Computational Statistics
https://doi.org/10.1007/s00180-024-01554-6

ORIGINAL PAPER

Bayes estimation of ratio of scale-like parameters for inverse Gaussian distributions and applications to classification

Ankur Chakraborty[1] · **Nabakumar Jana**[1]

Received: 31 October 2022 / Accepted: 9 September 2024

Abstract

We consider two inverse Gaussian populations with a common mean but different scale-like parameters, where all parameters are unknown. We construct noninformative priors for the ratio of the scale-like parameters to derive matching priors of different orders. Reference priors are proposed for different groups of parameters. The Bayes estimators of the common mean and ratio of the scale-like parameters are also derived. We propose confidence intervals of the conditional error rate in classifying an observation into inverse Gaussian distributions. A generalized variable-based confidence interval and the highest posterior density credible intervals for the error rate are computed. We estimate parameters of the mixture of these inverse Gaussian distributions and obtain estimates of the expected probability of correct classification. An intensive simulation study has been carried out to compare the estimators and expected probability of correct classification. Real data-based examples are given to show the practicality and effectiveness of the estimators.

Keywords Probability matching priors · Reference priors · Confidence interval · Generalised variable approach · Bayes classification rule · Conditional error rate

1 Introduction

The inverse Gaussian (IG) distribution has applications in various fields such as engineering, actuarial science, medical science, environmental, and management sciences. The IG distribution is a good choice for modeling data with a long right tail and a relatively small mean. 'Together with the normal and gamma distributions, the inverse Gaussian completes the trio of families that are both an exponential and a group family of distributions' (Lehmann and Casella 2006, p. 68). The IG

✉ Nabakumar Jana
nabakumarjana@iitism.ac.in

[1] Department of Mathematics and Computing, Indian Institute of Technology (Indian School of Mines), Dhanbad, Jharkhand 826004, India

Published online: 19 September 2024

distribution is widely applied in life-testing and reliability analysis. Consider a toy example for classifying an observation into IG distributions. Suppose the observed lifetimes of two types of electronic components used in an automatic machine are available. It is known that the lifetime of two kinds of components follow IG distributions and their mean lifetimes are equal. One of the components has failed, and the failure time has been recorded. The problem is identifying the component type based on the observed lifetime. If the component is of type-1 and the classification rule assigns it to type-2 or vice versa, it will be misclassified. It is desired that the error rate in classiifcation (ERC) be minimized. We aim to derive several confidence intervals (CIs) and credible interval of the conditional ERC using the training samples from each population.

Ahmad et al. (1991) first considered the model of k IG distributions having an equal mean μ and developed MLE, and Graybill-Deal type estimator of μ. They explored the decision-theoretic properties of the estimator. Gupta and Akman (1995) considered the mixture model of IG and length-biased IG distribution and derived the Bayes estimators of the model parameters. Since the Bayes estimator of the mean is not in explicit form, different numerical techniques were used to solve it. Karlis (2002) considered the mixture model of normal and IG distribution and derived the estimators of the parameters using the method of moments. He also calculated the MLE of the parameters using the EM algorithm. Tian and Wilding (2005) used a modified direct likelihood ratio statistic to derive the CI of the ratio of two means of IG distributions. They used reciprocal root IG distribution to simplify the CI. Sindhu et al. (2018) studied different properties of the mixture of half-normal distributions. They proposed Bayes estimators of the parameters of the mixture model using noninformative priors under different loss functions.

Noninformative priors provide satisfactory results when little or no prior information is available. A probability matching prior is a noninformative prior designed to match the frequentist coverage probabilities (CPs) of certain regions. This means that the posterior probability of a region will be equal to the frequentist CP of that region. Bernardo (1979) derived noninformative priors by separating the parameters of interest and nuisance parameters. This approach is known as the reference prior approach. Berger and Bernardo (1989) introduced the idea of reverse reference prior, in which the parameter of interest and nuisance parameters are pretended to be interchanged. Kim et al. (2006) used noninformative priors to study the Bayesian inference for a linear combination of normal means. They derived the second order probability matching priors for the linear combination of the normal means as a function of other nuisance parameters. They also showed that these priors match the alternative CPs up to the second order. Considering two IG populations having an equal mean, the noninformative priors for the parameters are not studied in the literature.

There is extensive literature on classifying observations into normal populations. A few articles are focused on classification into non-normal or skewed distributions. Amoh (1985) derived the estimated classification function for a mixture of IG distributions with a common scale-like parameter. They analyzed the efficiency of the classification function for small samples. Conde et al. (2005) proposed classification rules for two exponential distributions under the restrictions on parameter. The proposed rule has lower misclassification probabilities than the likelihood ratio-based rule. Batsidis and Zografos (2006) studied the classification techniques for elliptically contoured populations with a common scale matrix. They considered separate discriminant functions for complete and incomplete samples when missing data were observed. Their proposed discriminant function is a linear combination of those two discriminant functions. For small samples, different point estimators of error rate are nearly unbiased but not consistent. In such situations, interval estimation of the conditional error rate provides a better alternative. The conditional ERC measures the probability of misclassification given a training data set. Chung and Han (2009) derived CIs of the conditional and unconditional error rate for classifying into two or more p-dimensional normal populations with a common covariance matrix. The CIs are obtained using bootstrap and jackknife methods, which improve other methods such as binomial approximation, k-fold cross-validation, and parametric approach. Jana and Chakraborty (2023) investigated classification problem for several normal populations with an equal mean and different variances. They used the bootstrap and jackknife procedures to calculate the conditional error rate's CI. Jana and Kumar (2019) studied the classification problem for two IG populations in various cases under order restriction on parameters. They also derived the likelihood ratio-based rule for two IG populations without restrictions and generalized the same for k populations. Note that the estimation of conditional ERC has not been studied under the Bayesian framework for IG populations.

In estimating the ERC, one requires the estimation of a function of parameters. While dealing with multiple parameters, as in frequentist inference, one may encounter challenges in constructing suitable pivotal quantities that effectively remove nuisance parameters. In such cases, probability-matching priors can be used to create approximate CIs. The current study has two objectives. First, noninformative priors are derived for the ratio of scale-like parameters λ_is of two IG populations having a common mean. Second, the classification problem for this model has been considered to show applications of the estimators besides proposing other classical intervals and credible intervals of conditional ERC. Since the finite mixtures of two IG distributions are used to model data sets robustly, we study classification into mixture of two IG populations.

The paper is arranged as follows. Sect. 2 introduces IG distributions having an equal mean. In Sect. 3.1, we derive noninformative prior for the ratio of scale-like parameters through the orthogonal parametrization. In Sect. 3.2, Bayes estimation of parametric functions of these distributions is obtained. In Sect. 4.1, we derive the credible intervals and generalized variable-based CIs besides other CIs of conditional error rate. Section 4.2 considers classification of observations into a mixture of IG distributions. Sect. 5 presents a thorough simulation study together with real-world instances. Finally, some conclusions are made in Sect. 6.

2 Preliminaries

Consider two independent inverse Gaussian populations Π_1 and Π_2 having an equal mean μ and scale-like parameters λ_1 and λ_2 respectively. The probability density function (pdf) corresponding to the population Π_i is

$$f_i(x) = \sqrt{\frac{\lambda_i}{2\pi x^3}} \exp\left\{-\frac{\lambda_i(x-\mu)^2}{2\mu^2 x}\right\}, \quad x > 0, \mu > 0, \lambda_i > 0 \tag{1}$$

for $i = 1, 2$. Suppose $X_{i1}, X_{i2}, \dots, X_{in_i}$ $(n_i \geq 2)$ represent a random sample from the population Π_i. Denote $\bar{X}_i = n_i^{-1} \sum_{j=1}^{n_i} X_{ij}$ and $S_i^{-1} = n_i^{-1} \sum_{j=1}^{n_i} (X_{ij}^{-1} - \bar{X}_i^{-1})$ for $i = 1, 2$. Note that $(\bar{X}_1, \bar{X}_2, S_1, S_2)$ is a minimal sufficient statistic for the parameter $(\mu, \lambda_1, \lambda_2)$ of the distributions. [Chhikara and Folks (1989)]. The statistics S_1 and S_2 are also the MLEs of λ_1 and λ_2, respectively.

