad j = 1, \ldots, n \quad \text{and} \quad \sum_{j=1}^{n} y_j = m + n - 1.$$

We assume that $m \leq n$ without loss of generality. For the detail explanation of $Q_{q\times\ell} ={}_{q-2}C_{\ell-1}$, $q > \ell \geq 2$, see Appendix 4.

In case where $m = 2$ it is easy to obtain $T_{2\times n} = {}_nC_1 = n$ p-bases since $Y = \{2, 1, \ldots, 1\}$. Once the column which gives the row ratio is determined, we need to examine which element should belong to Z for each of the remaining columns.

In case where $m = 3$, we only have $Y = \{3, 1, \ldots, 1\}$ and $\{2, 2, 1, \ldots, 1\}$. For $Y = \{3, 1, \ldots, 1\}$, n p-bases are easily obtained as in the case $m = 2$. For the case where $Y = \{2, 2, 1, \ldots, 1\}$, we first determine a set of two columns which should have two elements of Z each. Then we consider the problem of a 3×2 table, which we can treat easily as discussed in the proof of Proposition 3.5. Since $Q_{3\times 2} = 1$, we have $n(n-1)/2$ p-bases for the case where $Y = \{2, 2, 1, \ldots, 1\}$.

In case where $m = 4$, we have $Y = \{4, 1, \ldots, 1\}$, $\{3, 2, 1, \ldots, 1\}$ and $\{2, 2, 2, 1, \ldots, 1\}$. We can treat the first and second cases similarly as in the case $m = 3$ and have $n+n(n-1)$ p-bases. For the case where $Y = \{2, 2, 2, 1, \ldots, 1\}$, we first determine a set of three columns which should have two elements of Z each. Since $Q_{4\times 3} = 1$, a p-basis exists uniquely for each set of three columns. However, in order to find the unique p-basis, we may have to examine all $T_{4\times 3} = 10$ p-bases of the 4×3 (or 3×4) table. An algorithm described in Appendix 3 (especially Lemma A.2 and Appendix 3.1) will be helpful to get the col$(2, 2, 2)$ p-basis of a 4×3 table. The col$(2, 2, 2)$ p-basis will be easily obtained if, for example, the col$(3, 2, 1)$ p-basis is available.

We stop the discussion with brief comments on the case of $5 \times n$ table. In this case we have to treat the cases $Y = \{3, 2, 2, 1, \ldots, 1\}$ and $\{2, 2, 2, 2, 1, \ldots, 1\}$. For the case where $Y = \{3, 2, 2, 1, \ldots, 1\}$, we may have to examine all $T_{5\times 3} = 15$ p-bases in order to find $Q_{5\times 3} = 3$ p-bases. For the case where $Y = \{2, 2, 2, 2, 1, \ldots, 1\}$, we may have to examine all $T_{5\times 4} = 35$ p-bases in order to find $Q_{5\times 4} = 1$ p-basis. An algorithm described in Appendix 3 will be helpful to get these p-bases

Shrinkage estimator

We first define our shrinkage estimator based on a fixed col$(y_1, \ldots, y_n)$ p-basis, where $(y_1, \ldots, y_n)$ satifies $y_j \geq 1, j = 1, \ldots, n$ and $\sum_{j=1}^{n} y_j = m+n-1$. To clarify the argument, we distinguish between variables and their observations till the end of the proof of Proposition 3.6. We denote the observation by x^o_{ij} and the variable simply by its subscripts (i, j). We denote the variable sets by X, Z and W and their observations by X^o, Z^o and W^o, respectively. We have shown in Corollary 3.1 that for an observed $m \times n$ table $X^o = \{x^o_{ij}, i = 1, \ldots, m,\ j = 1, \ldots, n\} \in S = \{X^o | x^o_{ij} \geq 0, i = 1, \ldots, m,\ j = 1, \ldots, n\}$, there exists a unique col$(y_1, \ldots, y_n)$ p-basis Z when Method 1 or Method 2 is applied. Thus, based on an observation X^o, we divide the set of mn variables $X = \{(i, j), i = 1, \ldots, m, j = 1, \ldots, n\}$ into the col$(y_1, \ldots, y_n)$ p-basis Z and its complement $W = X \setminus Z$. Since Z and W depend on the observation X^o, we may write them as $Z(X^o)$ and $W(X^o)$. Let L be the number of elements of the set

