Brazilian Journal of Probability and Statistics
2020, Vol. 34, No. 4, 728–751
https://doi.org/10.1214/19-BJPS459

Proper Bayes minimax estimation of parameters of Poisson distributions in the presence of unbalanced sample sizes

Yasuyuki Hamura and Tatsuya Kubokawa

University of Tokyo

Abstract. In this paper, we consider the problem of simultaneously estimating parameters of independent Poisson distributions in the presence of possibly unbalanced sample sizes under weighted standardized squared error loss. A class of heterogeneous Bayesian shrinkage estimators that utilize the unbalanced nature of sample sizes is proposed. To provide a theoretical justification, we first derive a necessary and sufficient condition for an estimator in the class to be proper Bayes and hence admissible and then obtain sufficient conditions for minimaxity that are compatible with the admissibility condition. Heterogeneous and homogeneous shrinkage estimators are compared by simulation. Several estimation methods are applied to data relating to the standardized mortality ratio.

1 Introduction

Since the work of Clevenson and Zidek (1975), simultaneous estimation of parameters of independent Poisson distributions has been studied by many authors including Tsui (1979), Tsui and Press (1982), Hwang (1982), and Chang and Shinozaki (2018). However, most of the existing work either concerns with the case of balanced sample sizes or deals with estimators in the unbalanced case which do not utilize the fact that the sample sizes are unbalanced. In this paper, we consider the estimation problem in the case of unbalanced sample sizes and construct shrinkage estimators whose shrinkage factors reflect the fact that the sample sizes are unbalanced.

Suppose that $X_1, \dots, X_m$ are mutually independent Poisson random variables with means $n_1\lambda_1, \dots, n_m\lambda_m$, respectively, and that $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_m) \in (0, \infty)^m$ is the unknown parameter while $n_1, \dots, n_m$ are positive known constants. This situation arises, for example, when for each $i = 1, \dots, m$, the observation X_i is the sum of $n_i (\in \mathbb{N})$ random sample from the Poisson distribution with mean λ_i. An example where $n_1, \dots, n_m$ are positive (possibly non-integer) real numbers is given in Komaki (2015). We treat the problem of estimating $\boldsymbol{\lambda}$ on the basis of $\boldsymbol{X} = (X_1, \dots, X_m)$.

In the balanced case with $n_1 = \cdots = n_m = 1$, the model becomes equivalent to that considered by Clevenson and Zidek (1975). When $n_1 = \cdots = n_m = 1$, for the avoidance of confusion, we use the different notation $Y_1 = X_1, \dots, Y_m = X_m$. Then, they showed that the estimator

$$\left(1 - \frac{\beta_0 + m - 1}{\sum_{i=1}^m Y_i + \beta_0 + m - 1}\right)(Y_1, \dots, Y_m) \tag{1.1}$$

is admissible for $1 < \beta_0$ and minimax for $m \ge 2$ and $0 \le \beta_0 \le m - 1$ relative to the loss function $\sum_{i=1}^m (d_i - \lambda_i)^2/\lambda_i$.

Using their result, we can readily verify that the estimator

$$\left(1 - \frac{\beta_0 + m - 1}{\sum_{i=1}^m X_i + \beta_0 + m - 1}\right)\left(\frac{X_1}{n_1}, \dots, \frac{X_m}{n_m}\right) \tag{1.2}$$

Key words and phrases. Admissibility, Bayes estimation, dominance, minimaxity, shrinkage prior, unbalanced sample sizes.

Received October 2018; accepted September 2019.

dominates the ML estimator $(X_1/n_1, \ldots, X_m/n_m)$ if $m \geq 2$ and $0 \leq \beta_0 \leq m-1$ under the loss

$$\sum_{i=1}^{m} \frac{n_i}{\lambda_i}(d_i - \lambda_i)^2. \tag{1.3}$$

However, the estimator given by (1.2) is not necessarily a natural shrinkage estimator from a practical point of view because the shrinkage factor $1-(\beta_0+m-1)/(\sum_{i=1}^{m} X_i+\beta_0+m-1)$ is common to all the samples irrespective of $\boldsymbol{n} = (n_1, \ldots, n_m)$. In many applications, one of the purposes of using shrinkage estimators is to reduce the instability of ML estimators. In the present setting, for all $i, j = 1, \ldots, m$ such that $n_i < n_j$, the ML estimator X_i/n_i tends to be more unstable than X_j/n_j since the variance of X_i/n_i is approximately n_j/n_i times the variance of X_j/n_j if $\lambda_i \approx \lambda_j$. In addition, for each $i = 1, \ldots, m$, the sample size n_i can be interpreted as representing the amount of information the observation X_i contains about the unknown parameter λ_i. Thus, it seems reasonable to use a shrinkage estimator such that it shrinks the ML estimator X_i/n_i more toward the origin than X_j/n_j for all $i, j = 1, \ldots, m$ such that $n_i < n_j$. Furthermore, it turns out in Section 2 that the estimator given by (1.2) with $m \geq 2$ and $\beta_0 \geq 0$ is the Bayes estimator with respect to a perhaps unnatural shrinkage prior which depends on $\boldsymbol{n}$ and puts less weight on the smaller values of λ_i than on the smaller values of λ_j for all $i, j = 1, \ldots, m$ such that $n_i < n_j$.

Thus, in the present paper, we consider the class of heterogeneous shrinkage estimators

$$\left(\{1-\phi_1(\boldsymbol{X})\}\frac{X_1}{n_1}, \ldots, \{1-\phi_m(\boldsymbol{X})\}\frac{X_m}{n_m}\right), \tag{1.4}$$

where the functions $\phi_1, \ldots, \phi_m \colon \{0, 1, 2, \ldots\}^m \to [0, 1]$ satisfy that $\phi_i(\boldsymbol{x}) > \phi_j(\boldsymbol{x})$ for all $\boldsymbol{x} = (x_1, \ldots, x_m) \in \{0, 1, 2, \ldots\}^m$ and $i, j = 1, \ldots, m$ such that $x_i, x_j \geq 1$ and $n_i < n_j$. We evaluate estimators under the general loss function given by

$$L_{\boldsymbol{c}}(\boldsymbol{d}, \boldsymbol{\lambda}) = \sum_{i=1}^{m} \frac{c_i}{\lambda_i}(d_i - \lambda_i)^2, \tag{1.5}$$

where $\boldsymbol{c} = (c_1, \ldots, c_m) \in (0, \infty)^m$ is a vector of weights possibly different form $\boldsymbol{n}$ and $\boldsymbol{d} = (d_1, \ldots, d_m)$ denotes a m-dimensional vector. For a discussion of the estimation of normal means in the presence of unequal weights as well as unequal variances, see Morris (1983).

Hamura and Kubokawa (2019) constructed shrinkage estimators of the form (1.4) by using a class of improper priors introduced by Komaki (2015). However, they did not prove the admissibility of the estimators. In this paper, we introduce a class of priors which includes both the proper priors of Clevenson and Zidek (1975) and the improper priors of Komaki (2015), construct proper Bayes estimators of the form (1.4), and derive sufficient conditions for the estimators to be minimax. The results for proper prior distributions are not straightforward generalizations of those for improper prior distributions. The main contribution of this paper is to construct Bayes estimators of the form (1.4) which are both admissible and minimax.

In Section 2, we introduce the class of priors mentioned above, derive a necessary and sufficient condition for a prior in the class to be proper, and express the corresponding Bayes estimators explicitly. In Section 3, we derive sufficient conditions for minimaxity. In Section 4, some Monte Carlo evidence is presented. In Section 5, we treat real data. All the proofs of the lemmas in Sections 2 and 3 are given in the Appendix.

2 A class of Bayes estimators

We begin by providing a class of priors which includes the priors of both Clevenson and Zidek (1975) and Komaki (2015). Let

$$\pi_{\alpha,\boldsymbol{\beta},\boldsymbol{\gamma};\beta_0,\gamma_0}(\boldsymbol{\lambda}) = \frac{\prod_{i=1}^m {\lambda_i}^{\beta_i-1}}{(\sum_{i=1}^m \lambda_i/\gamma_i)^\alpha} \int_0^\infty \frac{u^{\alpha-1+\beta_0}}{(u/\gamma_0 + \sum_{i=1}^m \lambda_i/\gamma_i)^{\beta_0}} e^{-u}\, du \tag{2.1}$$

for $\alpha > 0$, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_m) \in (0,\infty)^m$, $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_m) \in (0,\infty)^m$, $\beta_0 \geq 0$, and $\gamma_0 > 0$. By making the change of variables $u' = u/(\sum_{i=1}^m \lambda_i/\gamma_i)$, we can write (2.1) as

$$\pi_{\alpha,\boldsymbol{\beta},\boldsymbol{\gamma};\beta_0,\gamma_0}(\boldsymbol{\lambda}) = \left(\prod_{i=1}^m {\lambda_i}^{\beta_i-1}\right) \int_0^\infty \frac{u^{\alpha-1+\beta_0}}{(1+u/\gamma_0)^{\beta_0}} e^{-u\sum_{i=1}^m \lambda_i/\gamma_i}\, du. \tag{2.2}$$

The class of priors of Clevenson and Zidek (1975) is expressed as

$$\pi_{m-1,\boldsymbol{j},\boldsymbol{j};\beta_0,1}(\boldsymbol{\lambda}) = \frac{1}{(\sum_{i=1}^m \lambda_i)^{m-1}} \int_0^\infty \frac{u^{m-2+\beta_0}}{(u + \sum_{i=1}^m \lambda_i)^{\beta_0}} e^{-u}\, du, \tag{2.3}$$

where $\boldsymbol{j} = (1, \dots, 1) \in \mathbb{R}^m$, when $m \geq 2$ or $\beta_0 > 0$. The prior (2.3) is proper if $\beta_0 > 1$, as shown by Clevenson and Zidek (1975). On the other hand, the class of priors of Komaki (2015) is described by

$$\frac{\pi_{\alpha,\boldsymbol{\beta},\boldsymbol{\gamma};0,1}(\boldsymbol{\lambda})}{\Gamma(\alpha)} = \frac{\prod_{i=1}^m {\lambda_i}^{\beta_i-1}}{(\sum_{i=1}^m \lambda_i/\gamma_i)^\alpha}.$$

This prior is improper for all values of α, $\boldsymbol{\beta}$, and $\boldsymbol{\gamma}$, which can be verified by, for example, Lemma 2.1 below.

The following lemma gives a necessary and sufficient condition for the prior $\pi_{\alpha,\boldsymbol{\beta},\boldsymbol{\gamma};\beta_0,\gamma_0}$ to be proper.

Lemma 2.1. *The prior* $\pi_{\alpha,\boldsymbol{\beta},\boldsymbol{\gamma};\beta_0,\gamma_0}$ *satisfies*

$$\int \cdots \int_{(0,\infty)^m} \pi_{\alpha,\boldsymbol{\beta},\boldsymbol{\gamma};\beta_0,\gamma_0}(\boldsymbol{\lambda})\, d\boldsymbol{\lambda} < \infty$$

if and only if $\alpha < \sum_{i=1}^m \beta_i < \alpha + \beta_0$.

Next, we derive an explicit form of the Bayes estimator against the prior $\pi_{\alpha,\boldsymbol{\beta},\boldsymbol{\gamma};\beta_0,\gamma_0}$. To this end, we define

$$K(\boldsymbol{\gamma}, \boldsymbol{\xi}, \alpha; \gamma_0, \beta_0) = \int_0^\infty \frac{u^{\alpha-1}}{(1+u/\gamma_0)^{\beta_0}} \prod_{i=1}^m \frac{1}{(1+u/\gamma_i)^{\xi_i}}\, du$$

for $\boldsymbol{\xi} = (\xi_1, \dots, \xi_m) \in [0,\infty)^m$ such that $\beta_0 + \sum_{i=1}^m \xi_i > \alpha$. This function is a generalization of the function given by Komaki (2015) which generalizes the beta function. Indeed, when $\gamma_0 = \gamma_1 = \cdots = \gamma_m$, we have

$$K(\boldsymbol{\gamma}, \boldsymbol{\xi}, \alpha; \gamma_0, \beta_0) = {\gamma_0}^\alpha B(\alpha, \beta_0 + \xi_\cdot - \alpha) \tag{2.4}$$

for $\xi_\cdot = \sum_{i=1}^m \xi_i$. The function K satisfies the following properties. Let $\boldsymbol{e}_i$ denote the ith unit vector in $\mathbb{R}^m$, namely the ith row of the $m \times m$ identity matrix, for $i = 1, \dots, m$.

Lemma 2.2. *The following relations hold.*

(i)

$$\alpha K(\boldsymbol{\gamma}, \boldsymbol{\xi}, \alpha; \gamma_0, \beta_0) = \frac{\beta_0}{\gamma_0} K(\boldsymbol{\gamma}, \boldsymbol{\xi}, \alpha + 1; \gamma_0, \beta_0 + 1) + \sum_{i=1}^{m} \frac{\xi_i}{\gamma_i} K(\boldsymbol{\gamma}, \boldsymbol{\xi} + \boldsymbol{e}_i, \alpha + 1; \gamma_0, \beta_0). \tag{2.5}$$

(ii) *For* $i = 1, \ldots, m$,

$$K(\boldsymbol{\gamma}, \boldsymbol{\xi} + \boldsymbol{e}_i, \alpha + 1; \gamma_0, \beta_0) = \gamma_i \{ K(\boldsymbol{\gamma}, \boldsymbol{\xi}, \alpha; \gamma_0, \beta_0) - K(\boldsymbol{\gamma}, \boldsymbol{\xi} + \boldsymbol{e}_i, \alpha; \gamma_0, \beta_0) \}. \tag{2.6}$$

For the case of $\beta_0 = 0$, the relations (2.5) and (2.6) are given in Lemma 5 of Komaki (2015).

The following lemma gives some more properties of the function K and is crucial in Section 3 when we prove the existence of a heterogeneous shrinkage estimator which is both admissible and minimax.