3 Bayes estimation

3.1 Noninformative priors for the ratio of scale-like parameters

In Bayesian inference, a noninformative prior is used when limited or negligible information available about the data. To estimate a parameter of interest with some nuisance parameters, the orthogonality among them with respect to the expected Fisher information matrix plays an important role. A parameter θ_1 is said to be totally orthogonal to a set of parameters, say, $(\theta_2, \theta_3, \dots, \theta_k)$ if the information matrix is diagonal. The orthogonalization process and numerical simplification ensure that the corresponding rules are asymptotically independent. Noninformative priors help to achieve coverage error of function of parameters up to a particular order in the frequentist sense. Suppose the random samples $\underset{\sim}{x} = (x_1, \dots, x_{n_1})$ and $\underset{\sim}{y} = (y_1, \dots, y_{n_2})$ are drawn from IG(μ, λ_1) and IG(μ, λ_2) respectively. The log-likelihood function is

$$l(\mu, \lambda_1, \lambda_2 | \underset{\sim}{x}, \underset{\sim}{y}) = -\frac{n_1 + n_2}{2} \ln(2\pi) - \frac{3}{2}\left(\sum_{i=1}^{n_1} \ln x_i + \sum_{j=1}^{n_2} \ln y_j\right) + \frac{n_1}{2} \ln \lambda_1 + \frac{n_2}{2} \ln \lambda_2$$
$$- \sum_{i=1}^{n_1} \frac{\lambda_1 (x_i - \mu)^2}{2\mu^2 x_i} - \sum_{j=1}^{n_2} \frac{\lambda_2 (y_j - \mu)^2}{2\mu^2 y_j}. \tag{2}$$

Consider a prior distribution π to estimate the parameter $\theta_1 = \lambda_2/\lambda_1$. Let $\theta_1^\alpha(\pi;\mathbf{x})$ denotes the $(1-\alpha)$th percentile of the posterior distribution of θ_1, that is,

$P^\pi[\theta_1 \leq \theta_1^\alpha(\pi;\mathbf{x})|\mathbf{x}] = 1 - \alpha$, where θ_1 is the parameter of interest. We want to find the priors π for which $P^\pi\left[\theta_1 \leq \theta_1^\alpha(\pi;\mathbf{x})|\mathbf{x}\right] = 1 - \alpha + o(n^{-1})$ and make the prior a second-order matching prior. To find such priors, we consider the orthogonal parametrization techniques (Cox and Reid 1987; Tibshirani 1989). We find the orthogonal parameters θ_1, θ_2 and θ_3. Denote $\underset{\sim}{\theta} = (\theta_1, \theta_2, \theta_3)$. Under the transformations $\lambda_1 \to \phi_1,\ \lambda_2 \to \phi_1\psi,\ \mu \to \phi_2$, the expression (2) becomes

$$l(\phi_1, \phi_2, \psi | \underset{\sim}{x}, \underset{\sim}{y}) = -\frac{n_1 + n_2}{2} \ln(2\pi) - \frac{3}{2}\left(\sum_{i=1}^{n_1} \ln x_i + \sum_{j=1}^{n_2} \ln y_j\right) + \frac{n_1 + n_2}{2} \ln \phi_1 + \frac{n_2}{2} \ln \psi - \sum_{i=1}^{n_1} \frac{\phi_1 (x_i - \phi_2)^2}{2\phi_2^2 x_i} - \sum_{j=1}^{n_2} \frac{\phi_1 \psi (y_j - \phi_2)^2}{2\phi_2^2 y_j}.$$

The orthogonal equations to find ϕ_1, ϕ_2, ψ and subsequently the value of θ_2, θ_3 are provided in the supplementary material. Using orthogonal parametrization of the original parameters, we get $\theta_1 = \lambda_2/\lambda_1$, $\theta_2 = \lambda_1^{n_1} \lambda_2^{n_2}$, $\theta_3 = \mu$ which can be written as $\lambda_1 = \theta_1^{-n_2/(n_1+n_2)} \theta_2^{1/(n_1+n_2)}$, $\lambda_2 = \theta_1^{n_1/(n_1+n_2)} \theta_2^{1/(n_1+n_2)}$, $\mu = \theta_3$. Then the log-likelihood function (2) is written in the form of θ_1, θ_2 and θ_3 as

$$l(\underset{\sim}{\theta} | \underset{\sim}{x}, \underset{\sim}{y}) = \frac{\ln \theta_2}{2} - \theta_1^{-n_2/(n_1+n_2)} \theta_2^{1/(n_1+n_2)} \sum_{i=1}^{n_1} \frac{(x_i - \theta_3)^2}{2\theta_3^2 x_i} - \theta_1^{n_1/(n_1+n_2)} \theta_2^{1/(n_1+n_2)} \sum_{j=1}^{n_2} \frac{(y_j - \theta_3)^2}{2\theta_3^2 y_j} + c. \tag{3}$$

From Eq. (3), the elements of the information matrix are obtained as

$$I_{11} = -E\left[\frac{\partial^2 l}{\partial \theta_1^2}\right] = \frac{n_1 n_2 \theta_1^{-2}}{2(n_1 + n_2)}, \quad I_{22} = -E\left[\frac{\partial^2 l}{\partial \theta_2^2}\right] = \frac{\theta_2^{-2}}{2(n_1 + n_2)},$$

$$I_{33} = -E\left[\frac{\partial^2 l}{\partial \theta_3^2}\right] = \theta_2^{1/(n_1+n_2)} \theta_3^{-3} \left(n_1 \theta_1^{-n_2/(n_1+n_2)} + n_2 \theta_1^{n_1/(n_1+n_2)}\right),$$

$$I_{12} = -E\left[\frac{\partial^2 l}{\partial \theta_1 \partial \theta_2}\right] = 0, \quad I_{13} = -E\left[\frac{\partial^2 l}{\partial \theta_1 \partial \theta_3}\right] = 0, \quad I_{23} = -E\left[\frac{\partial^2 l}{\partial \theta_2 \partial \theta_3}\right] = 0.$$

Since θ_1 is orthogonal to θ_2 and θ_3, Tibshirani (1989) defines the class of first-order probability matching (FOPM) prior as

$$\pi^{(1)}(\underset{\sim}{\theta}) \propto \theta_1^{-1} d(\theta_2, \theta_3), \tag{4}$$

where $d(\theta_2, \theta_3)$ is an arbitrary differentiable function.

Theorem 1 *The second-order probability matching (SOPM) priors are given by* $\pi^{(2)}(\underset{\sim}{\theta}) = \theta_1^{-1} \theta_2^{-1} d(\theta_3)$, *where* $d(\theta_3)$ *is any smooth function of* θ_3.

Proof The proof of the theorem is provided in the supplementary material. □

If the following set of differential equations are satisfied, under orthogonal parametrization, a SOMP prior agrees with alternative CPs up to the second-order (Mukerjee and Reid 1999). The equations are

$$\frac{\partial}{\partial\theta_2}\left\{L_{112}I^{22}I_{11}^{-1/2}d(\theta_2,\theta_3)\right\}+\frac{\partial}{\partial\theta_3}\left\{L_{113}I^{33}I_{11}^{-1/2}d(\theta_2,\theta_3)\right\}=0,$$
$$\frac{\partial}{\partial\theta_2}\left\{L_{2,11}I^{22}I_{11}^{-1/2}d(\theta_2,\theta_3)\right\}+\frac{\partial}{\partial\theta_3}\left\{L_{3,11}I^{33}I_{11}^{-1/2}d(\theta_2,\theta_3)\right\}=0,$$
$$\frac{\partial}{\partial\theta_1}\left\{I_{11}^{-\frac{3}{2}}L_{111}\right\}=0 \text{ and } \frac{\partial}{\partial\theta_1}\left\{I_{11}^{-\frac{3}{2}}L_{1,11}\right\}=0.$$

Under this setup,

$$L_{111}=E\left[\frac{\partial^3 l}{\partial\theta_1^3}\right]=\frac{n_1n_2(n_1+2n_2)}{(n_1+n_2)^2}\theta_1^{-3},\quad L_{1,11}=E\left[\frac{\partial l}{\partial\theta_1}\frac{\partial^2 l}{\partial\theta_1^2}\right]=-\frac{n_1n_2^2}{(n_1+n_2)^2}\theta_1^{-3},$$
$$L_{2,11}=E\left[\frac{\partial l}{\partial\theta_2}\frac{\partial^2 l}{\partial\theta_1^2}\right]=\frac{n_1n_2(n_1+n_2+1)}{4(n_1+n_2)^2}\theta_1^{-2}\theta_2^{-1} \text{ and } L_{3,11}=E\left[\frac{\partial l}{\partial\theta_3}\frac{\partial^2 l}{\partial\theta_1^2}\right]=0.$$

After incorporating the expressions of $L_{111}, L_{1,11}, L_{2,11}, L_{3,11}$, the above equations hold.