$$\{Z(X^o) | X^o \in S\}.$$

As a matter of fact, L is the total number of col$(y_1, \ldots, y_n)$ bases of an $m \times n$ table since any col$(y_1, \ldots, y_n)$ basis is a col$(y_1, \ldots, y_n)$ p-basis for X^o belonging to some non-empty subset of S. Although a simple expression of L in terms of m, n and $(y_1, \ldots, y_n)$ is not available, the value of L may be obtained numerically. However, neither an expression nor the value of L is necessary for our discussion.

Now we give a partition of the total space S based on $\mathrm{col}(y_1, \ldots, y_n)$ p-bases. Since there exists L different choices of a basis Z, numbering them by some method, we denote them by $Z_\ell, \ell = 1, \ldots, L$. We may number L Z's by ordering them in terms of the variables (i, j)'s belonging to Z which themselves may be ordered as in Method 1 or 2. For a given $\mathrm{col}(y_1, \ldots, y_n)$ basis Z_ℓ let

$$S_\ell = \{X^o \in S | Z(X^o) = Z_\ell\},\ \ell = 1, \ldots, L.$$

Then we see that $\bigcup_{\ell=1}^{L} S_\ell = S$ and $S_1, \ldots, S_L$ are disjoint since $Z(X^o)$ is uniquely determined for any observation $X^o \in S$ from Corollary 3.1. Thus $S_1, \ldots, S_L$ gives a partition of S.

To explain our shrinkage estimator, we first consider the case where the observation $X^o \in S_\ell\,(\ell = 1, \ldots, L)$ since the shrinkage pattern is completely different among the regions S_ℓ, $\ell = 1, \ldots, L$. We note that $Z(X^o) = Z_\ell$ and $W(X^o) = X \setminus Z_\ell \equiv W_\ell$ for any $X^o \in S_\ell$. Let Z_ℓ^o and W_ℓ^o be the observed values of Z_ℓ and W_ℓ, respectively, which are both parts of the observation X^o. For the variable $(i, j) \in W_\ell$, let $\widehat{x}_{ij}(Z_\ell^o)$ be the estimated frequency based on Z_ℓ. Collecting all $\widehat{x}_{ij}(Z_\ell^o), (i, j) \in W_\ell$ together, we put them as $\widehat{W}_\ell(Z_\ell^o)$. Then we see from the definition of the p-basis that

$$S_\ell = \{X^o \in S | W_\ell^o \geq \widehat{W}_\ell(Z_\ell^o)\},$$

where it is determined by Method 1 or 2 whether the equality of each component of $W_\ell^o \geq \widehat{W}_\ell(Z_\ell^o)$ is included or not if it happens. Therefore when the value of $Z_\ell^o = z_\ell^o$ is fixed, the set of W_ℓ^o for which $X^o \in S_\ell$ is given by the $(m-1)(n-1)$ dimensional rectangular region $W_\ell^o \geq \widehat{W}_\ell(z_\ell^o)$. Thus Remark 3.1 to Lemma 3.1 will be applicable when we consider the shrinkage estimator.