Lemma 2.3. *Suppose that* $\alpha < \beta_0 + \sum_{i=1}^{m} \xi_i - 1$. *Then the following inequalities hold.*

(i) *For* $i = 1, \ldots, m$,

$$\frac{K(\boldsymbol{\gamma}, \boldsymbol{\xi}, \alpha + 1; \gamma_0, \beta_0)}{K(\boldsymbol{\gamma}, \boldsymbol{\xi} - \boldsymbol{e}_i, \alpha; \gamma_0, \beta_0)} \geq \frac{K(\boldsymbol{\gamma}, \boldsymbol{\xi} + \boldsymbol{e}_i, \alpha + 1; \gamma_0, \beta_0)}{K(\boldsymbol{\gamma}, \boldsymbol{\xi}, \alpha; \gamma_0, \beta_0)}. \tag{2.7}$$

Similarly,

$$\frac{K(\boldsymbol{\gamma}, \boldsymbol{\xi}, \alpha + 1; \gamma_0, \beta_0)}{K(\boldsymbol{\gamma}, \boldsymbol{\xi}, \alpha; \gamma_0, \beta_0 - 1)} \geq \frac{K(\boldsymbol{\gamma}, \boldsymbol{\xi}, \alpha + 1; \gamma_0, \beta_0 + 1)}{K(\boldsymbol{\gamma}, \boldsymbol{\xi}, \alpha; \gamma_0, \beta_0)}. \tag{2.8}$$

(ii) *For* $i = 1, \ldots, m$,

$$\frac{K(\boldsymbol{\gamma}, \boldsymbol{\xi} + \boldsymbol{e}_i, \alpha + 2; \gamma_0, \beta_0)}{K(\boldsymbol{\gamma}, \boldsymbol{\xi}, \alpha; \gamma_0, \beta_0)} \geq \frac{K(\boldsymbol{\gamma}, \boldsymbol{\xi}, \alpha + 1; \gamma_0, \beta_0)}{K(\boldsymbol{\gamma}, \boldsymbol{\xi}, \alpha; \gamma_0, \beta_0)} \frac{K(\boldsymbol{\gamma}, \boldsymbol{\xi} + \boldsymbol{e}_i, \alpha + 1; \gamma_0, \beta_0)}{K(\boldsymbol{\gamma}, \boldsymbol{\xi}, \alpha; \gamma_0, \beta_0)}. \tag{2.9}$$

Similarly,

$$\frac{K(\boldsymbol{\gamma}, \boldsymbol{\xi}, \alpha + 2; \gamma_0, \beta_0 + 1)}{K(\boldsymbol{\gamma}, \boldsymbol{\xi}, \alpha; \gamma_0, \beta_0)} \geq \frac{K(\boldsymbol{\gamma}, \boldsymbol{\xi}, \alpha + 1; \gamma_0, \beta_0)}{K(\boldsymbol{\gamma}, \boldsymbol{\xi}, \alpha; \gamma_0, \beta_0)} \frac{K(\boldsymbol{\gamma}, \boldsymbol{\xi}, \alpha + 1; \gamma_0, \beta_0 + 1)}{K(\boldsymbol{\gamma}, \boldsymbol{\xi}, \alpha; \gamma_0, \beta_0)}. \tag{2.10}$$

For $\boldsymbol{v} = (v_1, \ldots, v_m) \in \mathbb{R}^m$ and $\tilde{\boldsymbol{v}} = (\tilde{v}_1, \ldots, \tilde{v}_m) \in \mathbb{R}^m$, we write $\boldsymbol{v} \circ \tilde{\boldsymbol{v}} = (v_1 \tilde{v}_1, \ldots, v_m \tilde{v}_m)$. Let $\mathbb{N}_0 = \{0, 1, 2, \ldots\}$. For $\boldsymbol{x} = (x_1, \ldots, x_m) \in {\mathbb{N}_0}^m$ and $i = 1, \ldots, m$, we define

$$\phi_i^{(\alpha, \boldsymbol{\beta}, \boldsymbol{\gamma}; \beta_0, \gamma_0)}(\boldsymbol{x}) = \begin{cases} \dfrac{1}{n_i \gamma_i} \dfrac{K(\boldsymbol{n} \circ \boldsymbol{\gamma}, \boldsymbol{x} + \boldsymbol{\beta}, \alpha + \beta_0 + 1; \gamma_0, \beta_0)}{K(\boldsymbol{n} \circ \boldsymbol{\gamma}, \boldsymbol{x} + \boldsymbol{\beta} - \boldsymbol{e}_i, \alpha + \beta_0; \gamma_0, \beta_0)} & \\ \quad \text{if } x_i + \beta_i > 1 \text{ and } \sum_{j=1}^{m} (x_j + \beta_j) > \alpha + 1 & \\ 1 \quad \text{otherwise.} & \end{cases}$$

The following lemma gives an explicit form of the Bayes estimator based on $\pi_{\alpha, \boldsymbol{\beta}, \boldsymbol{\gamma}; \beta_0, \gamma_0}$.

Lemma 2.4. *Suppose $\alpha < \sum_{i=1}^m \beta_i$. Then the estimator $\widehat{\boldsymbol{\lambda}}^{(\alpha,\boldsymbol{\beta},\boldsymbol{\gamma};\beta_0,\gamma_0)}$ defined by*

$$\left(\{1-\phi_1^{(\alpha,\boldsymbol{\beta},\boldsymbol{\gamma};\beta_0,\gamma_0)}(\boldsymbol{X})\}\frac{X_1+\beta_1-1}{n_1},\dots,\{1-\phi_m^{(\alpha,\boldsymbol{\beta},\boldsymbol{\gamma};\beta_0,\gamma_0)}(\boldsymbol{X})\}\frac{X_m+\beta_m-1}{n_m}\right) \quad (2.11)$$

is the unique Bayes estimator of $\boldsymbol{\lambda}$ on the basis of $\boldsymbol{X}$ against the prior $\pi_{\alpha,\boldsymbol{\beta},\boldsymbol{\gamma};\beta_0,\gamma_0}$ under the loss function $L_{\boldsymbol{c}}$ given by (1.5).

It is worth noting that the Bayes estimator $\widehat{\boldsymbol{\lambda}}^{(\alpha,\boldsymbol{\beta},\boldsymbol{\gamma};\beta_0,\gamma_0)}$ is robust in the sense that it does not depend on $\boldsymbol{c}$. We remark that the estimator with $\boldsymbol{\beta}=\boldsymbol{j}$ shrinks the ML estimator toward the origin.

Let $\boldsymbol{\nu}=(1/n_1,\dots,1/n_m)$ be the vector whose elements are the reciprocals of the sample sizes, so that $\boldsymbol{n}\circ\boldsymbol{\nu}=\boldsymbol{j}$. Suppose that $m\ge 2$. Then the Bayes estimator with $\alpha=m-1$, $\boldsymbol{\beta}=\boldsymbol{j}$, $\boldsymbol{\gamma}=\boldsymbol{\nu}$, and $\gamma_0=1$, namely $\widehat{\boldsymbol{\lambda}}^{(m-1,\boldsymbol{j},\boldsymbol{\nu};\beta_0,1)}$, reduces to (1.2) by (2.4). Thus, (1.2) is the Bayes estimator against the prior

$$\pi_{m-1,\boldsymbol{j},\boldsymbol{\nu};\beta_0,1}(\boldsymbol{\lambda})=\frac{1}{(\sum_{i=1}^m n_i\lambda_i)^{m-1}}\int_0^\infty \frac{u^{m-2+\beta_0}}{(u+\sum_{i=1}^m n_i\lambda_i)^{\beta_0}}e^{-u}\,du.$$

In the context of shrinkage estimation, however, this choice of prior may be inappropriate since it depends on $\boldsymbol{n}$ and puts less weight on the smaller values of λ_i than on the smaller values of λ_j for all $i,j=1,\dots,m$ such that $n_i<n_j$. Indeed, the shrinkage factor of the resulting Bayes estimator (1.2) fails to reflect the fact that the sample size $\boldsymbol{n}$ is unbalanced.

Finally, we propose an estimator of the form (1.4) which shrinks the ML estimator X_i/n_i more toward the origin than X_j/n_j for all $i,j=1,\dots,m$ such that $n_i<n_j$. We consider the case where $\boldsymbol{\beta}=\boldsymbol{\gamma}=\boldsymbol{j}$ and $\alpha<m$ for $\boldsymbol{j}=(1,\dots,1)\in\mathbb{R}^m$. Then the prior is

$$\pi_{\alpha,\boldsymbol{j},\boldsymbol{j};\beta_0,\gamma_0}(\boldsymbol{\lambda})=\frac{1}{(\sum_{i=1}^m\lambda_i)^{\alpha}}\int_0^\infty\frac{u^{\alpha-1+\beta_0}}{(u/\gamma_0+\sum_{i=1}^m\lambda_i)^{\beta_0}}e^{-u}\,du,$$

which is a shrinkage prior symmetric in $\lambda_1,\dots,\lambda_m$. The resulting estimator can be expressed as

$$\widehat{\boldsymbol{\lambda}}^{(\alpha,\boldsymbol{j},\boldsymbol{j};\beta_0,\gamma_0)}=\left(\{1-\phi_1^{(\alpha,\boldsymbol{j},\boldsymbol{j};\beta_0,\gamma_0)}(\boldsymbol{X})\}\frac{X_1}{n_1},\dots,\{1-\phi_m^{(\alpha,\boldsymbol{j},\boldsymbol{j};\beta_0,\gamma_0)}(\boldsymbol{X})\}\frac{X_m}{n_m}\right), \quad (2.12)$$

where

$$\phi_i^{(\alpha,\boldsymbol{j},\boldsymbol{j};\beta_0,\gamma_0)}(\boldsymbol{x})=\begin{cases}\dfrac{1}{n_i}\dfrac{K(\boldsymbol{n},\boldsymbol{x}+\boldsymbol{j},\alpha+\beta_0+1;\gamma_0,\beta_0)}{K(\boldsymbol{n},\boldsymbol{x}+\boldsymbol{j}-\boldsymbol{e}_i,\alpha+\beta_0;\gamma_0,\beta_0)} & \text{if } x_i\ge 1\\ 1 & \text{if } x_i=0\end{cases} \quad (2.13)$$

for $\boldsymbol{x}=(x_1,\dots,x_m)\in\mathbb{N}_0{}^m$ and $i=1,\dots,m$. This shrinkage estimator has the following heterogeneity properties.

Lemma 2.5. *Let $\boldsymbol{x}=(x_1,\dots,x_m)\in\mathbb{N}_0{}^m$ and suppose $\alpha<m$.*

(i) *Let $i\in\{1,\dots,m\}$. Then*

$$0<\phi_i^{(\alpha,\boldsymbol{j},\boldsymbol{j};\beta_0,\gamma_0)}(\boldsymbol{x})\le 1. \quad (2.14)$$

Equality holds if and only if $x_i=0$.

(ii) *Let $i,j\in\{1,\dots,m\}$ and suppose $x_i,x_j\ge 1$. Then*

$$\phi_i^{(\alpha,\boldsymbol{j},\boldsymbol{j};\beta_0,\gamma_0)}(\boldsymbol{x})>\phi_j^{(\alpha,\boldsymbol{j},\boldsymbol{j};\beta_0,\gamma_0)}(\boldsymbol{x}) \quad \text{if and only if} \quad n_i<n_j. \quad (2.15)$$

(iii) *Let* $i \in \{1, \ldots, m\}$ *and suppose* $x_i \geq 1$. *Suppose further that* $\alpha < m - 2$. *Then*

$$\lim_{n_i \to \infty} \phi_i^{(\alpha, \boldsymbol{j}, \boldsymbol{j}; \beta_0, \gamma_0)}(\boldsymbol{x}) = 0. \tag{2.16}$$

In the case where $\boldsymbol{n} = \boldsymbol{j}$ and $\alpha + 1 = m \geq 2$ and $\gamma_0 = 1$, both of the estimators (1.2) and (2.12) coincide with the estimator (1.1) given by Clevenson and Zidek (1975). However, we propose the latter as an important generalization of (1.1) which satisfies the heterogeneity properties (2.14), (2.15), and (2.16).

3 Sufficient conditions for minimaxity

In this section, we derive sufficient conditions for the estimator $\widehat{\boldsymbol{\lambda}}^{(\alpha, \boldsymbol{\beta}, \boldsymbol{\gamma}; \beta_0, \gamma_0)}$ given by (2.11) to be minimax under the loss function $L_{\boldsymbol{c}}$ given by (1.5). Since it can be shown that the ML estimator $\widehat{\boldsymbol{\lambda}}^{\mathrm{ML}} = (X_1/n_1, \ldots, X_m/n_m)$ is the constant risk minimax estimator, it suffices to find conditions under which $\widehat{\boldsymbol{\lambda}}^{(\alpha, \boldsymbol{\beta}, \boldsymbol{\gamma}; \beta_0, \gamma_0)}$ dominates $\widehat{\boldsymbol{\lambda}}^{\mathrm{ML}}$. Hereafter, we restrict our attention to the case of $\alpha < m$ and $\boldsymbol{\beta} = \boldsymbol{j}$ and consider the shrinkage estimator

$$\widehat{\boldsymbol{\lambda}}^{(\alpha, \boldsymbol{j}, \boldsymbol{\gamma}; \beta_0, \gamma_0)} = \left(\{1 - \phi_1^{(\alpha, \boldsymbol{j}, \boldsymbol{\gamma}; \beta_0, \gamma_0)}(\boldsymbol{X})\} \frac{X_1}{n_1}, \ldots, \{1 - \phi_m^{(\alpha, \boldsymbol{j}, \boldsymbol{\gamma}; \beta_0, \gamma_0)}(\boldsymbol{X})\} \frac{X_m}{n_m} \right), \tag{3.1}$$

where

$$\phi_i^{(\alpha, \boldsymbol{j}, \boldsymbol{\gamma}; \beta_0, \gamma_0)}(\boldsymbol{x}) = \begin{cases} \dfrac{1}{n_i \gamma_i} \dfrac{K(\boldsymbol{n} \circ \boldsymbol{\gamma}, \boldsymbol{x} + \boldsymbol{j}, \alpha + \beta_0 + 1; \gamma_0, \beta_0)}{K(\boldsymbol{n} \circ \boldsymbol{\gamma}, \boldsymbol{x} + \boldsymbol{j} - \boldsymbol{e}_i, \alpha + \beta_0; \gamma_0, \beta_0)} & \text{if } x_i \geq 1 \\ 1 & \text{if } x_i = 0 \end{cases}$$

for $\boldsymbol{x} = (x_1, \ldots, x_m) \in \mathbb{N}_0{}^m$ and $i = 1, \ldots, m$.

The following result, due to Hudson (1978), is used in the proof of Theorem 3.1 below.

Lemma 3.1. *Let* $h \colon \mathbb{N}_0{}^m \to \mathbb{R}$ *and suppose that* $E_{\boldsymbol{\lambda}}[|h(\boldsymbol{X})|] < \infty$. *Then for all* $i = 1, \ldots, m$, *if* $h(\boldsymbol{x}) = 0$ *for all* $\boldsymbol{x} = (x_1, \ldots, x_m) \in \mathbb{N}_0{}^m$ *such that* $x_i = 0$, *we have*

$$E_{\boldsymbol{\lambda}}\left[\frac{h(\boldsymbol{X})}{n_i \lambda_i}\right] = E_{\boldsymbol{\lambda}}\left[\frac{h(\boldsymbol{X} + \boldsymbol{e}_i)}{X_i + 1}\right].$$

For simplicity of notation, we let $a_i = n_i \gamma_i$ and $C_i = (c_i/n_i)(1/a_i) = c_i/(n_i{}^2 \gamma_i)$ for $i = 1, \ldots, m$ and let $\underline{a} = \min_{1 \leq i \leq m} a_i$, $\overline{a} = \max_{1 \leq i \leq m} a_i$, $\underline{C} = \min_{1 \leq i \leq m} C_i$, $\overline{C} = \max_{1 \leq i \leq m} C_i$, and $C_\cdot = \sum_{i=1}^m C_i$. The following theorem, which will be proved later in this section, gives two sufficient conditions for the minimaxity of $\widehat{\boldsymbol{\lambda}}^{(\alpha, \boldsymbol{j}, \boldsymbol{\gamma}; \beta_0, \gamma_0)}$.

Theorem 3.1. *Assume that* $\alpha < m$ *and that* $\gamma_0 \leq \underline{a}$.

(i) *Suppose that*

$$\alpha + \beta_0 \leq \frac{2}{3} \left(\frac{C_\cdot}{\overline{C}} - 1 \right) \left(\frac{\beta_0}{m} + 1 \right). \tag{3.2}$$

Then the estimator $\widehat{\boldsymbol{\lambda}}^{(\alpha, \boldsymbol{j}, \boldsymbol{\gamma}; \beta_0, \gamma_0)}$ *is minimax under the loss* $L_{\boldsymbol{c}}$.