Note 1 It can be verified that $I_{11}^{-3/2}L_{111}$ is independent of θ_1. Hence the SOMPs proposed here are highest posterior density (HPD) matching priors up to the second-order.

First, we established that the parameters $\theta_1, \theta_2, \theta_3$ are orthogonal. Then, the reference priors for various groups of the ordering of the parameters $\theta_1, \theta_2, \theta_3$ are derived by following the work of Datta and Ghosh (1995).

Group ordering: $\{(\theta_1, \theta_2, \theta_3)\}$: The reference prior is of the form

$$\pi_1(\underset{\sim}{\theta}) \propto \theta_2^{-\frac{2(n_1+n_2)-1}{2(n_1+n_2)}}\theta_3^{-3/2}\left(n_1\theta_1^{-\frac{2n_1+3n_2}{2(n_1+n_2)}}+n_2\theta_1^{-\frac{n_1+2n_2}{2(n_1+n_2)}}\right).$$

$\{\theta_1, \theta_2, \theta_3\}, \{(\theta_1, \theta_2), \theta_3\}$: The form of the reference prior is given by

$$\pi_2(\underset{\sim}{\theta}) \propto (\theta_1\theta_2)^{-1}\theta_3^{-3/2}.$$

$\{\theta_1, (\theta_2, \theta_3)\}$: The reference prior is of the form

$$\pi_3(\underset{\sim}{\theta}) \propto \theta_1^{-1}\theta_2^{-\frac{2(n_1+n_2)-1}{2(n_1+n_2)}}\theta_3^{-3/2}.$$

$\{(\theta_1, \theta_3), \theta_2)\}$: The form of the reference prior is given below.

$$\pi_4(\underset{\sim}{\theta}) \propto \theta_2^{-1}\theta_3^{-3/2}\left(n_1\theta_1^{-\frac{2n_1+3n_2}{2(n_1+n_2)}} + n_2\theta_1^{-\frac{n_1+2n_2}{2(n_1+n_2)}}\right).$$

Note 2 Note that the proposed reference priors $\pi_2(\underset{\sim}{\theta})$ and $\pi_3(\underset{\sim}{\theta})$ are the FOPMs. The prior $\pi_2(\underset{\sim}{\theta})$ is the SOPM prior.

In Sect. 4.1, we obtain credible interval of ERC using the reference prior of the form $\pi_2(\underset{\sim}{\theta})$.

3.2 Bayes estimators of function of parameters

Let X be IG(μ, λ) distributed with the pdf of the form (1). The gamma family is a conjugate prior to the IG distribution with a known mean. However, when μ and λ are unknown, the conjugate prior is unknown for the IG distribution. We have considered two IG distributions having an equal mean but different λ_is where finding a conjugate prior distribution is a real challenge. The coefficient of variation of X is $\sqrt{\mu/\lambda}$. Considering the reparametrization of $\lambda = \mu\phi$, the pdf is written as

$$f(x;\mu,\phi) = \sqrt{\frac{\mu\phi}{2\pi x^3}}\exp\left\{-\frac{\phi(x-\mu)^2}{2\mu x}\right\},\ x>0, \mu>0, \phi>0. \quad (5)$$

Let $x_1, x_2, \dots, x_n$ be a random sample from the IG(μ, ϕ) distribution. The likelihood function based on $x_1, x_2, \dots, x_n$ is

$$l(\mu,\phi|\underset{\sim}{x}) = \left(\frac{\mu\phi}{2\pi}\right)^{n/2}\prod_{i=1}^{n} x_i^{-3/2}\exp(n\phi)\exp\left\{\frac{\phi n}{2}\left(\frac{\bar{x}}{\mu}+\mu\bar{x}_r\right)\right\},$$

where $\bar{x} = n^{-1}\sum_{i=1}^{n} x_i$ and $\bar{x}_r = n^{-1}\sum_{i=1}^{n} x_i^{-1}$. Assume the prior information about μ and ϕ is summarized in the density $\pi(\mu,\phi) = \pi(\mu|\phi)\pi(\phi)$, where

$$\pi(\mu|\phi) = \left(\frac{\eta\phi\omega}{2\pi}\right)^{1/2}\mu^{-3/2}\exp(\phi\omega)\exp\left\{-\frac{\phi\omega}{2}\left(\frac{\eta}{\mu}+\frac{\mu}{\eta}\right)\right\} \text{ and}$$
$$\pi(\phi) = \frac{a^\gamma}{\Gamma(\gamma)}\phi^{\gamma-1}e^{-a\phi},\ \phi>0.$$

Hence

$$\pi(\mu,\phi) = \left(\frac{\eta\omega}{2\pi}\right)^{1/2}\phi^{\gamma-1/2}e^{\phi(\omega-a)}\mu^{-3/2}\frac{a^\gamma}{\Gamma(\gamma)}\exp\left\{-\frac{\phi\omega}{2}\left(\frac{\eta}{\mu}+\frac{\mu}{\eta}\right)\right\},\ \ \mu,\phi>0.$$

The joint probability density function of $x_1, x_2, \dots, x_n, \mu, \phi$ is

$$\sqrt{\frac{\eta\omega}{(2\pi)^{n+1}}}\frac{a^{\gamma}}{\Gamma(\gamma)}\prod_{i=1}^{n}x_i^{-3/2}\mu^{\frac{n-3}{2}}\phi^{\gamma+\frac{n-1}{2}}\exp\{\phi(n+\omega-a)\}$$
$$\times\exp\left\{-\frac{\phi}{2}\left(\frac{n\bar{x}}{\mu}+n\mu\bar{x}_r+\frac{\eta\omega}{\mu}+\frac{\mu\omega}{\eta}\right)\right\}.$$

Now the joint posterior density function of (μ,ϕ) is given by

$$\pi(\mu,\phi|\underset{\sim}{x})=c\mu^{(n-3)/2}\cdot\phi^{(\gamma'-1)/2}\exp\left\{-\phi\left(\frac{\nu_1}{2\mu}-\nu_2+\frac{\nu_3\mu}{2}\right)\right\},\ \mu>0,\phi>0,$$

where c is the normalizing constant. The posterior density of ϕ is

$$\pi(\phi|\underset{\sim}{x})=c\phi^{(\gamma'-1)/2}\cdot\exp(\phi\nu_2)\int_0^{\infty}\mu^{\frac{n-3}{2}}\exp\left\{-\frac{\phi}{2}\left(\frac{\nu_1}{\mu}+\nu_3\mu\right)\right\}$$
$$=2c\phi^{\frac{\gamma'-1}{2}}\left(\frac{\nu_1}{\nu_3}\right)^{\frac{n-1}{4}}\exp(\phi\nu_2)K_{\frac{n-1}{2}}(\phi\sqrt{\nu_1\nu_3}),$$

where K_n is the Bessel function of third kind with index n. Next, we consider two IG distributions IG(μ,ϕ_1) and IG(μ,ϕ_2). The ratio of the coefficient of variation becomes ratio of λ_is of the distributions. We find an estimator of λ_1/λ_2. Since two populations are independent, the joint density of (ϕ_1,ϕ_2) given the data is $\pi(\phi_1,\phi_2|\underset{\sim}{x},\underset{\sim}{y})=\pi(\phi_1|\underset{\sim}{x})\pi(\phi_2|\underset{\sim}{y})$. The Bayes estimator of λ_1/λ_2 is

$$\int_0^{\infty}\int_0^{\infty}4c_1c_2\phi_1^{(\gamma_1'+1)/2}\phi_2^{(\gamma_2'-3)/2}\left(\frac{\nu_{11}}{\nu_{31}}\right)^{\frac{n_1-1}{4}}\left(\frac{\nu_{12}}{\nu_{32}}\right)^{\frac{n_2-1}{4}}\exp(\phi_1\nu_{21}+\phi_2\nu_{22})$$
$$\times K_{\frac{n_1-1}{2}}(\phi_1\sqrt{\nu_{11}\nu_{31}})K_{\frac{n_2-1}{2}}(\phi_2\sqrt{\nu_{12}\nu_{32}})\,d\phi_1\,d\phi_2.$$