Now we discuss the general case and propose our shrinkage estimator. Let $X^o \in S$ be observed and let $Z(X^o)$ be the unique $\mathrm{col}(y_1, \ldots, y_n)$ p-basis determined by X^o. Let Z^o be the observed value of $Z(X^o)$ which is a part of the observation X^o. Similarly, for $W(X^o) = X \setminus Z(X^o)$, let W^o be the observed value of $W(X^o)$ which is also a part of the observation X^o. Although it is not explicitly indicated, Z^o and W^o depend on the total observation X^o since the variable sets $Z(X^o)$ and $W(X^o)$ themselves depend on X^o. For the variable $(i, j) \in W(X^o)$, let $\hat{x}_{ij}(Z^o)$ be the estimated frequency based on $Z(X^o)$. Then the following shrinkage estimator is proposed.

$$\psi_{ij}(X^o) = \begin{cases} x_{ij}^o, & (i, j) \in Z(X^o), \\ x_{ij}^o - a\dfrac{x_{ij}^o - \hat{x}_{ij}(Z^o)}{T + d}, & (i, j) \in W(X^o), \end{cases}$$

where $T = \sum_{(i,j)\in W(X^o)}\{x_{ij}^o - \hat{x}_{ij}(Z^o)\}$, $a > 0$ and $d > 0$. Let $\Psi(X^o) = \{\psi_{ij}(X^o), i = 1, \ldots, m,\ j = 1, \ldots, n\}$ and we have following.

Proposition 3.6. Suppose that $(m-1)(n-1) \geq 2$. Then $\Psi(X^o)$ improves upon the unbiased estimator X^o if $0 < a \leq 2(mn - m - n)$ and $d \geq a/2$.

3.3 A numerical example

We take the 3×4 contingency table given by Table 3.4 as an example and illustrate the shrinkage estimator.

Table 3.4. A 3×4 table : observed frequencies and expected frequencies under independence.

	B_1	B_2	B_3	B_4	Total
A_1	5 (3.750)	2 (3.214)	3 (3.482)	5 (4.554)	15
A_2	3 (4.000)	7 (3.429)	2 (3.714)	4 (4.857)	16
A_3	6 (6.250)	3 (5.357)	8 (5.804)	8 (7.589)	25
Total	14	12	13	17	56

The number of protrusive bases is $T_{3\times4} = {}_5C_2 = 10$. The set of numbers of elements of a protrusive basis Z in each column of a 3×4 table X is $\{3,1,1,1\}$ or $\{2,2,1,1\}$. When we denote the numbers of elements of Z in four columns of X by (c_1, c_2, c_3, c_4), we only treat the two cases $(3,1,1,1)$ and $(2,2,1,1)$. The other cases are treated in the same way. Since $(4-1)(3-1) = 6$, we choose $a = 6-1 = 5$ and $d = a/2 = 2.5$ in the shrinkage estimator.

Case (3,1,1,1). The protrusive basis consists of $x_{11} = 5, x_{21} = 3, x_{31} = 6, x_{12} = 2, x_{13} = 3$ and $x_{14} = 5$ and the estimated frequencies of the other cells are given as $\hat{x}_{22} = 3 \times 2/5 = 1.2$, $\hat{x}_{32} = 6 \times 2/5 = 2.4$, $\hat{x}_{23} = 3 \times 3/5 = 1.8$, $\hat{x}_{33} = 6 \times 3/5 = 3.6$, $\hat{x}_{24} = 3 \times 5/5 = 3.0$ and $\hat{x}_{34} = 6 \times 5/5 = 6.0$ (Table 3.5).

Table 3.5. The table of the protrusive basis and estimated frequencies.

	B_1	B_2	B_3	B_4
A_1	5	2	3	5
A_2	3	$7 \to 1.2$	$2 \to 1.8$	$4 \to 3$
A_3	6	$3 \to 2.4$	$8 \to 3.6$	$8 \to 6$

Since $\sum_{x_{ij}\notin Z}(x_{ij} - \hat{x}_{ij}) + d = 16.5$, shrinkage factor $= 1 - 5/16.5 = 0.697$ and we have $\psi_{ij}^{(1)}(X), i = 1,2,3,\; j = 1,2,3,4$ as given in Table 3.6.