(ii) *Let* $\rho = \{\overline{C}(\alpha + \beta_0 + 1) - C_\cdot\}/\{\underline{C}(\alpha + \beta_0)\}$. *Suppose that* $0 \leq \rho \leq 1 - (1/2)(\overline{a}/\underline{a})$ *and that*

$$2\rho \left(\beta_0 + m + \frac{\underline{a}}{\overline{a}} \right) \leq \frac{\overline{C}}{\underline{C}} (\alpha + 2\beta_0 + 1) - \frac{C_\cdot}{\underline{C}} + 2 \frac{\underline{a}}{\overline{a}} - 1. \tag{3.3}$$

Then the estimator $\widehat{\boldsymbol{\lambda}}^{(\alpha, \boldsymbol{j}, \boldsymbol{\gamma}; \beta_0, \gamma_0)}$ *is minimax under the loss* $L_{\boldsymbol{c}}$.

Part (i) of the above theorem is a generalization of Theorem 3 of Hamura and Kubokawa (2019). They consider the case of $\beta_0 = 0$. In this case, the prior is improper by Lemma 2.1 but whenever $m \geq 2$, there is always a value of $\alpha > 0$ that satisfies the sufficient condition (3.2) for the minimaxity of the estimator $\widehat{\boldsymbol{\lambda}}^{(\alpha,\boldsymbol{j},\boldsymbol{\gamma};0,\underline{a})}$. On the other hand, when the prior is proper, assumption (3.2) implies $m < (2/3)(C_\cdot/\overline{C} - 1)(\beta_0/m + 1) \leq 2(C_\cdot/\overline{C} - 1)$. Therefore, there exist $C_1, \dots, C_m$ such that the condition (3.2) is violated for any choice of a proper prior.

Let $\underline{n} = \min_{1 \leq i \leq m} n_i$ and $\overline{n} = \max_{1 \leq i \leq m} n_i$. Let $C_i^* = c_i/{n_i}^2$ for $i = 1, \dots, m$ and define $\underline{C}^*$, $\overline{C}^*$, and $C_\cdot^*$ analogously. Combining Lemmas 2.1, 2.4, and 2.5 and Theorem 3.1, we obtain the following theorem.

Theorem 3.2. *Suppose that $\alpha < m < \alpha + \beta_0$ and that $\gamma_0 \leq \underline{n}$. Suppose further that one of the following two conditions holds*:

(i)

$$\alpha + \beta_0 \leq \frac{2}{3}\left(\frac{C_\cdot^*}{\overline{C}^*} - 1\right)\left(\frac{\beta_0}{m} + 1\right).$$

(ii)

$$\frac{\overline{C}^*(\alpha + \beta_0 + 1) - C_\cdot^*}{\underline{C}^*(\alpha + \beta_0)} \leq \min\left\{1 - \frac{1}{2}\frac{\overline{n}}{\underline{n}}, \frac{(\overline{C}^*/\underline{C}^*)(\alpha + 2\beta_0 + 1) - C_\cdot^*/\underline{C}^* + 2\underline{n}/\overline{n} - 1}{2(\beta_0 + m + \underline{n}/\overline{n})}\right\}.$$

Then the estimator $\widehat{\boldsymbol{\lambda}}^{(\alpha,\boldsymbol{j},\boldsymbol{j};\beta_0,\gamma_0)}$ given by (2.12) *is admissible and minimax under the loss $L_{\boldsymbol{c}}$. Furthermore, for all $\boldsymbol{x} = (x_1, \dots, x_m) \in {\mathbb{N}_0}^m$ and $i, j \in \{1, \dots, m\}$ such that $x_i, x_j \geq 1$, it satisfies* (2.14), (2.15), *and, if $\alpha < m - 2$,* (2.16).

It can be seen that there exists an admissible minimax shrinkage estimator that satisfies (2.14), (2.15), and (2.16) by, for example, applying part (i) of Theorem 3.2 to the case where $(\alpha, \beta_0, \gamma_0) = (1, m, \underline{n})$ and m is sufficiently large and $\underline{C}^*/\overline{C}^*$ is sufficiently close to 1. Furthermore, though the details are omitted here, it can be shown from part (i) of Theorem 3.2 that there exists $\alpha > 0$, $\beta_0 \geq 0$, and $\gamma_0 > 0$ such that the conclusion of Theorem 3.2 holds if $2 \leq m < (4/3)(C_\cdot^*/\overline{C}^* - 1)$. This condition reduces to

$$2 \leq m < \frac{4}{3}\left(\sum_{i=1}^{m} \frac{\underline{n}^k}{{n_i}^k} - 1\right)$$

with $k = 1$ when $c_i = n_i$ and with $k = 2$ when $c_i = 1$.

In the particular case of $\boldsymbol{n} = \boldsymbol{c} = \boldsymbol{\gamma} = \boldsymbol{j}$ and $\gamma_0 = 1$, the condition for $\widehat{\boldsymbol{\lambda}}^{(\alpha,\boldsymbol{j},\boldsymbol{j};\beta_0,\gamma_0)}$ to be admissible and minimax given in part (i) of Theorem 3.2 is

$$\alpha < m < \alpha + \beta_0 \quad \text{and} \quad \alpha + \left(1 - \frac{2}{3}\frac{m-1}{m}\right)\beta_0 \leq \frac{2}{3}(m-1), \tag{3.4}$$

whereas that given in part (ii) of Theorem 3.2 is

$$\alpha < m < \alpha + \beta_0 \leq 2(m-1) \quad \text{and} \quad \frac{\alpha + \beta_0 + 1 - m}{\alpha + \beta_0} \leq \frac{\alpha + 2\beta_0 - m + 2}{2(\beta_0 + m + 1)}. \tag{3.5}$$

Conditions (3.4) and (3.5) correspond to (3.2) and (3.3), respectively. The condition given by Clevenson and Zidek (1975) is

$$\alpha = m - 1 \quad \text{and} \quad 1 < \beta_0 \leq m - 1.$$

When $m \geq 2$ and $\alpha = m-1$, condition (3.4) is not satisfied for any values of β_0 but condition (3.5) becomes

$$1 < \beta_0 \leq (m-1)/3.$$

Thus, although the result of Clevenson and Zidek (1975) is not completely included, Theorem 3.1 or 3.2, which was derived for estimating $\boldsymbol{\lambda}$ when $\boldsymbol{n}$ is unbalanced, gives the sufficient condition which is close to that of Clevenson and Zidek (1975) even in the case of balanced sample sizes.

Proof of Theorem 3.1. Let $\Delta = E_{\boldsymbol{\lambda}}[L_{\boldsymbol{c}}(\widehat{\boldsymbol{\lambda}}^{(\alpha,\boldsymbol{j},\boldsymbol{\gamma};\beta_0,\gamma_0)}, \boldsymbol{\lambda})] - E_{\boldsymbol{\lambda}}[L_{\boldsymbol{c}}(\widehat{\boldsymbol{\lambda}}^{\mathrm{ML}}, \boldsymbol{\lambda})]$. From (3.1),

$$\Delta = E_{\boldsymbol{\lambda}}\Bigg[\sum_{i=1}^{m}\Bigg[\frac{c_i}{\lambda_i}\bigg\{\frac{X_i}{n_i} - \lambda_i - \frac{X_i}{n_i}\phi_i^{(\alpha,\boldsymbol{j},\boldsymbol{\gamma};\beta_0,\gamma_0)}(\boldsymbol{X})\bigg\}^2 - \frac{c_i}{\lambda_i}\bigg(\frac{X_i}{n_i} - \lambda_i\bigg)^2\Bigg]\Bigg]$$

$$= E_{\boldsymbol{\lambda}}\Bigg[\sum_{i=1}^{m}\bigg(\frac{n_i c_i}{n_i\lambda_i}\bigg[\bigg\{\frac{X_i}{n_i}\phi_i^{(\alpha,\boldsymbol{j},\boldsymbol{\gamma};\beta_0,\gamma_0)}(\boldsymbol{X})\bigg\}^2 - 2\bigg(\frac{X_i}{n_i}\bigg)^2\phi_i^{(\alpha,\boldsymbol{j},\boldsymbol{\gamma};\beta_0,\gamma_0)}(\boldsymbol{X})\bigg]$$

$$+ 2c_i\frac{X_i}{n_i}\phi_i^{(\alpha,\boldsymbol{j},\boldsymbol{\gamma};\beta_0,\gamma_0)}(\boldsymbol{X})\bigg)\Bigg],$$

which is, by application of Lemma 3.1,

$$\Delta = E_{\boldsymbol{\lambda}}\Bigg[\sum_{i=1}^{m}\bigg(\frac{n_i c_i}{X_i+1}\bigg[\bigg\{\frac{X_i+1}{n_i}\phi_i^{(\alpha,\boldsymbol{j},\boldsymbol{\gamma};\beta_0,\gamma_0)}(\boldsymbol{X}+\boldsymbol{e}_i)\bigg\}^2$$

$$- 2\bigg(\frac{X_i+1}{n_i}\bigg)^2\phi_i^{(\alpha,\boldsymbol{j},\boldsymbol{\gamma};\beta_0,\gamma_0)}(\boldsymbol{X}+\boldsymbol{e}_i)\bigg] + 2c_i\frac{X_i}{n_i}\phi_i^{(\alpha,\boldsymbol{j},\boldsymbol{\gamma};\beta_0,\gamma_0)}(\boldsymbol{X})\bigg)\Bigg].$$

Therefore, we can write the risk difference as $\Delta = E_{\boldsymbol{\lambda}}[I_1(\boldsymbol{X}) - 2I_2(\boldsymbol{X}) + 2I_3(\boldsymbol{X})]$, where

$$I_1(\boldsymbol{x}) = \sum_{i=1}^{m}\frac{c_i}{n_i}(x_i+1)\bigg\{\frac{1}{n_i\gamma_i}\frac{K(\boldsymbol{n}\circ\boldsymbol{\gamma}, \boldsymbol{x}+\boldsymbol{j}+\boldsymbol{e}_i, \alpha+\beta_0+1; \gamma_0, \beta_0)}{K(\boldsymbol{n}\circ\boldsymbol{\gamma}, \boldsymbol{x}+\boldsymbol{j}, \alpha+\beta_0; \gamma_0, \beta_0)}\bigg\}^2,$$

$$I_2(\boldsymbol{x}) = \sum_{i=1}^{m}\frac{c_i}{n_i}\frac{x_i+1}{n_i\gamma_i}\frac{K(\boldsymbol{n}\circ\boldsymbol{\gamma}, \boldsymbol{x}+\boldsymbol{j}+\boldsymbol{e}_i, \alpha+\beta_0+1; \gamma_0, \beta_0)}{K(\boldsymbol{n}\circ\boldsymbol{\gamma}, \boldsymbol{x}+\boldsymbol{j}, \alpha+\beta_0; \gamma_0, \beta_0)},$$

$$I_3(\boldsymbol{x}) = \begin{cases} 0 & \text{if } \boldsymbol{x} = \boldsymbol{0} \\ \displaystyle\sum_{i=1}^{m}\frac{c_i}{n_i}\frac{x_i}{n_i\gamma_i}\frac{K(\boldsymbol{n}\circ\boldsymbol{\gamma}, \boldsymbol{x}+\boldsymbol{j}, \alpha+\beta_0+1; \gamma_0, \beta_0)}{K(\boldsymbol{n}\circ\boldsymbol{\gamma}, \boldsymbol{x}+\boldsymbol{j}-\boldsymbol{e}_i, \alpha+\beta_0; \gamma_0, \beta_0)} & \text{otherwise,} \end{cases}$$

for $\boldsymbol{x} = (x_1, \ldots, x_m) \in \mathbb{N}_0{}^m$. We have $I_1(\boldsymbol{0}) - 2I_2(\boldsymbol{0}) + 2I_3(\boldsymbol{0}) < 0$ since

$$\frac{1}{n_i\gamma_i}\frac{K(\boldsymbol{n}\circ\boldsymbol{\gamma}, \boldsymbol{j}+\boldsymbol{e}_i, \alpha+\beta_0+1; \gamma_0, \beta_0)}{K(\boldsymbol{n}\circ\boldsymbol{\gamma}, \boldsymbol{j}, \alpha+\beta_0; \gamma_0, \beta_0)} \in [0, 1].$$

Thus, it is sufficient to show that $I_1(\boldsymbol{x}) - 2I_2(\boldsymbol{x}) + 2I_3(\boldsymbol{x}) \leq 0$ for all $\boldsymbol{x} \in \mathbb{N}_0{}^m \setminus \{\boldsymbol{0}\}$.

Fix $\boldsymbol{x} = (x_1, \ldots, x_m) \in \mathbb{N}_0{}^m \setminus \{\boldsymbol{0}\}$. Hereafter, for simplicity, we use the abbreviated notation

$$I_1 = I_1(\boldsymbol{x}), \qquad I_2 = I_2(\boldsymbol{x}), \qquad I_3 = I_3(\boldsymbol{x}),$$

$$I = I_1 - 2I_2 + 2I_3,$$

$$H(c) = \frac{K(\boldsymbol{n}\circ\boldsymbol{\gamma}, \boldsymbol{x}+\boldsymbol{j}, \alpha+\beta_0+c; \gamma_0, \beta_0)}{K(\boldsymbol{n}\circ\boldsymbol{\gamma}, \boldsymbol{x}+\boldsymbol{j}, \alpha+\beta_0; \gamma_0, \beta_0)},$$

$$H(0, c) = \frac{K(\boldsymbol{n}\circ\boldsymbol{\gamma}, \boldsymbol{x}+\boldsymbol{j}, \alpha+\beta_0+c; \gamma_0, \beta_0+1)}{K(\boldsymbol{n}\circ\boldsymbol{\gamma}, \boldsymbol{x}+\boldsymbol{j}, \alpha+\beta_0; \gamma_0, \beta_0)},$$

and

$$H(\pm i, c) = \frac{K(\boldsymbol{n} \circ \boldsymbol{\gamma}, \boldsymbol{x} + \boldsymbol{j} \pm \boldsymbol{e}_i, \alpha + \beta_0 + c; \gamma_0, \beta_0)}{K(\boldsymbol{n} \circ \boldsymbol{\gamma}, \boldsymbol{x} + \boldsymbol{j}, \alpha + \beta_0; \gamma_0, \beta_0)}$$

for $c = 0, 1, 2$ and $i = 1, \dots, m$ when well defined.