Next, we consider estimation of μ. The joint density function of μ,ϕ_1 and ϕ_2 is

$$\pi(\mu,\phi_1,\phi_2)=\frac{\eta\omega}{2\pi}\left\{\prod_{i=1}^{2}\phi_i^{\gamma-1/2}\right\}\mu^{-3}\left(\frac{a^{\gamma}}{\Gamma(\gamma)}\right)^2\exp\left\{\sum_{i=1}^{2}\phi_i\left(\omega-a-\frac{\omega}{2}\left(\frac{\eta}{\mu}+\frac{\mu}{\eta}\right)\right)\right\}.$$

Then the likelihood function is

$$L(\mu,\phi_1,\phi_2|\underset{\sim 1}{x},\underset{\sim 2}{x})=\prod_{i=1}^{2}\left\{\left(\frac{\phi_i\mu}{2\pi}\right)^{n_i/2}\prod_{j=1}^{n_i}x_{ij}^{-3/2}\right\}\exp\left[\sum_{i=1}^{2}\phi_i\left\{n_i-\frac{1}{2}\left(\frac{x_{ij}}{\mu}+\frac{\mu}{x_{ij}}\right)\right\}\right].$$

The posterior density of μ is

$$\pi(\mu|\phi_1,\phi_2)=\frac{\tilde{L}(\mu,\phi_1,\phi_2)}{\int_0^{\infty}\tilde{L}(\mu,\phi_1,\phi_2)\,d\mu},$$

where $\tilde{L}(\mu, \phi_1, \phi_2) = L(\mu, \phi_1, \phi_2 | \underset{\sim 1}{x}, \underset{\sim 2}{x}) \pi(\mu, \phi_1, \phi_2)$. The Bayes estimator of μ is

$$\hat{\mu} = \frac{\int_0^\infty \mu^{\frac{n_1+n_2}{2}-2} \exp\left[-\sum_{i=1}^{2} \frac{\phi_i}{2}\left\{\omega\left(\frac{\eta}{\mu}+\frac{\mu}{\eta}\right)+\sum_{j=1}^{n_i}\left(\frac{x_{ij}}{\mu}+\frac{\mu}{x_{ij}}\right)\right\}\right] d\mu}{\int_0^\infty \mu^{\frac{n_1+n_2}{2}-3} \exp\left[-\sum_{i=1}^{2} \frac{\phi_i}{2}\left\{\omega\left(\frac{\eta}{\mu}+\frac{\mu}{\eta}\right)+\sum_{j=1}^{n_i}\left(\frac{x_{ij}}{\mu}+\frac{\mu}{x_{ij}}\right)\right\}\right] d\mu}.$$

The estimator $\hat{\mu}$ is used to estimate the classification function and compare error rates in the next section.

4 Application to classification problem

The present section deals with applications of the considered IG distributions to the classification problem. Suppose $P(i|j)$ is the probability of misclassification of an observation from Π_j to Π_i and $C(i|j)$ is the coressponding cost of misclassification $i \neq j (= 1, 2)$. Given a new observation z, the following classification regions

$$\mathrm{S}_1 = \left\{ z : \frac{f_1(z)}{f_2(z)} \geq \frac{C(1|2)q_2}{C(2|1)q_1} \right\} \text{ and } \mathrm{S}_2 = \left\{ z : \frac{f_1(z)}{f_2(z)} < \frac{C(1|2)q_2}{C(2|1)q_1} \right\},$$

are obtained by minimizing the expected misclassification cost, where q_i represents the prior probability of belonging an observation into the population Π_i, $i = 1, 2$. The expected probability of correct classification (EPC) is defined as $\sum_{i=1}^{2} q_i P(i|i)$. Assume that $C(1|2) = C(2|1)$ and $q_1 = q_2$. Following (Anderson 2003), the Bayes classification rule (R) for assigning a new observation z is: classify z into Π_1 if $W > 0$, otherwise classify it to Π_2, where

$$W = \frac{(z-\mu)^2}{\mu^2 z}(\lambda_2 - \lambda_1) + \ln \lambda_1 - \ln \lambda_2.$$

4.1 Confidence intervals of ERC

We want to estimate the ERC for two such IG populations. Suppose $\gamma = (P_1 + P_2)/2 = (P[W < 0 | z \in \Pi_1] + P[W > 0 | z \in \Pi_2])/2$, denotes the unconditional error rate. Since $\mu, \lambda_1, \lambda_2$ are unknown, we use their estimates to get the conditional error rate $\gamma^* = (P_1^* + P_2^*)/2$, where $P_1^* = P(W < 0 | \bar{x}_1, \bar{x}_2, S_1, S_2; z \in \Pi_1)$ and $P_2^* = P(W \geq 0 | \bar{x}_1, \bar{x}_2, S_1, S_2; z \in \Pi_1)$. Then

$$\gamma^* = \begin{cases} [g(\hat{\lambda}_2 \hat{\lambda}^*) - g(\hat{\lambda}_1 \hat{\lambda}^*) + 1]/2, & \text{if} \hat{\lambda}_1 > \hat{\lambda}_2 \\ [g(\hat{\lambda}_1 \hat{\lambda}^*) - g(\hat{\lambda}_2 \hat{\lambda}^*) + 1]/2, & \text{if} \hat{\lambda}_1 < \hat{\lambda}_2, \end{cases}$$

where $\hat{\lambda}_i$ is an estimator of λ_i, $g(\cdot)$ is the cumulative distribution function of χ^2_1 and $\hat{\lambda}^* = (\ln\hat{\lambda}_1 - \ln\hat{\lambda}_2)/(\hat{\lambda}_1 - \hat{\lambda}_2)$.

An estimator $\hat{\gamma}^*$ of γ^* is obtained from a set of random samples from the populations and the corresponding B estimators $\hat{\gamma}^*_1, \dots, \hat{\gamma}^*_B$ of γ^* is calculated from B bootstrap samples. Let s^2_γ denote the sample variance of γ^*_i's and $\hat{\gamma}^*_{(i)}$ is the ordered estimates of γ^*_i's. Four types of $100(1-2\eta)\%$ CIs are presented below:

(a) The conditional CI of γ^* using the symmetric method is $(\hat{\gamma}^* - z_\eta s_\gamma, \hat{\gamma}^* + z_\eta s_\gamma)$.
(b) The conditional CI using the percentile method is given by $(\hat{\gamma}^*_{(r)}, \hat{\gamma}^*_{(s)})$. [Jana and Chakraborty (2023)].
(c) Suppose q denotes the number of bootstrap estimates of γ^* that are smaller than $\hat{\gamma}^*$. Define $z_0 = \phi^{-1}(q/B)$, $\eta_{BL} = \Phi(2z_0 - z_\eta)$ and $\eta_{BR} = \Phi(2z_0 + z_\eta)$, where $\Phi(z_\eta) = 1 - \eta$ and Φ is the standard normal distribution function. The conditional CI using the bias-corrected percentile method is given by $(\hat{\gamma}^*_{(j)}, \hat{\gamma}^*_{(k)})$, where $j = (B+1)\eta_{BL}$ and $k = (B+1)\eta_{BR}$. The conditional error rate using the accelerated bias-corrected percentile (Abcp) method is given by $(\hat{\gamma}^*_{(u)}, \hat{\gamma}^*_{(v)})$, where $u = (B+1)\eta_{AL}$ and $v = (B+1)\eta_{AR}$. We refer to Jana and Chakraborty (2023) for details.
(d) We use the jackknife resampling method to create the following confidence interval (CI) for the conditional error rate. [see Jana and Chakraborty (2023)]

$$\left\{\hat{\gamma}^{**} - t_{n-1,\alpha/2}\sqrt{\frac{\sum_{i=1}^{n}(\hat{\gamma}^{**}_i - \hat{\gamma}^{**})^2}{n(n-1)}}, \hat{\gamma}^{**} + t_{n-1,\alpha/2}\sqrt{\frac{\sum_{i=1}^{n}(\hat{\gamma}^{**}_i - \hat{\gamma}^{**})^2}{n(n-1)}}\right\},$$

Next, the pivotal quantities for the parameters proposed by Ye et al. (2010) are used to obtain the CI. Suppose $T(X, x, \theta_1, \theta_2)$ is a generalized pivot quantity for the parameter of interest θ_1, where x denotes the observed value of the random variable X and θ_2 is the nuisance parameter. Then $T(X, x, \theta_1, \theta_2)$ must satisfy the following conditions:

1. The distribution function of $T(X, x, \theta_1, \theta_2)$ is free from the unknown parameters.
2. The observed value $T(x, x, \theta_1, \theta_2)$ of the pivot quantity $T(X, x, \theta_1, \theta_2)$ is free from the parameter θ_2.