Table 3.6. Shrinkage estimates $\psi_{ij}^{(1)}(X), i = 1,2,3,\; j = 1,2,3,4$ for the case $(3,1,1,1)$.

	B_1	B_2	B_3	B_4
A_1	5	2	3	5
A_2	3	5.242	1.939	3.697
A_3	6	2.818	6.667	7.394

Case (2,2,1,1). Since $\frac{2}{5} < \frac{3}{6} < \frac{7}{3}$, $x_{21} = 3, x_{31} = 6, x_{12} = 2$ and $x_{32} = 3$ are included in the basis and thus $x_{23} = 2$ and $x_{14} = 5$ are included further. Then we have $\hat{x}_{11} =$

$2 \times 6/3 = 4.0$, $\hat{x}_{22} = 3 \times 3/6 = 1.5$, $\hat{x}_{13} = 2 \times 6/3 \times 2/3 = 2.33$, $\hat{x}_{33} = 6 \times 3/2 = 4.0$, $\hat{x}_{24} = 3 \times 3/6 \times 5/2 = 3.75$ and $\hat{x}_{34} = 3 \times 5/2 = 7.5$. Since $\sum_{x_{ij} \notin Z}(x_{ij} - \hat{x}_{ij}) + d = 14.083$, shrinkage factor $= 1 - 5/14.083 = 0.645$ and we have $\psi_{ij}^{(5)}(X), i = 1, 2, 3,\ j = 1, 2, 3, 4$ as given in Table 3.7.

Table 3.7. Shrinkage estimates $\psi_{ij}^{(5)}(X), i = 1, 2, 3,\ j = 1, 2, 3, 4$ for the case $(2, 2, 1, 1)$.

	B_1	B_2	B_3	B_4
A_1	4.645	2	2.882	5
A_2	3	5.047	2	3.911
A_3	6	3	6.580	7.822

Finally by averaging 10 estimates we have $\psi_{ij}(X) = \sum_{k=1}^{10} \psi_{ij}^{(k)}(X)/10$ as given in Table 3.8.

Table 3.8. Shrinkage estimates $\psi_{ij}(X), i = 1, 2, 3,\ j = 1, 2, 3, 4$.

	B_1	B_2	B_3	B_4
A_1	4.598	1.963	2.898	4.882
A_2	2.928	5.306	1.994	3.835
A_3	5.911	2.971	6.806	7.741

For comparison we have applied the shrinkage estimator (3.1) which shrink x_{ij} to its expected frequency $\tilde{x}_{ij}$ when $x_{ij} \geq \tilde{x}_{ij}$, although it is not shown to improve upon x_{ij}. Since 5 $x_{ij}{}'$s are larger than their respective expected frequencies in our case, we set $p = 5,\ a = p - 1 = 4,\ d = a/2 = 2$. Thus we have $\sum_i \sum_j (x_{ij} - \tilde{x}_{ij})^+ + d = 9.875$ and shrinkage factor $= 1 - 4/9.875 = 0.595$. The resulting estimates are given in Table 3.9. We note that they are quite close to those given in Table 3.8.

Table 3.9. Shrinkage estimates when we shrink 5 observed frequencies which are larger than their respective expected frequencies.

	B_1	B_2	B_3	B_4
A_1	4.494	2	3	4.819
A_2	3	5.553	2	4
A_3	6	3	7.110	7.834

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] Chang, Y.-T. and Shinozaki, N. (2022). Simultaneous estimation of Poisson means in two-way contingency tables under normalized squared error loss. *Japanese J. Statist. and Data Analysis*, 52, 577-628, https://doi.org/10.1007/s42081-022-00169-9.

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

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

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

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

[8] 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.

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

[10] 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.

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

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

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

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

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

[16] 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 also thanks Professor Miyata Manabu, Mejiro university for his support.

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

**Shinozaki Nobuo
Keio University
Kanagawa 223-8522
JAPAN