For part (i), we have

$$I_1 \le \overline{C} H(1) \sum_{i=1}^{m} \frac{x_i + 1}{a_i} H(i, 1).$$

By part (ii) of Lemma 2.2, we obtain

$$I_2 = \sum_{i=1}^{m} \frac{c_i}{n_i} \frac{x_i + 1}{a_i} H(1) - \sum_{i=1}^{m} C_i \frac{x_i + 1}{a_i} H(i, 2)$$

and

$$\begin{aligned} I_3 &= \sum_{i=1}^{m} \frac{c_i}{n_i} \frac{x_i}{a_i} \left[H(1) - \left\{ H(1) - \frac{H(1)}{H(-i, 0)} \right\} \right] \\ &= \sum_{i=1}^{m} \frac{c_i}{n_i} \frac{x_i}{a_i} H(1) - \sum_{i=1}^{m} \frac{c_i}{n_i} \frac{x_i}{a_i} \frac{1}{a_i} \frac{H(1)}{H(-i, 0)} H(1). \end{aligned}$$

Then, from part (i) of Lemma 2.3,

$$\begin{aligned} I_3 &\le \sum_{i=1}^{m} \frac{c_i}{n_i} \frac{x_i}{a_i} H(1) - \sum_{i=1}^{m} \frac{c_i}{n_i} \frac{x_i}{a_i} \frac{1}{a_i} H(i, 1) H(1) \qquad (3.6) \\ &\le \sum_{i=1}^{m} \frac{c_i}{n_i} \frac{x_i}{a_i} H(1) - \sum_{i=1}^{m} \frac{c_i}{n_i} \frac{x_i + 1}{a_i} \frac{1}{a_i} H(i, 1) H(1) + \overline{C} \sum_{i=1}^{m} \frac{1}{a_i} H(i, 1) H(1). \end{aligned}$$

Therefore,

$$\begin{aligned} I &\le \overline{C} H(1) \sum_{i=1}^{m} \frac{x_i + 1}{a_i} H(i, 1) - 2C.H(1) \\ &\quad + 2 \sum_{i=1}^{m} C_i \frac{x_i + 1}{a_i} \{ H(i, 2) - H(i, 1) H(1) \} + 2\overline{C} \sum_{i=1}^{m} \frac{1}{a_i} H(i, 1) H(1) \\ &\le \overline{C} H(1) \sum_{i=1}^{m} \frac{x_i + 1}{a_i} H(i, 1) - 2C.H(1) \\ &\quad + 2\overline{C} \sum_{i=1}^{m} \frac{x_i + 1}{a_i} \{ H(i, 2) - H(i, 1) H(1) \} + 2\overline{C} \sum_{i=1}^{m} \frac{1}{a_i} H(i, 1) H(1), \end{aligned}$$

where the second inequality follows since $H(i, 2) - H(i, 1) H(1) \ge 0$ for all $i = 1, \dots, m$ by part (ii) of Lemma 2.3. By applying part (i) of Lemma 2.2, we obtain

$$\begin{aligned} I &\le \overline{C} H(1) \left\{ \alpha + \beta_0 - \frac{\beta_0}{\gamma_0} H(0, 1) \right\} - 2C.H(1) + 2\overline{C} \sum_{i=1}^{m} \frac{1}{a_i} H(i, 1) H(1) \\ &\quad + 2\overline{C} \left\{ (\alpha + \beta_0 + 1) H(1) - \frac{\beta_0}{\gamma_0} H(0, 2) - (\alpha + \beta_0) H(1) + \frac{\beta_0}{\gamma_0} H(0, 1) H(1) \right\} \end{aligned}$$

$$\le \overline{C}H(1)\left\{\alpha+\beta_0-\sum_{i=1}^{m}\frac{\beta_0/m}{a_i}H(i,1)\right\}-2C.H(1)$$
$$+2\overline{C}\sum_{i=1}^{m}\frac{1}{a_i}H(i,1)H(1)+2\overline{C}H(1)$$
$$=\{\overline{C}(\alpha+\beta_0+2)-2C.\}H(1)+\overline{C}\sum_{i=1}^{m}\frac{2-\beta_0/m}{a_i}H(i,1)H(1),$$

where the second inequality follows from the assumption that $\gamma_0 \le \underline{a}$ and part (ii) of Lemma 2.3. Since

$$\alpha+\beta_0=\sum_{i=1}^{m}\left\{\frac{x_i+1}{a_i}H(i,1)+\frac{\beta_0/m}{\gamma_0}H(0,1)\right\}\ge\sum_{i=1}^{m}\frac{1+\beta_0/m}{a_i}H(i,1)$$

by part (i) of Lemma 2.2 and the assumption that $\gamma_0 \le \underline{a}$ and since assumption (3.2) implies $2-\beta_0/m \ge 0$, we conclude that

$$I/H(1)\le\overline{C}(\alpha+\beta_0+2)-2C.+\overline{C}\frac{2-\beta_0/m}{1+\beta_0/m}(\alpha+\beta_0)\le 0,$$

where the second inequality follows from assumption (3.2). This completes the proof of part (i).

For part (ii), let $\underline{i}\in\{1,\dots,m\}$ be an index such that $a_{\underline{i}}=\underline{a}$. Then we have

$$I_1-2I_2\le\left\{\frac{1}{\underline{a}}H(\underline{i},1)-2\right\}I_2. \tag{3.7}$$

Note that, by part (ii) and part (i) of Lemma 2.2,

$$I_2=\sum_{i=1}^{m}C_i(x_i+1)H(1)-\sum_{i=1}^{m}C_i(x_i+1)\frac{1}{a_i}H(i,2)$$
$$\ge\sum_{i=1}^{m}C_i(x_i+1)H(1)-\overline{C}\sum_{i=1}^{m}(x_i+1)\frac{1}{a_i}H(i,2)$$
$$=\left\{\sum_{i=1}^{m}C_i(x_i+1)-\overline{C}(\alpha+\beta_0+1)\right\}H(1)+\overline{C}\frac{\beta_0}{\gamma_0}H(0,2). \tag{3.8}$$

Since $(1/\underline{a})H(\underline{i},1)\le 1$, it follows from (3.7) and (3.8) that

$$I_1-2I_2\le\left\{\frac{1}{\underline{a}}H(\underline{i},1)-2\right\}$$
$$\times\left[\left\{\sum_{i=1}^{m}C_i(x_i+1)-\overline{C}(\alpha+\beta_0+1)\right\}H(1)+\overline{C}\frac{\beta_0}{\gamma_0}H(0,2)\right]$$
$$=-\left\{2-\frac{1}{\underline{a}}H(\underline{i},1)\right\}\sum_{i=1}^{m}C_ix_iH(1)-\left\{2-\frac{1}{\underline{a}}H(\underline{i},1)\right\}\overline{C}\frac{\beta_0}{\gamma_0}H(0,2)$$
$$+\left\{2-\frac{1}{\underline{a}}H(\underline{i},1)\right\}\{\overline{C}(\alpha+\beta_0+1)-C.\}H(1). \tag{3.9}$$

Combining (3.6) and (3.9) gives

$$I \le \frac{1}{\underline{a}} H(\underline{i}, 1) \sum_{i=1}^{m} C_i x_i H(1) - \Big\{2 - \frac{1}{\underline{a}} H(\underline{i}, 1)\Big\} \overline{C} \frac{\beta_0}{\gamma_0} H(0, 2)$$
$$+ \Big\{2 - \frac{1}{\underline{a}} H(\underline{i}, 1)\Big\} \{\overline{C}(\alpha + \beta_0 + 1) - C.\} H(1) - 2 \sum_{i=1}^{m} C_i \frac{x_i}{a_i} H(i, 1) H(1). \quad (3.10)$$

Note that

$$\sum_{i=1}^{m} C_i \frac{x_i}{a_i} H(i, 1) \ge \frac{\underline{a}}{\overline{a}} \frac{1}{\underline{a}} H(\underline{i}, 1) \sum_{i=1}^{m} C_i x_i$$

and that

$$\sum_{i=1}^{m} C_i \frac{x_i}{a_i} H(i, 1) \ge \underline{C} \sum_{i=1}^{m} \frac{x_i + 1}{a_i} H(i, 1) - \underline{C} \sum_{i=1}^{m} \frac{1}{a_i} H(i, 1)$$
$$= \underline{C}(\alpha + \beta_0) - \underline{C} \frac{\beta_0}{\gamma_0} H(0, 1) - \underline{C} \sum_{i=1}^{m} \frac{1}{a_i} H(i, 1) \quad (3.11)$$

by part (i) of Lemma 2.2. Then we have

$$I \le \frac{1}{\underline{a}} H(\underline{i}, 1) \sum_{i=1}^{m} C_i x_i H(1) - \Big\{2 - \frac{1}{\underline{a}} H(\underline{i}, 1)\Big\} \overline{C} \frac{\beta_0}{\gamma_0} H(0, 2)$$
$$+ \Big\{2 - \frac{1}{\underline{a}} H(\underline{i}, 1)\Big\} \{\overline{C}(\alpha + \beta_0 + 1) - C.\} H(1)$$
$$- 2\Big(1 - \frac{1}{2}\frac{\overline{a}}{\underline{a}} - \rho\Big) \sum_{i=1}^{m} C_i \frac{x_i}{a_i} H(i, 1) H(1) - 2 \frac{1}{2} \frac{\overline{a}}{\underline{a}} \frac{\underline{a}}{\overline{a}} \frac{1}{\underline{a}} H(\underline{i}, 1) \sum_{i=1}^{m} C_i x_i H(1)$$
$$- 2\rho \Big\{\underline{C}(\alpha + \beta_0) - \underline{C} \frac{\beta_0}{\gamma_0} H(0, 1) - \underline{C} \sum_{i=1}^{m} \frac{1}{a_i} H(i, 1)\Big\} H(1)$$
$$= -\Big\{2 - \frac{1}{\underline{a}} H(\underline{i}, 1)\Big\} \overline{C} \frac{\beta_0}{\gamma_0} H(0, 2) - \{\overline{C}(\alpha + \beta_0 + 1) - C.\} \frac{1}{\underline{a}} H(\underline{i}, 1) H(1)$$
$$+ 2\rho \underline{C} \frac{\beta_0}{\gamma_0} H(0, 1) H(1) + 2\rho \underline{C} \sum_{i=1}^{m} \frac{1}{a_i} H(i, 1) H(1)$$
$$- 2\Big(1 - \frac{1}{2}\frac{\overline{a}}{\underline{a}} - \rho\Big) \sum_{i=1}^{m} C_i \frac{x_i}{a_i} H(i, 1) H(1)$$
$$\le -\Big\{2 - \frac{1}{\underline{a}} H(\underline{i}, 1)\Big\} \overline{C} \frac{\beta_0}{\gamma_0} H(0, 2) - \{\overline{C}(\alpha + \beta_0 + 1) - C.\} \frac{1}{\underline{a}} H(\underline{i}, 1) H(1)$$
$$+ 2\rho \underline{C} \frac{\beta_0}{\gamma_0} H(0, 1) H(1) + 2\rho \underline{C} m \frac{1}{\underline{a}} H(\underline{i}, 1) H(1)$$
$$- 2\Big(1 - \frac{1}{2}\frac{\overline{a}}{\underline{a}} - \rho\Big) \underline{C} \frac{\underline{a}}{\overline{a}} \frac{1}{\underline{a}} H(\underline{i}, 1) H(1)$$

since $0 \le \rho \le 1 - (1/2)(\overline{a}/\underline{a})$ by assumption and since $\boldsymbol{x} \ne \boldsymbol{0}$. Now since $(1/\underline{a}) H(\underline{i}, 1) \le 1$ and since

$$\frac{1}{\gamma_0} H(0, 2) \ge \frac{1}{\gamma_0} H(0, 1) H(1) \ge \frac{1}{\underline{a}} H(\underline{i}, 1) H(1)$$

by part (ii) of Lemma 2.3, it follows that

$$
\begin{aligned}
&-\Big\{2-\frac{1}{\underline{a}}H(\underline{i},1)\Big\}\overline{C}\frac{\beta_0}{\gamma_0}H(0,2)+2\rho\underline{C}\frac{\beta_0}{\gamma_0}H(0,1)H(1) \\
&\le -\overline{C}\frac{\beta_0}{\gamma_0}H(0,1)H(1)+2\rho\underline{C}\frac{\beta_0}{\gamma_0}H(0,1)H(1) \\
&\le -(\overline{C}-2\rho\underline{C})\frac{\beta_0}{\underline{a}}H(\underline{i},1)H(1), \qquad (3.12)
\end{aligned}
$$

where we have used the fact that $\overline{C}-2\rho\underline{C}\ge\underline{C}(1-2\rho)\ge 0$ by assumption. Thus,

$$
\begin{aligned}
I/\Big\{\frac{1}{\underline{a}}H(\underline{i},1)H(1)\Big\} &\le -(\overline{C}-2\rho\underline{C})\beta_0-\{\overline{C}(\alpha+\beta_0+1)-C.\}+2\rho\underline{C}m \\
&\quad -2\Big(1-\frac{1}{2}\frac{\overline{a}}{\underline{a}}-\rho\Big)\underline{C}\frac{\underline{a}}{\overline{a}}.
\end{aligned}
$$

The right-hand side of the above inequality is not positive by assumption (3.3). This completes the proof of part (ii). □

Remark 3.1. The major difference of the setting considered above from that considered by Hamura and Kubokawa (2019) is that now the parameter β_0 may take on positive values in (3.8), yielding the additional terms in (3.9). In the present setting, we need to evaluate the factor $(1/\gamma_0)H(0,2)$ appropriately. Indeed, if (3.9) is replaced by

$$
\begin{aligned}
I_1-2I_2 &\le -\Big\{2-\frac{1}{\underline{a}}H(\underline{i},1)\Big\}\sum_{i=1}^{m}C_i x_i H(1) \\
&\quad +\Big\{2-\frac{1}{\underline{a}}H(\underline{i},1)\Big\}\{\overline{C}(\alpha+\beta_0+1)-C.\}H(1),
\end{aligned}
$$

then it leads to a sufficient condition that is incompatible with the condition for propriety given in Lemma 2.1. Note also that the term $\overline{C}(\alpha+\beta_0+1)-C.$ is positive if the prior is proper satisfying $\alpha<m<\alpha+\beta_0$, while it is nonpositive in the case they consider. Since $(1/\gamma_0)H(0,2)$ can be very small compared to $H(1)$ in general, it is not straightforward to extend their results to the case of proper priors. We evaluate the third term on the right side of (3.10) by using (3.11), and then apply part (ii) of Lemma 2.3 to the second term in (3.10) in order to evaluate the secondary terms deriving from the last two terms in (3.11). Thus, Lemma 2.3 is important for the above proof of the existence of a heterogeneous shrinkage estimator that is both admissible and minimax.