Suppose $\bar{x}_i$ and s_i are the observed values of $\bar{X}_i$ and S_i, respectively. A generalized pivot quantity for λ_i is defined as

$$T_i = \frac{n_i\lambda_i S_i}{n_i s_i} \sim \frac{\chi^2_{n_i-1}}{n_i s_i},\ i = 1, 2.$$

Define $T = \dfrac{(n_1-1)n_2S_2}{(n_2-1)n_1S_1}$. Note that T follows a F-distribution with degrees of freedom with $(n_1-1),(n_2-1)$ and T is a generalized variable for the parameter λ_1/λ_2. We propose the following Algorithm for deriving CI of error rate.

Algorithm 1

Step 1. Generate $x_1, x_2, \ldots, x_{n_1} \sim$IG$(\mu, \lambda_1)$ and $y_1, y_2, \ldots, y_{n_2} \sim$IG$(\mu, \lambda_2)$.

Step 2. Find $\bar{x}, \bar{y}, S_1$ and S_2.

Step 3. Generate a random sample from $F(n_1-1, n_2-1)$.

Step 4. Let $T^* = \log T/(T-1)$. Consider T^* as an estimate of λ^*.

Step 5. Calculate γ^*.

Step 6. Repeat Steps 3 to 5, B times.

Step 7. Find the order statistic corresponding to the γ^*s. The $100(1-\alpha)\%$ CI of γ^* is $(\gamma^*_{(1)}, \gamma^*_{[(1-\alpha)*B]})$.

In Sect. 5.1, we perform a detailed simulation study to compare the CIs using the proposed methods for two IG populations having an equal mean. Comparisons of the CIs for three populations using bootstrap and jackknife techniques are also studied.

In addition, we propose to use Bayesian credible intervals to estimate the classification error rate. The HPD credible intervals are the shortest credible intervals that contain the true error rate with a certain probability. All points within the HPD interval have a higher posterior probability than any points outside the interval. The technique to compute such intervals was introduced by Chen and Shao (1999), involving the utilization of samples generated from the posterior density through the Markov chain Monte Carlo method. Assume that $x_1, x_2, \ldots, x_{n_1}$ and $y_1, y_2, \ldots, y_{n_2}$ are random samples originating from IG(μ, λ_1) and IG(μ, λ_2) distributions, respectively. The joint density function of $\underset{\sim}{x}$ and $\underset{\sim}{y}$, as per the reparameterization detailed in Sect. 3.1, is expressed as

$$f(\underset{\sim}{x}, \underset{\sim}{y}|\underset{\sim}{\theta}) = \frac{\theta_2^{\frac{1}{2}}}{(2\pi)^{\frac{n}{2}} \prod_{i=1}^{n_1} x_i^{\frac{3}{2}} \prod_{j=1}^{n_2} y_j^{\frac{3}{2}}} \exp\left[-\frac{\theta_2^{\frac{1}{n}}}{2\theta_3^2}\left\{\sum_{i=1}^{n_1} \frac{\theta_1^{-\frac{n_2}{n}}(x_i-\theta_3)^2}{x_i} + \sum_{j=1}^{n_2} \frac{\theta_1^{\frac{n_1}{n}}(y_j-\theta_3)^2}{y_j}\right\}\right],$$

where $n = n_1 + n_2$. Given the second-order matching prior of the form $\pi_2(\underset{\sim}{\theta}) \propto \theta_1^{-1}\theta_2^{-1}\theta_3^{-3/2}$, we derive the posterior density of θ_1 as follows:

$$\pi(\theta_1|\underset{\sim}{x}, \underset{\sim}{y}, \theta_2, \theta_3) = \frac{\theta_1^{-1} \exp\left[-\frac{\theta_2^{\frac{1}{n}}}{2\theta_3^2}\left\{\sum_{i=1}^{n_1} \frac{\theta_1^{-\frac{n_2}{n}}(x_i-\theta_3)^2}{x_i} + \sum_{j=1}^{n_2} \frac{\theta_1^{\frac{n_1}{n}}(y_j-\theta_3)^2}{y_j}\right\}\right]}{\int_0^\infty \theta_1^{-1} \exp\left[-\frac{\theta_2^{\frac{1}{n}}}{2\theta_3^2}\left\{\sum_{i=1}^{n_1} \frac{\theta_1^{-\frac{n_2}{n}}(x_i-\theta_3)^2}{x_i} + \sum_{j=1}^{n_2} \frac{\theta_1^{\frac{n_1}{n}}(y_j-\theta_3)^2}{y_j}\right\}\right] d\theta_1}.$$

To generate samples from the posterior density, we use the estimates of θ_2 and θ_3 as

$$\hat{\theta}_2 = \left\{ \frac{1}{n_1} \sum_{i=1}^{n_1} \left(\frac{1}{x_i} - \frac{1}{\bar{x}} \right) \right\}^{-n_1} \left\{ \frac{1}{n_2} \sum_{j=1}^{n_2} \left(\frac{1}{y_j} - \frac{1}{\bar{y}} \right) \right\}^{-n_2}$$

$$\text{and } \hat{\theta}_3 = \frac{1}{n_1 + n_2} \left(\sum_{i=1}^{n_1} x_i + \sum_{j=1}^{n_2} y_j \right),$$

respectively. Note that the error rate γ^* is a function of θ_1 only and is written as

$$\gamma^* = \begin{cases} \left[g\left(\frac{\ln \theta_1}{\theta_1 - 1} \right) - g\left(\frac{\theta_1 \ln \theta_1}{\theta_1 - 1} \right) + 1 \right] / 2, & \text{if } \theta_1 \geq 1 \\ \left[g\left(\frac{\theta_1 \ln \theta_1}{\theta_1 - 1} \right) - g\left(\frac{\ln \theta_1}{\theta_1 - 1} \right) + 1 \right] / 2, & \text{if } \theta_1 < 1, \end{cases}$$

where $g(\cdot)$ is the cumulative distribution function of χ_1^2. Due to the complexity of the posterior density, we use the algorithm proposed by Chen and Shao (1999) to generate random samples from this density and subsequently find the credible interval for γ^*. Given random samples $\underset{\sim}{x}$ and $\underset{\sim}{y}$, the following procedure outlines the necessary steps to compute a $100(1 - \alpha)\%$ credible interval for γ^*.

Algorithm 2

Step 1. Define a current value $\theta_1^{(0)}$ from the target density function.

Step 2. Generate a random sample z_* from a proposed density $h(x|\theta_1^{(j)})$, $j = 0, 1, 2, \ldots$. Find the acceptance probability

$$\kappa(\theta_1^{(j)}, z_*) = \min \left\{ 1, \frac{\pi(z_*) h(\theta_1^{(j)} | z_*)}{\pi(\theta_1^{(j)}) h(z_* | \theta_1^{(j)})} \right\}.$$

Step 3. Generate a random sample u from U(0, 1) distribution. If $u < \kappa(\theta_1^{(j)}, z_*)$, set $\theta_1^{(j+1)} = z_*$; otherwise, set $\theta_1^{(j+1)} = \theta_1^{(j)}$.

Step 4. Obtain the error rate γ^* by putting $\theta_1 = \theta_1^{(j)}$.

Step 5. Set $j = j + 1$.

Step 6. Repeat steps 2-5, M times.

Step 7. From the γ^* values, obtain the order statistics $\gamma^*_{(1)}, \gamma^*_{(2)}, \ldots, \gamma^*_{(M)}$.

Then the $100(1 - \alpha)\%$ HPD credible interval for γ^* is given by $(\gamma^*_{([\frac{\alpha}{2}M])}, \gamma^*_{([(1-\frac{\alpha}{2})M])})$.

In Sect. 5.1, we apply this algorithm to compute the credible intervals for γ^*.