Remark 3.2. In theory, we can obtain a sufficient condition that generalizes part 4 of Theorem 2.5 of Clevenson and Zidek (1975). By part (i) of Lemma 2.2, we have

$$
\begin{aligned}
\alpha+\beta_0 &= \sum_{i=1}^{m}\frac{x_i+1}{a_i}H(i,1)+\frac{\beta_0}{\gamma_0}H(0,1)\ge(x.+m+\beta_0)\frac{1}{\overline{a}}H(\overline{i},1), \\
(\alpha+\beta_0+1)H(1) &= \sum_{i=1}^{m}\frac{x_i+1}{a_i}H(i,2)+\frac{\beta_0}{\gamma_0}H(0,2)\le(x.+m+\beta_0)\frac{1}{\gamma_0}H(0,2),
\end{aligned}
$$

where $x. = \sum_{i=1}^{m} x_i$ and $\overline{i}\in\{1,\dots,m\}$ is an index such that $a_{\overline{i}}=\overline{a}$. Therefore, it follows that

$$
\frac{1}{\underline{a}}H(\underline{i},1)\le\frac{\overline{a}}{\underline{a}}\frac{\alpha+\beta_0}{x.+m+\beta_0}\le\frac{\overline{a}}{\underline{a}}\frac{\alpha+\beta_0}{1+m+\beta_0}
$$

and that

$$\frac{1}{\gamma_0}H(0,2) \geq \frac{\alpha+\beta_0+1}{\alpha+\beta_0}\frac{1}{\overline{a}}H(\overline{i},1)H(1) \geq \frac{\alpha+\beta_0+1}{\alpha+\beta_0}\frac{\gamma_0}{\overline{a}}\frac{1}{\gamma_0}H(0,1)H(1).$$

Hence, (3.12) can be replaced by

$$\begin{aligned}
&-\Big\{2-\frac{1}{\underline{a}}H(\underline{i},1)\Big\}\overline{C}\frac{\beta_0}{\gamma_0}H(0,2)+2\rho\underline{C}\frac{\beta_0}{\gamma_0}H(0,1)H(1)\\
&\leq -\Big[2-\min\Big\{1,\frac{\overline{a}}{\underline{a}}\frac{\alpha+\beta_0}{1+m+\beta_0}\Big\}\Big]\overline{C}\frac{\beta_0}{\gamma_0}H(0,1)H(1)\max\Big\{1,\frac{\alpha+\beta_0+1}{\alpha+\beta_0}\frac{\gamma_0}{\overline{a}}\Big\}\\
&\quad+2\rho\underline{C}\frac{\beta_0}{\gamma_0}H(0,1)H(1)\\
&=-\Big(\Big[2\overline{C}-\overline{C}\min\Big\{1,\frac{\overline{a}}{\underline{a}}\frac{\alpha+\beta_0}{1+m+\beta_0}\Big\}\Big]\max\Big\{1,\frac{\alpha+\beta_0+1}{\alpha+\beta_0}\frac{\gamma_0}{\overline{a}}\Big\}-2\rho\underline{C}\Big)\\
&\quad\times\frac{\beta_0}{\gamma_0}H(0,1)H(1)\\
&\leq-\Big(\Big[2\overline{C}-\overline{C}\min\Big\{1,\frac{\overline{a}}{\underline{a}}\frac{\alpha+\beta_0}{1+m+\beta_0}\Big\}\Big]\max\Big\{1,\frac{\alpha+\beta_0+1}{\alpha+\beta_0}\frac{\gamma_0}{\overline{a}}\Big\}-2\rho\underline{C}\Big)\\
&\quad\times\frac{\beta_0}{\underline{a}}H(\underline{i},1)H(1),
\end{aligned}$$

which leads to a condition generalizing the sufficient condition of Clevenson and Zidek (1975) for the balanced case.

Remark 3.3. We can also evaluate the risk of the Bayes estimator with respect to the prior $\pi_{\alpha,\boldsymbol{\beta},\boldsymbol{\gamma};\beta_0,\gamma_0}$ under the loss function

$$\tilde{L}_{\boldsymbol{c}}(\boldsymbol{d},\boldsymbol{\lambda})=\sum_{i=1}^m c_i\lambda_i\Big(\frac{d_i}{\lambda_i}-1-\log\frac{d_i}{\lambda_i}\Big)=\sum_{i=1}^m c_i\Big(d_i-\lambda_i-\lambda_i\log\frac{d_i}{\lambda_i}\Big),$$

which is the loss function considered by Ghosh and Yang (1988) for the balanced case and by Hamura and Kubokawa (2019) for the unbalanced case. The Bayes estimator is given by

$$\widetilde{\boldsymbol{\lambda}}^{(\alpha,\boldsymbol{\beta},\boldsymbol{\gamma};\beta_0,\gamma_0)}=\widetilde{\boldsymbol{\lambda}}^{(\boldsymbol{\beta})}\circ(1-\tilde{\phi}_1^{(\alpha,\boldsymbol{\beta},\boldsymbol{\gamma};\beta_0,\gamma_0)}(\boldsymbol{X}),\dots,1-\tilde{\phi}_m^{(\alpha,\boldsymbol{\beta},\boldsymbol{\gamma};\beta_0,\gamma_0)}(\boldsymbol{X}))$$

for $\alpha<\sum_{i=1}^m\beta_i$, where $\widetilde{\boldsymbol{\lambda}}^{(\boldsymbol{\beta})}=((X_1+\beta_1)/n_1,\dots,(X_m+\beta_m)/n_m)$ is the Bayes estimator against the improper prior $\pi_{\boldsymbol{\beta}}(\boldsymbol{\lambda})=\prod_{i=1}^m\lambda_i{}^{\beta_i-1}$ and where

$$\tilde{\phi}_i^{(\alpha,\boldsymbol{\beta},\boldsymbol{\gamma};\beta_0,\gamma_0)}(\boldsymbol{X})=\frac{1}{n_i\gamma_i}\frac{K(\boldsymbol{n}\circ\boldsymbol{\gamma},\boldsymbol{X}+\boldsymbol{\beta}+\boldsymbol{e}_i,\alpha+\beta_0+1;\gamma_0,\beta_0)}{K(\boldsymbol{n}\circ\boldsymbol{\gamma},\boldsymbol{X}+\boldsymbol{\beta},\alpha+\beta_0;\gamma_0,\beta_0)}$$

determines the amount of shrinkage for $i=1,\dots,m$. A calculation similar to that in the proof of Theorem 3.1 shows that the risk difference between the two estimators is

$$E_{\boldsymbol{\lambda}}[\tilde{L}_{\boldsymbol{c}}(\widetilde{\boldsymbol{\lambda}}^{(\alpha,\boldsymbol{\beta},\boldsymbol{\gamma};\beta_0,\gamma_0)},\boldsymbol{\lambda})]-E_{\boldsymbol{\lambda}}[L_{\boldsymbol{c}}(\widetilde{\boldsymbol{\lambda}}^{(\boldsymbol{\beta})},\boldsymbol{\lambda})]=E_{\boldsymbol{\lambda}}[\tilde{D}^{(\alpha,\boldsymbol{\beta},\boldsymbol{\gamma};\beta_0,\gamma_0)}(\boldsymbol{X})],$$

where $\tilde{D}^{(\alpha,\boldsymbol{\beta},\boldsymbol{\gamma};\beta_0,\gamma_0)}(\boldsymbol{0})=-\sum_{i=1}^m C_i\beta_i K(\boldsymbol{n}\circ\boldsymbol{\gamma},\boldsymbol{\beta}+\boldsymbol{e}_i,\alpha+\beta_0+1;\gamma_0,\beta_0)/K(\boldsymbol{n}\circ\boldsymbol{\gamma},\boldsymbol{\beta},\alpha+\beta_0;\gamma_0,\beta_0)$ and

$$\begin{aligned}
\tilde{D}^{(\alpha,\boldsymbol{\beta},\boldsymbol{\gamma};\beta_0,\gamma_0)}(\boldsymbol{x})=&-\sum_{i=1}^m\frac{c_i}{n_i}\frac{x_i+\beta_i}{n_i\gamma_i}\frac{K(\boldsymbol{n}\circ\boldsymbol{\gamma},\boldsymbol{x}+\boldsymbol{\beta}+\boldsymbol{e}_i,\alpha+\beta_0+1;\gamma_0,\beta_0)}{K(\boldsymbol{n}\circ\boldsymbol{\gamma},\boldsymbol{x}+\boldsymbol{\beta},\alpha+\beta_0;\gamma_0,\beta_0)}\\
&+\sum_{i=1}^m\frac{c_i}{n_i}x_i\log\Big\{1+\frac{1}{n_i\gamma_i}\frac{K(\boldsymbol{n}\circ\boldsymbol{\gamma},\boldsymbol{x}+\boldsymbol{\beta},\alpha+\beta_0+1;\gamma_0,\beta_0)}{K(\boldsymbol{n}\circ\boldsymbol{\gamma},\boldsymbol{x}+\boldsymbol{\beta},\alpha+\beta_0;\gamma_0,\beta_0)}\Big\}
\end{aligned}$$

for $\boldsymbol{x} = (x_1, \ldots, x_m) \in \mathbb{N}_0{}^m \setminus \{\boldsymbol{0}\}$. Using Lemma 2.2 to evaluate the first term on the right and applying the inequality $\log(1+\xi) \le \xi$ for $\xi \ge 0$ to the second term will lead to a sufficient condition for $\widetilde{\boldsymbol{\lambda}}^{(\alpha,\boldsymbol{\beta},\boldsymbol{\gamma};\beta_0,\gamma_0)}$ to improve on $\widetilde{\boldsymbol{\lambda}}^{(\boldsymbol{\beta})}$ that is similar to the condition of Theorem 1 of Hamura and Kubokawa (2019) and incompatible with the condition for propriety. In contrast, applying the sharper inequality $\log(1+\xi) \le \xi - \xi^2/\{2(1+\xi)\}$ for $\xi \ge 0$ will lead to a result applicable to proper Bayes estimators. This shaper inequality is similar to the inequality of Lemma 3.1 of Dey, Ghosh and Srinivasan (1987), which is used by Ghosh and Yang (1988).

Remark 3.4. The class of proper Bayes minimax estimators will be broadened by replacing the factor $u^{\alpha-1+\beta_0}/(1+u/\gamma_0)^{\beta_0}$ in (2.2) with $u^{\beta_\cdot}\psi(u)$, where $\beta_\cdot = \sum_{i=1}^m \beta_i$ and ψ is a proper density on $(0,\infty)$. This class of priors is considered by Ghosh and Parsian (1981) for the balanced case with $\boldsymbol{\beta} = \boldsymbol{\gamma} = \boldsymbol{j}$. One choice for ψ is the exponential density $\psi(u) = e^{-u/\gamma_0}$ for $u > 0$. The details are omitted.

4 Simulation study

In this section, we investigate through simulation the numerical performance of the risk functions of the Bayes estimators given in Section 2 under the loss function $L_{\boldsymbol{c}}$ given by (1.5) with $\boldsymbol{c} = \boldsymbol{n}$ or $\boldsymbol{c} = \boldsymbol{j}$. For the case of $\boldsymbol{c} = \boldsymbol{n}$, the estimators which we compare are the following five:

ML: the ML estimator $\widehat{\boldsymbol{\lambda}}^{\mathrm{ML}} = (X_1/n_1, \ldots, X_m/n_m)$,
PB1: the proper Bayes estimator $\widehat{\boldsymbol{\lambda}}^{\mathrm{PB1}} = \widehat{\boldsymbol{\lambda}}^{(m-1,\boldsymbol{j},\boldsymbol{\nu};2,1)}$ given by (1.2) with $\beta_0 = 2$,
GB1: the generalized Bayes estimator $\widehat{\boldsymbol{\lambda}}^{\mathrm{GB1}} = \widehat{\boldsymbol{\lambda}}^{(m-1,\boldsymbol{j},\boldsymbol{\nu};0,1)}$ given by (1.2) with $\beta_0 = 0$,
PB2: the proper Bayes estimator $\widehat{\boldsymbol{\lambda}}^{\mathrm{PB2}} = \widehat{\boldsymbol{\lambda}}^{(m-1,\boldsymbol{j},\boldsymbol{j};2,\underline{n})}$ given by (2.12) with $(\alpha, \beta_0, \gamma_0) = (m-1, 2, \underline{n})$,
GB2: the generalized Bayes estimator $\widehat{\boldsymbol{\lambda}}^{\mathrm{GB2}} = \widehat{\boldsymbol{\lambda}}^{(m-1,\boldsymbol{j},\boldsymbol{j};0,\underline{n})}$ given by (2.12) with $(\alpha, \beta_0, \gamma_0) = (m-1, 0, \underline{n})$.

For the case of $\boldsymbol{c} = \boldsymbol{j}$, the estimators which we compare are the above five estimators and the following two:

PB3: the proper Bayes estimator $\widehat{\boldsymbol{\lambda}}^{\mathrm{PB3}} = \widehat{\boldsymbol{\lambda}}^{(m-1,\boldsymbol{j},\boldsymbol{\nu}\circ\boldsymbol{\nu};2,1/\overline{n})}$ given by (2.11) with $\boldsymbol{\beta} = \boldsymbol{j}$, $\boldsymbol{\gamma} = \boldsymbol{\nu} \circ \boldsymbol{\nu}$, and $(\alpha, \beta_0, \gamma_0) = (m-1, 2, 1/\overline{n})$,
GB3: the generalized Bayes estimator $\widehat{\boldsymbol{\lambda}}^{\mathrm{GB3}} = \widehat{\boldsymbol{\lambda}}^{(m-1,\boldsymbol{j},\boldsymbol{\nu}\circ\boldsymbol{\nu};0,1/\overline{n})}$ given by (2.11) with $\boldsymbol{\beta} = \boldsymbol{j}$, $\boldsymbol{\gamma} = \boldsymbol{\nu} \circ \boldsymbol{\nu}$, and $(\alpha, \beta_0, \gamma_0) = (m-1, 0, 1/\overline{n})$.

The imbalanced cases in $\boldsymbol{a} = (n_1\gamma_1, \ldots, n_m\gamma_m)$ and $\boldsymbol{C} = (c_1/(n_1{}^2\gamma_1), \ldots, c_m/(n_m{}^2\gamma_m))$ are summarized in Table 1. We consider the two estimators $\widehat{\boldsymbol{\lambda}}^{\mathrm{PB3}}$ and $\widehat{\boldsymbol{\lambda}}^{\mathrm{GB3}}$ for $\boldsymbol{c} = \boldsymbol{j}$ in order to include the case where $\boldsymbol{C} = \boldsymbol{j}$.

Table 1 *Imbalanced cases in* $\boldsymbol{a}$ *and* $\boldsymbol{C}$

$\boldsymbol{c}$	$\boldsymbol{\gamma}$	$\boldsymbol{a}$	$\boldsymbol{C}$
$\boldsymbol{n}$	$\boldsymbol{\nu}$	$\boldsymbol{j}$	$\boldsymbol{j}$
$\boldsymbol{n}$	$\boldsymbol{j}_m$	$\boldsymbol{n}$	$\boldsymbol{\nu}$
$\boldsymbol{j}_m$	$\boldsymbol{\nu}$	$\boldsymbol{j}$	$\boldsymbol{\nu}$
$\boldsymbol{j}_m$	$\boldsymbol{j}_m$	$\boldsymbol{n}$	$\boldsymbol{\nu} \circ \boldsymbol{\nu}$
$\boldsymbol{j}_m$	$\boldsymbol{\nu} \circ \boldsymbol{\nu}$	$\boldsymbol{\nu}$	$\boldsymbol{j}$

When $\boldsymbol{c} = \boldsymbol{n}$, the homogeneous proper Bayes estimator $\widehat{\boldsymbol{\lambda}}^{\mathrm{PB1}}$ is always admissible and, by part (ii) of Theorem 3.1, minimax. On the other hand, the heterogeneous proper Bayes estimator $\widehat{\boldsymbol{\lambda}}^{\mathrm{PB2}}$ is admissible, but the minimaxity is not clear, because Theorem 3.1 cannot always be applied when $\boldsymbol{n}$ is unbalanced. However, the conditions for the minimaxity of $\widehat{\boldsymbol{\lambda}}^{\mathrm{PB2}}$ given in Theorem 3.1 are somewhat restrictive especially when the sample sizes are unbalanced, and it is worth investigating the performance of $\widehat{\boldsymbol{\lambda}}^{\mathrm{PB2}}$. The generalized Bayes estimators $\widehat{\boldsymbol{\lambda}}^{\mathrm{GB1}}$, $\widehat{\boldsymbol{\lambda}}^{\mathrm{GB2}}$, and $\widehat{\boldsymbol{\lambda}}^{\mathrm{GB3}}$ are similar to the corresponding proper Bayes estimators $\widehat{\boldsymbol{\lambda}}^{\mathrm{PB1}}$, $\widehat{\boldsymbol{\lambda}}^{\mathrm{PB2}}$, and $\widehat{\boldsymbol{\lambda}}^{\mathrm{PB3}}$ but whether or not the generalized Bayes estimators are admissible is not clear.

We set $m = 30$ and $(n_i, \lambda_i) = (\underline{n}, \lambda^{(1)})$ for $i = 1, \dots, 15$ and $(n_i, \lambda_i) = (\overline{n}, \lambda^{(2)})$ for $i = 16, \dots, 30$ and we generate random numbers of $\boldsymbol{X}$ for $(\underline{n}, \overline{n}) = (1, 1), (0.5, 2), (0.1, 10)$ and $(\lambda^{(1)}, \lambda^{(2)}) = (1, 1), (3, 3), (1, 3), (3, 1)$. For each estimator $\widehat{\boldsymbol{\lambda}}$, we obtain approximated values of the risk function $E_{\boldsymbol{\lambda}}[L_{\boldsymbol{c}}(\widehat{\boldsymbol{\lambda}}, \boldsymbol{\lambda})]$ by simulation with 100,000 replications. The integrals are calculated via the Monte Carlo simulation with 100,000 replications. The percentage relative improvement in average loss (PRIAL) of an estimator $\widehat{\boldsymbol{\lambda}}$ over $\widehat{\boldsymbol{\lambda}}^{\mathrm{ML}}$ is defined by

$$\mathrm{PRIAL} = 100\{E_{\boldsymbol{\lambda}}[L_{\boldsymbol{c}}(\widehat{\boldsymbol{\lambda}}^{\mathrm{ML}}, \boldsymbol{\lambda})] - E_{\boldsymbol{\lambda}}[L_{\boldsymbol{c}}(\widehat{\boldsymbol{\lambda}}, \boldsymbol{\lambda})]\}/E_{\boldsymbol{\lambda}}[L_{\boldsymbol{c}}(\widehat{\boldsymbol{\lambda}}^{\mathrm{ML}}, \boldsymbol{\lambda})].$$

For the case of $\boldsymbol{c} = \boldsymbol{n}$, Table 2 reports values of the risks of the estimators with values of PRIAL given in parentheses. When $(\underline{n}, \overline{n}) = (1, 1)$, the risk values of $\widehat{\boldsymbol{\lambda}}^{\mathrm{PB1}}$ and $\widehat{\boldsymbol{\lambda}}^{\mathrm{PB2}}$ are the same because $\widehat{\boldsymbol{\lambda}}^{\mathrm{PB1}} = \widehat{\boldsymbol{\lambda}}^{\mathrm{PB2}}$. When $(\underline{n}, \overline{n}) = (0.5, 2)$, the risk values of $\widehat{\boldsymbol{\lambda}}^{\mathrm{PB2}}$ are smaller than those of $\widehat{\boldsymbol{\lambda}}^{\mathrm{PB1}}$ except when $(\lambda^{(1)}, \lambda^{(2)}) = (3, 1)$. When $(\underline{n}, \overline{n}) = (0.1, 10)$, all risk the values of $\widehat{\boldsymbol{\lambda}}^{\mathrm{PB2}}$ are much smaller than those of $\widehat{\boldsymbol{\lambda}}^{\mathrm{PB1}}$, and the improvement of $\widehat{\boldsymbol{\lambda}}^{\mathrm{PB2}}$ is significant. In addition, when $(\underline{n}, \overline{n}) = (0.1, 10)$, $\widehat{\boldsymbol{\lambda}}^{\mathrm{PB2}}$ has the largest values of PRIAL while $\widehat{\boldsymbol{\lambda}}^{\mathrm{PB1}}$ has the smallest values of PRIAL. These results suggest that the heterogeneous shrinkage estimators can enjoy substantial improvement over the homogeneous shrinkage estimators in the more unbalanced cases. The risk values of the proper Bayes estimators are almost the same as the corresponding risk values of their generalized Bayes counterparts.

For the case of $\boldsymbol{c} = \boldsymbol{j}$, Table 3 reports values of the risks of the estimators with values of PRIAL given in parentheses. The performance of the five estimators $\widehat{\boldsymbol{\lambda}}^{\mathrm{ML}}$, $\widehat{\boldsymbol{\lambda}}^{\mathrm{PB1}}$, $\widehat{\boldsymbol{\lambda}}^{\mathrm{GB1}}$, $\widehat{\boldsymbol{\lambda}}^{\mathrm{PB2}}$, and $\widehat{\boldsymbol{\lambda}}^{\mathrm{GB2}}$ is almost the same as in the previous case. The estimators $\widehat{\boldsymbol{\lambda}}^{\mathrm{PB3}}$ and $\widehat{\boldsymbol{\lambda}}^{\mathrm{GB3}}$, which satisfy the condition $C_1 = \cdots = C_m$, have the largest risk values for $(\underline{n}, \overline{n}) = (0.5, 2), (0.1, 10)$.

Table 2 *Risks of the estimators ML, PB1, GB1, PB2, and GB2 for* $\boldsymbol{c} = \boldsymbol{n}$. (*Values of PRIAL of PB1, GB1, PG2, and GB2 are given in parentheses*)

$(\underline{n}, \overline{n})$	$(\lambda^{(1)}, \lambda^{(2)})$	ML	PB1	GB1	PB2	GB2
(1, 1)	(1, 1)	30.01	15.35 (48.83)	15.37 (48.77)	15.35 (48.83)	15.37 (48.77)
	(3, 3)	29.99	22.83 (23.89)	22.82 (23.91)	22.83 (23.89)	22.82 (23.92)
	(1, 3)	30.00	20.38 (32.09)	20.38 (32.09)	20.38 (32.08)	20.38 (32.08)
	(3, 1)	30.03	20.37 (32.16)	20.37 (32.15)	20.37 (32.16)	20.37 (32.15)
(0.5, 2)	(1, 1)	30.00	17.03 (43.21)	17.05 (43.17)	15.34 (48.88)	15.35 (48.83)
	(3, 3)	30.00	23.99 (20.03)	23.98 (20.06)	22.16 (26.14)	22.16 (26.13)
	(1, 3)	29.98	23.24 (22.47)	23.24 (22.49)	19.16 (36.09)	19.21 (35.93)
	(3, 1)	30.00	19.47 (35.10)	19.47 (35.09)	20.53 (31.57)	20.50 (31.66)
(0.1, 10)	(1, 1)	30.11	25.37 (15.73)	25.36 (15.75)	15.18 (49.56)	15.19 (49.56)
	(3, 3)	30.00	28.25 (5.85)	28.24 (5.87)	18.05 (39.85)	18.05 (39.84)
	(1, 3)	29.96	28.22 (5.80)	28.21 (5.83)	16.23 (45.82)	16.25 (45.77)
	(3, 1)	29.99	25.37 (15.42)	25.36 (15.45)	17.54 (41.53)	17.53 (41.54)

Table 3 *Risks of the estimators ML, PB*1, *GB*1, *PB*2, *GB*2, *PB*3, *and GB*3 *for* $\mathbf{c} = \boldsymbol{j}$. (*Values of PRIAL of PB*1, *GB*1, *PG*2, *GB*2, *PB*3, *and GB*3 *are given in parentheses*)

$(\underline{n}, \overline{n})$	$(\lambda^{(1)}, \lambda^{(2)})$	ML	PB1	GB1	PB2	GB2	PB3	GB3
(1, 1)	(1, 1)	30.01	15.37 (48.79)	15.39 (48.73)	15.37 (48.79)	15.38 (48.74)	15.37 (48.79)	15.38 (48.74)
	(3, 3)	30.03	22.86 (23.89)	22.85 (23.91)	22.86 (23.88)	22.85 (23.90)	22.86 (23.88)	22.85 (23.90)
	(1, 3)	30.04	20.38 (32.13)	20.38 (32.13)	20.38 (32.14)	20.38 (32.14)	20.38 (32.14)	20.38 (32.14)
	(3, 1)	30.04	20.38 (32.16)	20.38 (32.16)	20.38 (32.17)	20.38 (32.16)	20.38 (32.17)	20.38 (32.16)
(0.5, 2)	(1, 1)	37.52	17.76 (52.66)	18.01 (51.98)	15.30 (59.21)	15.32 (59.16)	24.55 (34.57)	24.82 (33.84)
	(3, 3)	37.57	27.54 (26.70)	27.77 (26.08)	24.79 (34.00)	24.81 (33.96)	32.99 (12.18)	33.07 (11.98)
	(1, 3)	37.61	25.35 (32.60)	25.69 (31.68)	18.44 (50.98)	18.63 (50.46)	32.73 (12.98)	32.82 (12.73)
	(3, 1)	37.54	23.55 (37.25)	23.62 (37.07)	25.10 (33.13)	24.91 (33.65)	26.86 (28.44)	27.03 (27.98)
(0.1, 10)	(1, 1)	151.49	105.39 (30.43)	107.62 (28.96)	15.06 (90.06)	15.07 (90.05)	150.85 (0.43)	150.86 (0.42)
	(3, 3)	151.62	133.27 (12.10)	134.32 (11.41)	36.35 (76.02)	36.40 (75.99)	151.42 (0.13)	151.42 (0.13)
	(1, 3)	152.12	133.46 (12.27)	134.54 (11.56)	17.90 (88.23)	18.09 (88.11)	151.93 (0.13)	151.93 (0.13)
	(3, 1)	151.37	106.84 (29.42)	108.95 (28.02)	38.55 (74.53)	38.47 (74.59)	150.72 (0.43)	150.74 (0.42)

In particular, when $(\underline{n}, \overline{n}) = (0.1, 10)$, these estimators have the values of PRIAL almost equal to zero.

5 Application

In this section, several estimation methods considered in the previous sections are applied to data relating to the standardized mortality ratio (SMR). (For the SMR, see, for example, Clayton and Kaldor (1987).) More specifically, the data consist of actual and expected numbers of deaths of females from a specific cause in $m = 72$ districts in a prefecture in Japan during the 5 years from 2008 to 2012. For $i = 1, \dots, m$, the actual and expected numbers of deaths in the ith district are denoted by x_i and n_i, respectively. Each component of an estimator $\boldsymbol{\lambda} = (\hat{\lambda}_1, \dots, \hat{\lambda}_m)$ is a measure of relative risk in a district calculated from the data.

We here consider only the three estimators $\widehat{\boldsymbol{\lambda}}^{\mathrm{ML}} = (\hat{\lambda}_1^{\mathrm{ML}}, \dots, \hat{\lambda}_m^{\mathrm{ML}})$, $\widehat{\boldsymbol{\lambda}}^{\mathrm{PB1}} = (\hat{\lambda}_1^{\mathrm{PB1}}, \dots, \hat{\lambda}_m^{\mathrm{PB1}})$, and $\widehat{\boldsymbol{\lambda}}^{\mathrm{PB2}} = (\hat{\lambda}_1^{\mathrm{PB2}}, \dots, \hat{\lambda}_m^{\mathrm{PB2}})$ given in Section 4. Integrals are calculated via the Monte Carlo simulation with 100,000 replications. The data and the estimates for all the $m = 72$ districts are given in Tables 4 and 5.

The values of the ratio $\hat{\lambda}_i^{\mathrm{PB2}} / \hat{\lambda}_i^{\mathrm{PB1}}$ for all $i = 1, \dots, m$ are plotted in Figure 1. For $i = 1, \dots, m$, the heterogeneous estimator $\hat{\lambda}_i^{\mathrm{PB2}}$ shrinks the ML estimator $\hat{\lambda}_i^{\mathrm{ML}}$ toward the origin more than the homogeneous estimator $\hat{\lambda}_i^{\mathrm{PB1}}$ if $n_i \lessapprox 50$ and less than $\hat{\lambda}_i^{\mathrm{PB1}}$ if $n_i \gtrapprox 50$.

Appendix

All the proofs of the lemmas in Sections 2 and 3 are given here. For $\boldsymbol{v} = (v_1, \dots, v_m) \in \mathbb{R}^m$ and $\tilde{\boldsymbol{v}} = (\tilde{v}_1, \dots, \tilde{v}_m) \in \mathbb{R}^m$, we write the inner product $v_1 \tilde{v}_1 + \dots + v_m \tilde{v}_m$ as $\boldsymbol{v} \cdot \tilde{\boldsymbol{v}}$.

Proof of Lemma 2.1. Let $J = \int \dots \int_{(0,\infty)^m} \pi_{\alpha,\boldsymbol{\beta},\boldsymbol{\gamma};\beta_0,\gamma_0}(\boldsymbol{\lambda})\, d\boldsymbol{\lambda}$. From (2.2), it follows that

$$J = \int_0^\infty \left\{ \frac{u^{\alpha-1+\beta_0}}{(1+u/\gamma_0)^{\beta_0}} \int \dots \int_{(0,\infty)^m} \left(\prod_{i=1}^m {\lambda_i}^{\beta_i - 1} \right) e^{-u \sum_{i=1}^m \lambda_i/\gamma_i}\, d\boldsymbol{\lambda} \right\} du.$$

By making the change of variables

$$(\theta_1, \dots, \theta_{m-1}, \Lambda) = \left(\lambda_1 \left(\sum_{i=1}^m \lambda_i \right)^{-1}, \dots, \lambda_{m-1} \left(\sum_{j=1}^m \lambda_j \right)^{-1}, \sum_{i=1}^m \lambda_i \right),$$

we obtain

$$\begin{aligned} J &= \int_0^\infty \left\{ \frac{u^{\alpha-1+\beta_0}}{(1+u/\gamma_0)^{\beta_0}} \right. \\ &\quad \left. \times \int \dots \int_{D \times (0,\infty)} \left(\Lambda^{\beta_\cdot - m} \prod_{i=1}^m {\theta_i}^{\beta_i - 1} \right) e^{-\Lambda u \sum_{i=1}^m \theta_i/\gamma_i} \Lambda^{m-1}\, d\theta_1 \cdots d\theta_{m-1}\, d\Lambda \right\} du \\ &= \int_0^\infty \left\{ \frac{u^{\alpha-1+\beta_0}}{(1+u/\gamma_0)^{\beta_0}} \right. \\ &\quad \left. \times \int \dots \int_D \left(\prod_{i=1}^m {\theta_i}^{\beta_i - 1} \right) \Gamma(\beta_\cdot) \left(u \sum_{i=1}^m \theta_i/\gamma_i \right)^{-\beta_\cdot} d\theta_1 \cdots d\theta_{m-1} \right\} du, \end{aligned} \tag{A.1}$$