4.2 Classification into mixture of IG distributions

In this section, we study estimation of the parameters of a mixture of two inverse Gaussian (MTIG) distributions and the corresponding discriminant function. The density of the MTIG distribution having an equal mean parameter is

$$f(x; \Theta) = \sum_{j=1}^{2} p_j f_j(x; \Theta_j), \quad p_1 + p_2 = 1, \tag{6}$$

where $\Theta = (p_1, \mu, \phi_1, \phi_2), \Theta_j = (\mu, \phi_j), j = 1, 2$ and $f_j(x;\Theta_j)$ is the pdf of the jth univariate IG distribution as given in (5). Suppose $x_1, x_2, \dots, x_n$ is a random sample drawn from the MTIG distribution with density (6). The likelihood function based on the random sample is $L(\Theta) = \prod_{i=1}^{n} \sqrt{(\mu/(2\pi x_i^3))Q_i}$, where

$$Q_i = \sum_{j=1}^{2} p_j \sqrt{\phi_j} \exp\left\{ -\frac{\phi_j (x_i - \mu)^2}{2\mu x_i} \right\}.$$

The log-likelihood function is $l(\Theta) = -\sum_{i=1}^{n} \{\ln(2\pi x_i^3)\}/2 + \sum_{i=1}^{n} \ln Q_i$. The likelihood equations are given by

$$\sum_{i=1}^{n} \frac{1}{Q_i}\left[\sqrt{\phi_1} \exp\left\{ -\frac{\phi_1 (x_i - \mu)^2}{2\mu x_i} \right\} - \sqrt{\phi_2} \exp\left\{ -\frac{\phi_2 (x_i - \mu)^2}{2\mu x_i} \right\}\right] = 0,$$

$$\sum_{i=1}^{n} \frac{(x_i^2 - \mu^2)}{x_i Q_i}\left[p_1 \phi_1^{3/2} \exp\left\{ -\frac{\phi_1 (x_i - \mu)^2}{2\mu x_i} \right\} + p_2 \phi_2^{3/2} \exp\left\{ -\frac{\phi_2 (x_i - \mu)^2}{2\mu x_i} \right\}\right] + n\mu = 0,$$

$$\sum_{i=1}^{n} \frac{p_1}{Q_i}\left[\exp\left\{ -\frac{\phi_1 (x_i - \mu)^2}{2\mu x_i} \right\}\left\{ 1 - \frac{\phi_1 (x_i - \mu)^2}{\mu x_i} \right\}\right] = 0,$$

$$\sum_{i=1}^{n} \frac{p_2}{Q_i}\left[\exp\left\{ -\frac{\phi_2 (x_i - \mu)^2}{2\mu x_i} \right\}\left\{ 1 - \frac{\phi_2 (x_i - \mu)^2}{\mu x_i} \right\}\right] = 0.$$

Finding estimates of the parameters from the above equations is not always possible due to the computational complexity. So, we use the EM algorithm to find the MLEs of the parameters $\Theta = (p_1, \mu, \phi_1, \phi_2)$. Let $\mathbf{x} = (x_1, \dots, x_n)$ be an observed sample from the mixture distribution $f(x;\Theta)$. Assume the component from which x_i originates is unknown. This missing data is labeled as y_{ij}, where y_{ij} signifies x_i derived from the jth component for $j = 1, 2$. The log-likelihood function based on the observed data is

$$l(p_1, \mu, \phi_1, \phi_2 | \mathbf{x}) = \sum_{i=1}^{n} \log \sum_{j=1}^{2} p_j f_j(x_i | \mu, \phi_j).$$

The complete data log-likelihood satisfies

$$l_c(p_1, \mu, \phi_1, \phi_2 | \mathbf{x}, \mathbf{y}) = \sum_{i=1}^{n} \sum_{j=1}^{2} y_{ij} \left\{ \log p_j + \frac{1}{2} \log(\mu\phi_j) - \frac{1}{2} \log(2\pi x_i^3) - \frac{\phi_j (x_i - \mu)^2}{2\mu x_i} \right\}.$$

In EM algorithm, we apply two steps namely an expectation step (E-step) and a maximization step (M-step) alternately to get the sequence $\{(p_1^{(k)}, \mu^{(k)}, \phi_1^{(k)}, \phi_2^{(k)})\}_{k\in\mathbb{N}}$ of estimators. We continue to generate the terms of the sequence till it converges and maximizes the likelihood function. Dempster et al. (1977) proved the convergence of the EM algorithm to a local maxima. Let w_{ij} be the conditional probability that x_i arises from the mixture component indexed j, having density $f_j(\cdot | \mu, \phi_j)$, given the sample data $\mathbf{x}$. Using Bayes rule, for each i and j, we have

$$w_{ij} = \frac{p_j f_j\left(x_i|\mu^{(k)}, \phi_j^{(k)}\right)}{\sum_{j=1}^{2} p_j f_j\left(x_i|\mu^{(k)}, \phi_j^{(k)}\right)}.$$

We mention the kth iteration scheme for the E-step and M-step of the EM algorithm.

E-step: For $i = 1, 2, \dots, n$ and $j = 1, 2$, evaluate

$$w_{ij}^{(k)} = \frac{p_j^{(k)} f_j(x_i|\mu^{(k)}, \phi_j^{(k)})}{\sum_{j=1}^{2} p_j^{(k)} f_j(x_i|\mu^{(k)}, \phi_j^{(k)})}.$$

M-step: By replacing y_{ij} with w_{ij}, we need to maximize the log likelihood of the complete data.

$$p_1^{(k+1)} = \frac{1}{n}\sum_{i=1}^{n} w_{ij}^{(k)}, \quad \mu^{(k+1)} = \frac{n + \sqrt{n^2 + 4\left(\sum_{i=1}^{n}\sum_{j=1}^{2}\phi_j^{(k)} w_{ij}^{(k)} x_i^{-1}\right)\left(\sum_{i=1}^{n}\sum_{j=1}^{2}\phi_j^{(k)} w_{ij}^{(k)} x_i\right)}}{2\left(\sum_{i=1}^{n}\sum_{j=1}^{2}\phi_j^{(k)} w_{ij}^{(k)} x_i^{-1}\right)},$$

$$\phi_j^{(k+1)} = \frac{\mu^{(k)}\sum_{i=1}^{n} w_{ij}^{(k)}}{\sum_{i=1}^{n} w_{ij}^{(k)}(x_i - \mu^{(k)})^2 x_i^{-1}}, \quad j = 1, 2.$$

To find the estimator $(\hat{p}_1, \hat{\mu}, \hat{\phi}_1, \hat{\phi}_2)$ of $\Theta = (p_1, \mu, \phi_1, \phi_2)$, we choose $\Theta^{(0)} = (p_1^{(0)}, \mu^{(0)}, \phi_1^{(0)}, \phi_2^{(0)})$ as an initial value based on the sample and repeat the E-step and M-step simultaneously until it converges. After the convergence, last iteration values are considered as the estimates $(\hat{p}_1, \hat{\mu}, \hat{\phi}_1, \hat{\phi}_2)$. Using the relation $\lambda_j = \mu\phi_j$, $j = 1, 2$, we obtain the estimates $(\hat{\lambda}_1, \hat{\lambda}_2)$ of the parameters (λ_1, λ_2). In Sect. 5.2, we consider several parameter combinations for the MTIG distributions and computed their estimates. These estimates are used as plug-in estimate in finding estimated EPC.

5 Numerical results

5.1 Interval estimation of conditional ERC

In this section, a detailed simulation study has been performed to compute the CI of the conditional ERC into one of two independent IG distributions having an equal mean but different λ_is. We generate random samples with sizes n_1 and n_2 from populations Π_1 and Π_2, respectively. Using the methods mentioned in Sect. 4.1, we calculate the CIs for sample sizes $(n_1, n_2) = (5, 5), (10, 10), (20, 20)$. The AUL and ALL

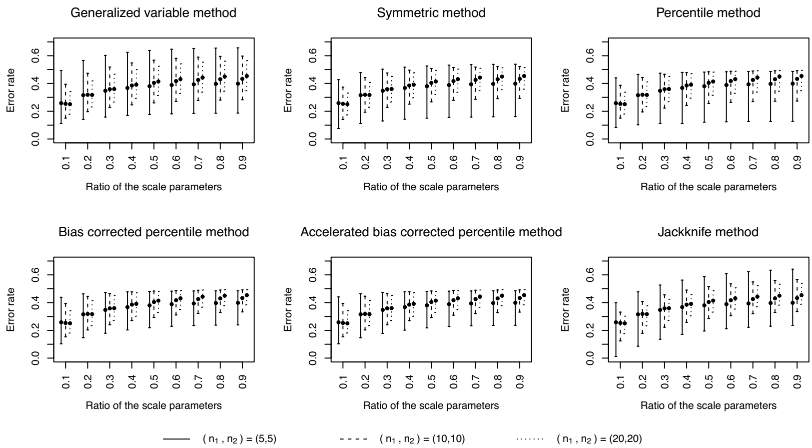

Fig. 1 Estimated conditional error rate and the CI for $(n_1, n_2) = (5, 5), (10, 10), (20, 20)$