Table 4 *The data and the estimates of relative risk for* $i = 1, \dots, 36$

i	x_i	n_i	$\hat{\lambda}_i^{\mathrm{ML}}$	$\hat{\lambda}_i^{\mathrm{PB1}}$	$\hat{\lambda}_i^{\mathrm{PB2}}$
1	49	49.65	0.99	0.97	0.97
2	64	66.60	0.96	0.94	0.95
3	69	63.83	1.08	1.06	1.07
4	79	87.49	0.90	0.89	0.89
5	47	48.60	0.97	0.95	0.95
6	35	42.58	0.82	0.81	0.80
7	66	76.12	0.87	0.85	0.86
8	75	72.67	1.03	1.01	1.02
9	49	55.24	0.89	0.87	0.87
10	54	64.75	0.83	0.82	0.82
11	192	182.16	1.05	1.04	1.05
12	349	269.71	1.29	1.27	1.29
13	48	40.54	1.18	1.16	1.16
14	47	45.94	1.02	1.00	1.00
15	62	54.53	1.14	1.12	1.12
16	38	32.79	1.16	1.14	1.13
17	31	31.41	0.99	0.97	0.96
18	81	78.79	1.03	1.01	1.02
19	57	52.49	1.09	1.07	1.07
20	62	57.09	1.09	1.07	1.07
21	21	23.03	0.91	0.90	0.88
22	83	67.53	1.23	1.21	1.21
23	116	111.32	1.04	1.02	1.03
24	51	41.87	1.22	1.20	1.19
25	41	36.28	1.13	1.11	1.10
26	21	17.72	1.19	1.16	1.13
27	59	47.77	1.24	1.21	1.21
28	13	9.42	1.38	1.36	1.26
29	20	11.98	1.67	1.64	1.55
30	22	23.76	0.93	0.91	0.89
31	14	15.09	0.93	0.91	0.87
32	23	13.38	1.72	1.69	1.61
33	14	9.72	1.44	1.42	1.31
34	5	3.28	1.53	1.50	1.19
35	52	52.79	0.99	0.97	0.97
36	6	7.03	0.85	0.84	0.75

where θ_m denotes $1 - (\theta_1 + \cdots + \theta_{m-1})$ and $D = \{(\zeta_1, \dots, \zeta_{m-1}) \in (0,1)^{m-1} : \zeta_1 + \cdots + \zeta_{m-1} < 1\}$. Let $\overline{\gamma} = \max_{1 \le i \le m} \gamma_i$ and $\underline{\gamma} = \min_{1 \le i \le m} \gamma_i$. Then, from (A.1),

$$\begin{aligned} J \le \int_0^\infty \Bigg\{ & \frac{u^{\alpha-1+\beta_0}}{(1+u/\gamma_0)^{\beta_0}} \\ & \times \int \cdots \int_D \left(\prod_{i=1}^m \theta_i{}^{\beta_i - 1} \right) \Gamma(\beta.) u^{-\beta.} \left(\overline{\gamma} \Big/ \sum_{i=1}^m \theta_i \right)^{\beta.} d\theta_1 \cdots d\theta_{m-1} \Bigg\} du \\ = \Gamma(\beta.) \overline{\gamma}^{\beta.} & \left\{ \int \cdots \int_D \left(\prod_{i=1}^m \theta_i{}^{\beta_i - 1} \right) d\theta_1 \cdots d\theta_{m-1} \right\} \int_0^\infty \frac{u^{\alpha-1+\beta_0-\beta.}}{(1+u/\gamma_0)^{\beta_0}} du, \end{aligned}$$

Table 5 *The data and the estimates of relative risk for $i = 37, \ldots, 72$*

i	x_i	n_i	$\hat{\lambda}_i^{\mathrm{ML}}$	$\hat{\lambda}_i^{\mathrm{PB1}}$	$\hat{\lambda}_i^{\mathrm{PB2}}$
37	7	9.78	0.72	0.70	0.65
38	7	7.05	0.99	0.98	0.88
39	10	12.32	0.81	0.80	0.75
40	41	54.59	0.75	0.74	0.74
41	15	9.24	1.62	1.59	1.47
42	11	9.67	1.14	1.12	1.04
43	15	17.81	0.84	0.83	0.80
44	119	124.74	0.95	0.94	0.95
45	103	89.18	1.16	1.13	1.14
46	25	24.95	1.00	0.98	0.97
47	61	55.91	1.09	1.07	1.07
48	83	70.76	1.17	1.15	1.16
49	45	36.92	1.22	1.20	1.19
50	141	127.72	1.10	1.08	1.10
51	151	156.31	0.97	0.95	0.96
52	22	16.53	1.33	1.31	1.26
53	98	83.55	1.17	1.15	1.16
54	37	38.18	0.97	0.95	0.95
55	39	32.97	1.18	1.16	1.15
56	29	28.27	1.03	1.01	0.99
57	20	19.84	1.01	0.99	0.96
58	21	25.64	0.82	0.80	0.79
59	72	52.02	1.38	1.36	1.36
60	19	31.88	0.60	0.59	0.58
61	29	22.59	1.28	1.26	1.23
62	15	8.82	1.70	1.67	1.54
63	9	12.31	0.73	0.72	0.68
64	118	111.11	1.06	1.04	1.05
65	52	37.12	1.40	1.38	1.37
66	59	60.27	0.98	0.96	0.96
67	30	29.67	1.01	0.99	0.98
68	155	181.29	0.86	0.84	0.85
69	51	56.42	0.90	0.89	0.89
70	75	89.61	0.84	0.82	0.83
71	75	78.53	0.96	0.94	0.94
72	43	34.05	1.26	1.24	1.23

the right-hand side of which is finite if $\alpha < \beta_{\cdot} < \alpha + \beta_0$. Similarly,

$$J \geq \Gamma(\beta_{\cdot})\underline{\gamma}^{\beta_{\cdot}} \left\{ \int \cdots \int_D \left(\prod_{i=1}^{m} \theta_i{}^{\beta_i - 1} \right) d\theta_1 \cdots d\theta_{m-1} \right\} \int_0^\infty \frac{u^{\alpha - 1 + \beta_0 - \beta_{\cdot}}}{(1 + u/\gamma_0)^{\beta_0}} du = \infty$$

if the condition $\alpha < \beta_{\cdot} < \alpha + \beta_0$ does not hold, and the proof is complete. □

Proof of Lemma 2.2. For part (i), we have by integration by parts that

$$K(\boldsymbol{\gamma}, \boldsymbol{\xi}, \alpha; \gamma_0, \beta_0)$$
$$= \left[\frac{u^\alpha}{\alpha} \frac{1}{(1 + u/\gamma_0)^{\beta_0}} \prod_{i=1}^{m} \frac{1}{(1 + u/\gamma_i)^{\xi_i}} \right]_0^\infty$$

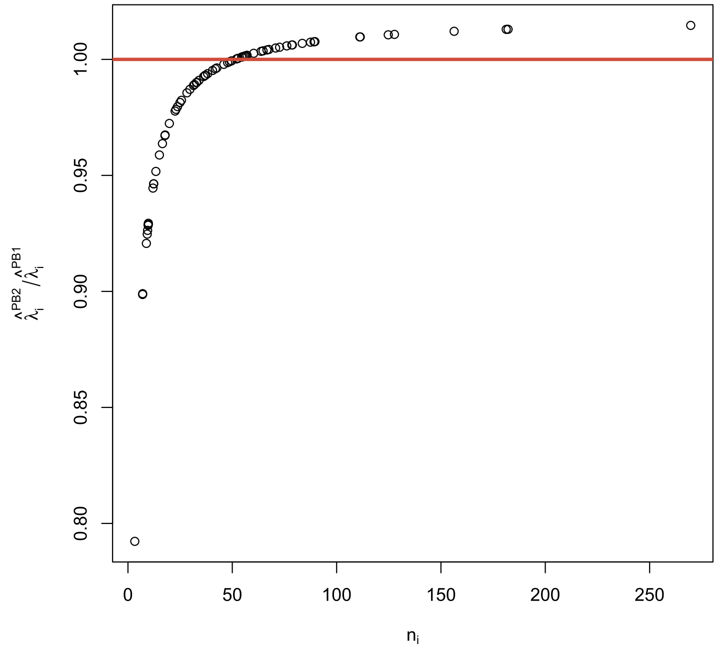

Figure 1 *The ratio* $\hat{\lambda}_i^{\mathrm{PB2}}/\hat{\lambda}_i^{\mathrm{PB1}}$.

$$
-\int_0^\infty \frac{u^\alpha}{\alpha}\left(\frac{-\beta_0/\gamma_0}{1+u/\gamma_0}+\sum_{i=1}^m \frac{-\xi_i/\gamma_i}{1+u/\gamma_i}\right)\frac{1}{(1+u/\gamma_0)^{\beta_0}}\prod_{i=1}^m \frac{1}{(1+u/\gamma_i)^{\xi_i}}\,du
$$

$$
=0+\frac{1}{\alpha}\left\{\frac{\beta_0}{\gamma_0}K(\boldsymbol{\gamma},\boldsymbol{\xi},\alpha+1;\gamma_0,\beta_0+1)+\sum_{i=1}^m \frac{\xi_i}{\gamma_i}K(\boldsymbol{\gamma},\boldsymbol{\xi}+\boldsymbol{e}_i,\alpha+1;\gamma_0,\beta_0)\right\}.
$$

Part (ii) follows since

$$
\begin{aligned}
&K(\boldsymbol{\gamma},\boldsymbol{\xi}+\boldsymbol{e}_i,\alpha+1;\gamma_0,\beta_0)\\
&\quad=\int_0^\infty \frac{u^{\alpha-1}}{(1+u/\gamma_0)^{\beta_0}}\frac{u}{1+u/\gamma_i}\prod_{j=1}^m \frac{1}{(1+u/\gamma_j)^{\xi_j}}\,du\\
&\quad=\int_0^\infty \frac{u^{\alpha-1}}{(1+u/\gamma_0)^{\beta_0}}\gamma_i\left(1-\frac{1}{1+u/\gamma_i}\right)\prod_{j=1}^m \frac{1}{(1+u/\gamma_j)^{\xi_j}}\,du\\
&\quad=\gamma_i\{K(\boldsymbol{\gamma},\boldsymbol{\xi},\alpha;\gamma_0,\beta_0)-K(\boldsymbol{\gamma},\boldsymbol{\xi}+\boldsymbol{e}_i,\alpha;\gamma_0,\beta_0)\}
\end{aligned}
$$

for $i=1,\dots,m$. This completes the proof. □

Proof of Lemma 2.3. For part (i), let $f(u)=u^{\alpha-1}(1+u/\gamma_0)^{-\beta_0}\prod_{i=1}^m(1+u/\gamma_i)^{-\xi_i}$ for $u>0$ and let $\Delta_K=K(\boldsymbol{\gamma},\boldsymbol{\xi},\alpha+1;\gamma_0,\beta_0)K(\boldsymbol{\gamma},\boldsymbol{\xi},\alpha;\gamma_0,\beta_0)-K(\boldsymbol{\gamma},\boldsymbol{\xi},\alpha+1;\gamma_0,\beta_0+1)K(\boldsymbol{\gamma},\boldsymbol{\xi},\alpha;\gamma_0,\beta_0-1)$. Note that

$$
\begin{aligned}
\Delta_K&=\int_0^\infty uf(u)\,du\int_0^\infty f(u)\,du-\int_0^\infty \frac{u}{1+u/\gamma_0}f(u)\,du\int_0^\infty\left(1+\frac{u}{\gamma_0}\right)f(u)\,du\\
&=\int_0^\infty uf(u)\,du\int_0^\infty f(u)\,du
\end{aligned}
$$

$$
\begin{aligned}
&-\gamma_0\int_0^\infty\Big(1-\frac{1}{1+u/\gamma_0}\Big)f(u)\,du\int_0^\infty\Big(1+\frac{u}{\gamma_0}\Big)f(u)\,du\\
&=-\gamma_0\Big\{\int_0^\infty f(u)\,du\Big\}^2+\gamma_0\int_0^\infty\frac{1}{1+u/\gamma_0}f(u)\,du\int_0^\infty\Big(1+\frac{u}{\gamma_0}\Big)f(u)\,du.
\end{aligned}
$$

Then it follows from the Cauchy–Schwarz inequality that $\Delta_K\geq 0$, which can be rewritten as (2.8). The inequality (2.7) can be similarly shown. Next, we prove part (ii). From (2.8), we have

$$
0\leq\frac{K(\boldsymbol{\gamma},\boldsymbol{\xi},\alpha+1;\gamma_0,\beta_0)}{K(\boldsymbol{\gamma},\boldsymbol{\xi},\alpha;\gamma_0,\beta_0-1)}-\frac{K(\boldsymbol{\gamma},\boldsymbol{\xi},\alpha+1;\gamma_0,\beta_0+1)}{K(\boldsymbol{\gamma},\boldsymbol{\xi},\alpha;\gamma_0,\beta_0)}.
$$

By adding and subtracting $K(\boldsymbol{\gamma},\boldsymbol{\xi},\alpha+1;\gamma_0,\beta_0)/K(\boldsymbol{\gamma},\boldsymbol{\xi},\alpha;\gamma_0,\beta_0)$, we obtain

$$
\begin{aligned}
0&\leq-\frac{K(\boldsymbol{\gamma},\boldsymbol{\xi},\alpha+1;\gamma_0,\beta_0)}{K(\boldsymbol{\gamma},\boldsymbol{\xi},\alpha;\gamma_0,\beta_0)}\Big(1-\frac{K(\boldsymbol{\gamma},\boldsymbol{\xi},\alpha;\gamma_0,\beta_0)}{K(\boldsymbol{\gamma},\boldsymbol{\xi},\alpha;\gamma_0,\beta_0-1)}\Big)\\
&\quad+\frac{1}{\gamma_0}\frac{K(\boldsymbol{\gamma},\boldsymbol{\xi},\alpha+2;\gamma_0,\beta_0+1)}{K(\boldsymbol{\gamma},\boldsymbol{\xi},\alpha;\gamma_0,\beta_0)}\\
&=-\frac{K(\boldsymbol{\gamma},\boldsymbol{\xi},\alpha+1;\gamma_0,\beta_0)}{K(\boldsymbol{\gamma},\boldsymbol{\xi},\alpha;\gamma_0,\beta_0)}\frac{1}{\gamma_0}\frac{K(\boldsymbol{\gamma},\boldsymbol{\xi},\alpha+1;\gamma_0,\beta_0)}{K(\boldsymbol{\gamma},\boldsymbol{\xi},\alpha;\gamma_0,\beta_0-1)}\\
&\quad+\frac{1}{\gamma_0}\frac{K(\boldsymbol{\gamma},\boldsymbol{\xi},\alpha+2;\gamma_0,\beta_0+1)}{K(\boldsymbol{\gamma},\boldsymbol{\xi},\alpha;\gamma_0,\beta_0)}\\
&\leq-\frac{K(\boldsymbol{\gamma},\boldsymbol{\xi},\alpha+1;\gamma_0,\beta_0)}{K(\boldsymbol{\gamma},\boldsymbol{\xi},\alpha;\gamma_0,\beta_0)}\frac{1}{\gamma_0}\frac{K(\boldsymbol{\gamma},\boldsymbol{\xi},\alpha+1;\gamma_0,\beta_0+1)}{K(\boldsymbol{\gamma},\boldsymbol{\xi},\alpha;\gamma_0,\beta_0)}\\
&\quad+\frac{1}{\gamma_0}\frac{K(\boldsymbol{\gamma},\boldsymbol{\xi},\alpha+2;\gamma_0,\beta_0+1)}{K(\boldsymbol{\gamma},\boldsymbol{\xi},\alpha;\gamma_0,\beta_0)},
\end{aligned}
$$

where the second inequality follows from (2.8), and thus (2.10) follows. The inequality (2.9) can be similarly shown. The proof of Lemma 2.3 is complete. □