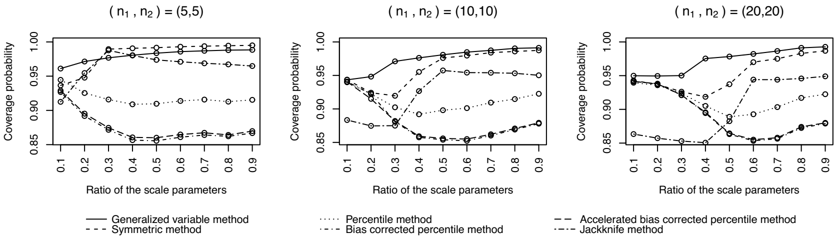

Fig. 2 Coverage probability of the interval of conditional error rate for $(n_1, n_2) = (5, 5), (10, 10), (20, 20)$

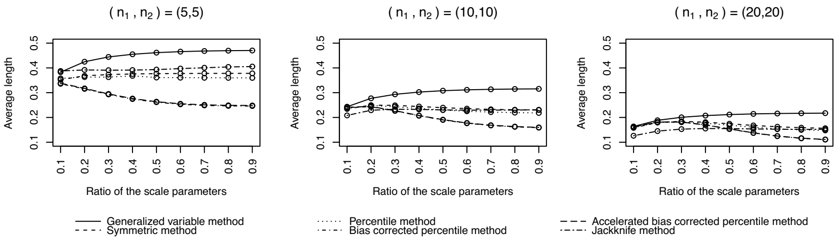

Fig. 3 Average length of the interval of conditional error rate for $(n_1, n_2) = (5, 5), (10, 10), (20, 20)$

values are computed using Monte Carlo simulations with 20,000 replications. The expected length of the CI is the difference between AUL and ALL. Since the ERC and the corresponding CI depend on the ratio λ_1/λ_2 but are independent of the mean parameter, we calculate the CIs by varying λ_1/λ_2 from 0.1 to 0.9. The CIs using different methods are compared in terms of average lengths and CPs.

From Figs. 1, 2 and 3, it follows that the generalized variable-based CI produces better CPs than other CIs. For $(n_1, n_2) = (5, 5)$, the CI using symmetric method marginally dominates the generalized interval in terms of CP when $\lambda_1/\lambda_2 > 0.3$. As the sample size increases, CI using the generalized variable approach outperforms other methods. However, the average length of the CI using the generalized variable approach is longer than other methods. For all the methods described, the length of the CI decreases as the sample size increases. The bias-corrected percentile and Abcp methods perform poorly in terms of CP. But the bias-corrected percentile and Abcp CIs become more efficient than other CIs as the ratio λ_1/λ_2 approaches one in terms of interval length.

Tables 1 and 2 refer to the estimated ERC for three populations. It also suggests that the symmetric and percentile methods perform better than other methods. We have analyzed the computational time required to derive CIs for error rates. Table 3 presents the execution time, measured in seconds, necessary for the bootstrap method, jackknife method, and generalized variable approach to compute CIs for error rates across various sample sizes. Each entry reflects the time for a single iteration. The bootstrap method requires more computational time compared to the other two methods. Conversely, the jackknife method consistently demonstrates the least computational time across all parameter combinations and sample sizes. Although the computational time for these methods is generally reasonable, the generalized variable approach balances efficiency and performance. It consumes slightly more time than the jackknife method but delivers comparable or even better CP than the bootstrap method.

Next, Algorithm 2 is used to compute 90% and 95% credible intervals for the error rate considering $N(\theta_1^{(j)}, 1)$ as the proposed density when sample size $(n_1, n_2) = (20, 20), (25, 25)$. Credible intervals are calculated for λ_1/λ_2 values ranging from 0.1 to 0.9. As the error rate remains unaffected by μ, μ is set as one during the simulation. Monte Carlo samples are generated using 20,000 replications and setting $M = 1000$.

Figs. 4 and 5 visualize the error rates and the corresponding credible interval bands. Note that the average length of credible intervals is the smallest among all intervals explored in this study. With increasing sample sizes, the average interval length decreases while the CP increases. As the ratio λ_1/λ_2 approaches one, both CPs and average lengths exhibit a consistent increment. It is evident that the CPs of the credible intervals are nearly equivalent to those obtained from other methods. However, the distinctive feature lies in shorter interval lengths compared to alternative approaches. It is worth noting that this method may perform poorly for small samples. When the sample size from each population is greater than twenty, the HPD credible interval is recommended in estimating the ERC.

5.2 Estimates of the parameters using EM algorithm

As discussed in Sect. 4.2, we use EM algorithm to find the MLEs of $\Theta = (p_1, \mu, \phi_1, \phi_2)$ for different values of the parameters $p_1 (= 0.1, 0.3)$, $\lambda_1 (= 0.3, 0.7)$, $\lambda_2 (= 1, 3)$. Since

Table 1 90% CIs for conditional error rate

$(\mu, \lambda_1, \lambda_2, \lambda_3)$	Method[a]	$(n_1, n_2, n_3) = (5, 5, 5)$				$(n_1, n_2, n_3) = (10, 10, 10)$				$(n_1, n_2, n_3) = (20, 20, 20)$			
		ALL	AUL	AL	CP	ALL	AUL	AL	CP	ALL	AUL	AL	CP
		$\gamma^* = 0.5788$				$\gamma^* = 0.6032$				$\gamma^* = 0.6036$			
(1,0.1,0.2,0.3)	Sym	0.3425	0.8089	0.4664	0.9586	0.4358	0.7812	0.3454	0.9508	0.4810	0.7378	0.2568	0.9551
	Per	0.3396	0.7981	0.4585	0.9198	0.4352	0.7767	0.3415	0.9508	0.4833	0.7382	0.2549	0.9551
	Bcp	0.3052	0.7030	0.3978	0.8465	0.4039	0.7100	0.3061	0.8976	0.4643	0.6919	0.2276	0.8969
	Abcp	0.3048	0.7033	0.3985	0.8465	0.4037	0.7101	0.3064	0.8976	0.4644	0.6921	0.2277	0.8969
	Jack	0.2673	0.8690	0.6017	0.9945	0.3591	0.7958	0.4367	0.9855	0.4132	0.7329	0.3197	0.9619
		$\gamma^* = 0.5604$				$\gamma^* = 0.5733$				$\gamma^* = 0.5876$			
(1,0.5,1,1.5)	Sym	0.3224	0.8031	0.4807	0.9803	0.3962	0.7337	0.3375	0.9779	0.4704	0.7058	0.2354	0.9528
	Per	0.3197	0.7961	0.4764	0.9580	0.3948	0.7295	0.3347	0.9574	0.4708	0.7041	0.2333	0.9528
	Bcp	0.2746	0.6980	0.4234	0.8887	0.3836	0.6922	0.3086	0.9107	0.4559	0.6753	0.2194	0.9182
	Abcp	0.2742	0.6980	0.4238	0.8887	0.3835	0.6923	0.3088	0.9107	0.4558	0.6754	0.2196	0.9182
	Jack	0.2397	0.8618	0.6221	0.9883	0.3571	0.7819	0.4248	0.9919	0.4366	0.7318	0.2952	0.9772
		$\gamma^* = 0.5821$				$\gamma^* = 0.5893$				$\gamma^* = 0.5960$			
(1,1,1.5,2.5)	Sym	0.3643	0.8306	0.4663	0.9624	0.4216	0.7542	0.3326	0.9713	0.4796	0.7044	0.2248	0.9635
	Per	0.3590	0.8206	0.4616	0.9624	0.4197	0.7495	0.3298	0.9713	0.4789	0.7013	0.2224	0.9635
	Bcp	0.2862	0.7034	0.4172	0.8579	0.3937	0.7018	0.3081	0.9551	0.4704	0.6793	0.2089	0.9086
	Abcp	0.2855	0.7036	0.4181	0.8579	0.3934	0.7018	0.3084	0.9551	0.4702	0.6795	0.2093	0.9086
	Jack	0.2653	0.8733	0.6080	0.9972	0.3906	0.8023	0.4117	0.9927	0.4696	0.7520	0.2824	0.9793

[a]Bcp: bias-corrected percentile method; Jack: Jackknife method; Sym: symmetric method; Per: percentile method; ALL: average lower limit; AUL: average upper limit; AL: average length; CP: coverage probability