Proof of Lemma 2.4. Let $\boldsymbol{x}=(x_1,\dots,x_m)\in\mathbb{N}_0{}^m$. For $i=1,\dots,m$, the posterior mean of $1/\lambda_i$ with respect to the observation $\boldsymbol{X}=\boldsymbol{x}$ and the prior $\pi_{\alpha,\boldsymbol{\beta},\boldsymbol{\gamma};\beta_0,\gamma_0}$, denoted $E^{\boldsymbol{\lambda}|\boldsymbol{X}}[1/\lambda_i|\boldsymbol{X}=\boldsymbol{x}]$, is given by

$$
\frac{\int_0^\infty[\frac{u^{\alpha-1+\beta_0}}{(1+u/\gamma_0)^{\beta_0}}\int\cdots\int_{(0,\infty)^m}\{\frac{1}{\lambda_i}(\prod_{j=1}^m{\lambda_j}^{x_j+\beta_j-1}e^{-n_j\lambda_j})e^{-u\sum_{j=1}^m\lambda_j/\gamma_j}\}\,d\boldsymbol{\lambda}]\,du}{\int_0^\infty[\frac{u^{\alpha-1+\beta_0}}{(1+u/\gamma_0)^{\beta_0}}\int\cdots\int_{(0,\infty)^m}\{(\prod_{j=1}^m{\lambda_j}^{x_j+\beta_j-1}e^{-n_j\lambda_j})e^{-u\sum_{j=1}^m\lambda_j/\gamma_j}\}\,d\boldsymbol{\lambda}]\,du},
$$

which can be rewritten as

$$
\begin{aligned}
&\frac{\int_0^\infty\frac{u^{\alpha-1+\beta_0}}{(1+u/\gamma_0)^{\beta_0}}\prod_{j=1}^m\int_0^\infty{\lambda_j}^{x_j+\beta_j-1-\delta_{ij}}e^{-\lambda_j(n_j+u/\gamma_j)}\,d\lambda_j\,du}{\int_0^\infty\frac{u^{\alpha-1+\beta_0}}{(1+u/\gamma_0)^{\beta_0}}\prod_{j=1}^m\int_0^\infty{\lambda_j}^{x_j+\beta_j-1}e^{-\lambda_j(n_j+u/\gamma_j)}\,d\lambda_j\,du}\\
&=\frac{\int_0^\infty\frac{u^{\alpha-1+\beta_0}}{(1+u/\gamma_0)^{\beta_0}}\prod_{j=1}^m\frac{\Gamma(x_j+\beta_j-\delta_{ij})}{(n_j+u/\gamma_j)^{x_j+\beta_j-\delta_{ij}}}\,du}{\int_0^\infty\frac{u^{\alpha-1+\beta_0}}{(1+u/\gamma_0)^{\beta_0}}\prod_{j=1}^m\frac{\Gamma(x_j+\beta_j)}{(n_j+u/\gamma_j)^{x_j+\beta_j}}\,du}\\
&=n_i\frac{\Gamma(x_i+\beta_i-1)}{\Gamma(x_i+\beta_i)}\frac{K(\boldsymbol{n}\circ\boldsymbol{\gamma},\boldsymbol{x}+\boldsymbol{\beta}-\boldsymbol{e}_i,\alpha+\beta_0;\gamma_0,\beta_0)}{K(\boldsymbol{n}\circ\boldsymbol{\gamma},\boldsymbol{x}+\boldsymbol{\beta},\alpha+\beta_0;\gamma_0,\beta_0)},
\end{aligned}
$$

where $\delta_{ij} = \boldsymbol{e}_i \cdot \boldsymbol{e}_j$ for $j = 1, \dots, m$, $\Gamma(t) = \infty$ for $t \le 0$, and $K(\boldsymbol{n} \circ \boldsymbol{\gamma}, \boldsymbol{0} + \boldsymbol{\beta} - \boldsymbol{e}_i, \alpha + \beta_0; \gamma_0, \beta_0) = \infty$ for $\alpha \ge \sum_{j=1}^m \beta_j - 1$. Similarly, we have

$$E^{\boldsymbol{\lambda}|\boldsymbol{X}}[\lambda_i | \boldsymbol{X} = \boldsymbol{x}] = \frac{1}{n_i} \frac{\Gamma(x_i + \beta_i + 1)}{\Gamma(x_i + \beta_i)} \frac{K(\boldsymbol{n} \circ \boldsymbol{\gamma}, \boldsymbol{x} + \boldsymbol{\beta} + \boldsymbol{e}_i, \alpha + \beta_0; \gamma_0, \beta_0)}{K(\boldsymbol{n} \circ \boldsymbol{\gamma}, \boldsymbol{x} + \boldsymbol{\beta}, \alpha + \beta_0; \gamma_0, \beta_0)},$$

which is finite. Hence, for all $\boldsymbol{d} = (d_1, \dots, d_m) \in \mathbb{R}^m$, we have

$$\begin{aligned}
&E^{\boldsymbol{\lambda}|\boldsymbol{X}}[L_{\boldsymbol{c}}(\boldsymbol{d}, \boldsymbol{\lambda}) | \boldsymbol{X} = \boldsymbol{x}] \\
&= \sum_{i=1}^m c_i E^{\boldsymbol{\lambda}|\boldsymbol{X}} \left[\frac{{d_i}^2}{\lambda_i} - 2d_i + \lambda_i \middle| \boldsymbol{X} = \boldsymbol{x} \right] \\
&= \sum_{i \in S} c_i \left\{ E^{\boldsymbol{\lambda}|\boldsymbol{X}} \left[\frac{1}{\lambda_i} \middle| \boldsymbol{X} = \boldsymbol{x} \right] \left(d_i - \frac{1}{E^{\boldsymbol{\lambda}|\boldsymbol{X}}[\frac{1}{\lambda_i} | \boldsymbol{X} = \boldsymbol{x}]} \right)^2 + A_i \right\} \\
&\quad + \sum_{i \in S^c} c_i \left({d_i}^2 \cdot \infty - 2d_i + E^{\boldsymbol{\lambda}|\boldsymbol{X}}[\lambda_i | \boldsymbol{X} = \boldsymbol{x}] \right),
\end{aligned}$$

where $S = \{i \in \{1, \dots, m\} : E^{\boldsymbol{\lambda}|\boldsymbol{X}}[1/\lambda_i | \boldsymbol{X} = \boldsymbol{x}] < \infty\}$ and $A_i = -(E^{\boldsymbol{\lambda}|\boldsymbol{X}}[1/\lambda_i | \boldsymbol{X} = \boldsymbol{x}])^{-1} + E^{\boldsymbol{\lambda}|\boldsymbol{X}}[\lambda_i | \boldsymbol{X} = \boldsymbol{x}]$ for $i \in S$. Therefore, $E^{\boldsymbol{\lambda}|\boldsymbol{X}}[L_{\boldsymbol{c}}(\boldsymbol{d}, \boldsymbol{\lambda}) | \boldsymbol{X} = \boldsymbol{x}]$ is finite if and only if $d_i = 0$ for all $i \in S^c$. Furthermore, in this case, it is minimized if and only if $d_i = (E^{\boldsymbol{\lambda}|\boldsymbol{X}}[1/\lambda_i | \boldsymbol{X} = \boldsymbol{x}])^{-1}$ for all $i \in S$. Thus, $E^{\boldsymbol{\lambda}|\boldsymbol{X}}[L_{\boldsymbol{c}}(\boldsymbol{d}, \boldsymbol{\lambda}) | \boldsymbol{X} = \boldsymbol{x}]$ is uniquely minimized at

$$\boldsymbol{d} = \sum_{i \in S} \frac{x_i + \beta_i - 1}{n_i} \frac{K(\boldsymbol{n} \circ \boldsymbol{\gamma}, \boldsymbol{x} + \boldsymbol{\beta}, \alpha + \beta_0; \gamma_0, \beta_0)}{K(\boldsymbol{n} \circ \boldsymbol{\gamma}, \boldsymbol{x} + \boldsymbol{\beta} - \boldsymbol{e}_i, \alpha + \beta_0; \gamma_0, \beta_0)} \boldsymbol{e}_i,$$

which can be expressed as

$$\left(\{1 - \phi_1^{(\alpha, \boldsymbol{\beta}, \boldsymbol{\gamma}; \beta_0, \gamma_0)}(\boldsymbol{x})\} \frac{x_1 + \beta_1 - 1}{n_1}, \dots, \{1 - \phi_m^{(\alpha, \boldsymbol{\beta}, \boldsymbol{\gamma}; \beta_0, \gamma_0)}(\boldsymbol{x})\} \frac{x_m + \beta_m - 1}{n_m} \right)$$

by part (ii) of Lemma 2.2. Thus, the desired result is obtained. □

Proof of Lemma 2.5. For part (i), suppose $x_i \ge 1$. Then

$$\begin{aligned}
&n_i K(\boldsymbol{n}, \boldsymbol{x} + \boldsymbol{j} - \boldsymbol{e}_i, \alpha + \beta_0; \gamma_0, \beta_0) \\
&= \int_0^\infty \frac{u^{\alpha + \beta_0 - 1}}{(1 + u/\gamma_0)^{\beta_0}} (n_i + u) \prod_{j=1}^m \frac{1}{(1 + u/n_j)^{x_j + 1}} \, du \\
&> \int_0^\infty \frac{u^{\alpha + \beta_0}}{(1 + u/\gamma_0)^{\beta_0}} \prod_{j=1}^m \frac{1}{(1 + u/n_j)^{x_j + 1}} \, du \\
&= K(\boldsymbol{n}, \boldsymbol{x} + \boldsymbol{j}, \alpha + \beta_0 + 1; \gamma_0, \beta_0).
\end{aligned} \tag{A.2}$$

This shows the desired result. Part (ii) follows immediately from (A.2). For part (iii), let $f_i(u) = (1 + u/\gamma_0)^{-\beta_0} \prod_{j \ne i} (1 + u/n_j)^{-(x_j + 1)}$ for $u > 0$ and let $k = 0, 1$. Then we have that

$$0 \le u^{\alpha + \beta_0 - k} \frac{1}{(1 + u/n_i)^{x_i + 1 - k}} f_i(u) \uparrow u^{\alpha + \beta_0 - k} f_i(u)$$

as $n_i \to \infty$ for every $u > 0$. Since $\alpha < m-2$, it follows from the dominated convergence theorem that

$$\begin{aligned}K(\boldsymbol{n}, \boldsymbol{x}+\boldsymbol{j}-k\boldsymbol{e}_i, \alpha+\beta_0+1-k; \gamma_0, \beta_0) &= \int_0^\infty u^{\alpha+\beta_0-k} \frac{1}{(1+u/n_i)^{x_i+1-k}} f_i(u)\,du \\ &\to \int_0^\infty u^{\alpha+\beta_0-k} f_i(u)\,du \in (0,\infty)\end{aligned}$$

as $n_i \to \infty$, and this completes the proof. □

Proof of Lemma 3.1. It can be seen that

$$\begin{aligned}&E_{\boldsymbol{\lambda}}\Big[\frac{h(\boldsymbol{X}+\boldsymbol{e}_i)}{X_i+1}\Big] \\ &= \sum_{\boldsymbol{x}\in\mathbb{N}_0{}^m} \frac{h(\boldsymbol{x}+\boldsymbol{e}_i)}{x_i+1} \prod_{j=1}^m \frac{(n_j\lambda_j)^{x_j}}{x_j!} e^{-n_j\lambda_j} \\ &= \sum_{\boldsymbol{x}\in\mathbb{N}_0{}^m} \frac{h(\boldsymbol{x})}{n_i\lambda_i} \prod_{j=1}^m \frac{(n_j\lambda_j)^{x_j}}{x_j!} e^{-n_j\lambda_j} - \sum_{\boldsymbol{x}\in\mathbb{N}_0{}^m, \boldsymbol{x}\cdot\boldsymbol{e}_i=0} \frac{h(\boldsymbol{x})}{n_i\lambda_i} \prod_{j=1}^m \frac{(n_j\lambda_j)^{x_j}}{x_j!} e^{-n_j\lambda_j} \\ &= E_{\boldsymbol{\lambda}}\Big[\frac{h(\boldsymbol{X})}{n_i\lambda_i}\Big],\end{aligned}$$

which proves Lemma 3.1. □

Acknowledgments

We would like to thank the Editor, the Associate Editor, and the reviewer for their helpful comments which improved on the presentation of the paper. Research of the second author was supported in part by Grant-in-Aid for Scientific Research (18K11188, 15H01943 and 26330036) from Japan Society for the Promotion of Science.

References

Chang, Y.-T. and Shinozaki, N. (2018). New types of shrinkage estimators of Poisson means under the normalized squared error loss. *Communications in Statistics Theory and Methods* **0**, 1–15. MR3957584 https://doi.org/10.1080/03610926.2018.1423699

Clayton, D. and Kaldor, J. (1987). Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics* **43**, 671–681.

Clevenson, M. L. and Zidek, J. V. (1975). Simultaneous estimation of the means of independent Poisson laws. *Journal of the American Statistical Association* **70**, 698–705. MR0394962

Dey, D. K., Ghosh, M. and Srinivasan, C. (1987). Simultaneous estimation of parameters under entropy loss. *Journal of Statistical Planning and Inference* **15**, 347–363. MR0879215 https://doi.org/10.1016/0378-3758(86)90108-4

Ghosh, M. and Parsian, A. (1981). Bayes minimax estimation of multiple Poisson parameters. *Journal of Multivariate Analysis* **11**, 280–288. MR0618791 https://doi.org/10.1016/0047-259X(81)90115-9

Ghosh, M. and Yang, M.-C. (1988). Simultaneous estimation of Poisson means under entropy loss. *The Annals of Statistics* **16**, 278–291. MR0924871 https://doi.org/10.1214/aos/1176350705

Hamura, Y. and Kubokawa, T. (2019). Simultaneous estimation of parameters of Poisson distributions with unbalanced sample sizes. *Japanese Journal of Statistics and Data Science*. To appear. MR4024621 https://doi.org/10.1007/s42081-019-00039-x

Hudson, H. M. (1978). A natural identity for exponential families with applications in multiparameter estimation. *The Annals of Statistics* **6**, 473–484. MR0467991

Hwang, J. T. (1982). Improving upon standard estimators in discrete exponential families with applications to Poisson and negative binomial cases. *The Annals of Statistics* **10**, 857–867. MR0663437

Komaki, F. (2015). Simultaneous prediction for independent Poisson processes with different durations. *Journal of Multivariate Analysis* **141**, 35–48. MR3390057 https://doi.org/10.1016/j.jmva.2015.06.008

Morris, C. N. (1983). Parametric empirical Bayes inference: Theory and applications. *Journal of the American Statistical Association* **78**, 47–55. MR0696849

Tsui, K. W. (1979). Multiparameter estimation of discrete exponential distributions. *Canadian Journal of Statistics* **7**, 193–200. MR0570539 https://doi.org/10.2307/3315119

Tsui, K. W. and Press, S. J. (1982). Simultaneous estimation of several Poisson parameters under K-normalized squared error loss. *The Annals of Statistics* **10**, 93–100. MR0642721

Graduate School of Economics
University of Tokyo
7-3-1 Hongo
Bunkyo-ku
Tokyo 113-0033
Japan
E-mail: yasu.stat@gmail.com

Faculty of Economics
University of Tokyo
7-3-1 Hongo
Bunkyo-ku
Tokyo 113-0033
Japan
E-mail: tatsuya@e.u-tokyo.ac.jp