Table 2 95% CIs for conditional error rate

$(\mu, \lambda_1, \lambda_2, \lambda_3)$	Method[a]	$(n_1, n_2, n_3) = (5, 5, 5)$				$(n_1, n_2, n_3) = (10, 10, 10)$				$(n_1, n_2, n_3) = (20, 20, 20)$			
		ALL	AUL	AL	CP	ALL	AUL	AL	CP	ALL	AUL	AL	CP
		$\gamma^* = 0.5788$				$\gamma^* = 0.6032$				$\gamma^* = 0.6036$			
(1,0.1,0.2,0.3)	Sym	0.2978	0.8536	0.5558	0.9836	0.4027	0.8143	0.4116	0.9882	0.4564	0.7624	0.3060	0.9775
	Per	0.2926	0.8348	0.5422	0.9836	0.3997	0.8061	0.4064	0.9899	0.4588	0.7627	0.3039	0.9775
	Bcp	0.2729	0.7314	0.4585	0.8465	0.3784	0.7317	0.3533	0.9073	0.4453	0.7101	0.2648	0.9353
	Abcp	0.2724	0.7318	0.4594	0.8465	0.3782	0.7318	0.3536	0.9073	0.4453	0.7102	0.2649	0.9353
	Jack	0.2018	0.9345	0.7327	0.9989	0.3146	0.8403	0.5257	0.9985	0.3816	0.7645	0.3829	0.9899
		$\gamma^* = 0.5604$				$\gamma^* = 0.5733$				$\gamma^* = 0.5876$			
(1,0.5,1,1.5)	Sym	0.2764	0.8491	0.5727	0.9803	0.3852	0.7910	0.4058	0.9866	0.4479	0.7284	0.2805	0.9834
	Per	0.2716	0.8358	0.5642	0.9803	0.3820	0.7861	0.4041	0.9866	0.4469	0.7273	0.2804	0.9834
	Bcp	0.2413	0.7311	0.4898	0.8989	0.3579	0.7207	0.3628	0.9563	0.4363	0.6923	0.2560	0.9566
	Abcp	0.2408	0.7313	0.4905	0.8989	0.3577	0.7207	0.3630	0.9563	0.4362	0.6924	0.2562	0.9566
	Jack	0.1720	0.9295	0.7575	0.9987	0.3138	0.8252	0.5114	0.9977	0.4074	0.7609	0.3535	0.9948
		$\gamma^* = 0.5821$				$\gamma^* = 0.5893$				$\gamma^* = 0.5960$			
(1,1,1.5,2.5)	Sym	0.3197	0.8752	0.5555	0.9911	0.3898	0.7860	0.3962	0.9905	0.4581	0.7259	0.2678	0.9843
	Per	0.3094	0.8561	0.5467	0.9842	0.3847	0.7803	0.3956	0.9905	0.4553	0.7227	0.2674	0.9843
	Bcp	0.2553	0.7353	0.4800	0.9581	0.3668	0.7253	0.3585	0.9551	0.4507	0.6943	0.2436	0.9533
	Abcp	0.2546	0.7354	0.4808	0.9581	0.3665	0.7254	0.3589	0.9595	0.4505	0.6943	0.2438	0.9533
	Jack	0.1991	0.9395	0.7404	0.9997	0.3487	0.8443	0.4956	0.9976	0.4417	0.7798	0.3381	0.9943

Table 3 Computational time for bootstrap method, jackknife method, and generalized variable approach

(n_1, n_2)	Method	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
(5,5)	Boot	0.043760	0.049385	0.056099	0.050489	0.054422	0.074274	0.054871	0.053766	0.057713
	Jack	0.000661	0.000624	0.000643	0.000635	0.000717	0.000691	0.000599	0.000644	0.000626
	GV	0.013182	0.012645	0.011910	0.011352	0.016278	0.013562	0.013929	0.012574	0.012084
(10,10)	Boot	0.055932	0.047457	0.051607	0.044530	0.043685	0.050766	0.050090	0.062374	0.061510
	Jack	0.001224	0.001229	0.001275	0.001161	0.001243	0.001144	0.001260	0.001810	0.001209
	GV	0.013357	0.013394	0.009824	0.010754	0.011317	0.010735	0.009347	0.010314	0.010421
(20,20)	Boot	0.052394	0.052922	0.047230	0.047875	0.055450	0.056864	0.064131	0.065101	0.044105
	Jack	0.003570	0.003832	0.003344	0.002442	0.002374	0.003400	0.003535	0.004144	0.004125
	GV	0.013076	0.011728	0.010884	0.010248	0.010104	0.008982	0.008944	0.011811	0.008847

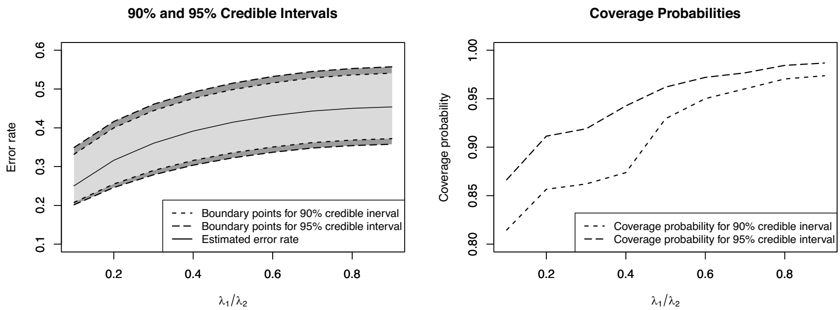

Fig. 4 Credible intervals of the conditional error rate with the CPs for $(n_1, n_2) = (20, 20)$

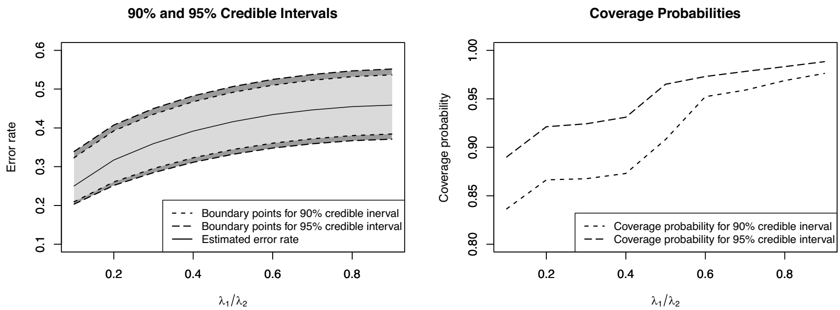

Fig. 5 Credible intervals of the conditional error rate with CPs for $(n_1, n_2) = (25, 25)$

the exact value of μ does not affect the estimates of other parameters, we assume μ as five throughout the simulation study. For each parameter combination, we consider sample of sizes $n = 40, 50, 60, 70, 80, 90, 100$ to find the estimates of $\Theta = (p_1, \mu, \phi_1, \phi_2)$. The algorithm is repeated until the computed difference between two successive iterations is $\leq 10^{-3}$. For every parameter combination, 5000 replications are used, and the average of the estimates are taken as the estimated value. Next, 100 new samples are generated from the MTIG distribution. Using the estimators as plug-in estimators in the classification function, we obtain the probability of correct classification. Repeating this procedure 100 times and taking the average, we finally get the EPC values. The following observations are made from Table 4.

As the sample size increases, EPC increases in every case. For the parameter combination, when p_1 is 0.1, the EPC values are higher than the EPC values obtained for $p_1 = 0.3$. The MSEs of the estimators for the parameters decrease as the sample size increases. The EPC value is higher if the difference between λ_1 and λ_2 is higher.

Table 4 Estimates of the parameters and EPC for mixture of IG distributions

n	$p = 0.1$	$\mu = 5$	$\lambda_1 = 0.3$	$\lambda_2 = 1.2$	EPC	$p = 0.3$	$\mu = 5$	$\lambda_1 = 0.3$	$\lambda_2 = 1.2$	EPC
40	0.4338 (0.1114)	5.1289 (0.0166)	0.8914 (0.3498)	1.9229 (0.5226)	0.7580	0.4235 (0.0153)	5.1933 (0.0374)	0.5672 (0.0714)	1.5969 (0.1575)	0.6917
50	0.4256 (0.1060)	5.0885 (0.0078)	0.8575 (0.3108)	1.6894 (0.2395)	0.7977	0.4162 (0.0135)	5.1267 (0.0161)	0.5438 (0.0594)	3.5004 (5.2920)	0.6367
60	0.4146 (0.0990)	5.0785 (0.0062)	0.8222 (0.2726)	1.4176 (0.0474)	0.8443	0.4133 (0.0128)	5.1109 (0.0123)	0.5274 (0.0517)	1.4275 (0.0518)	0.7037
70	0.4120 (0.0974)	5.0744 (0.0055)	0.8072 (0.2573)	1.8676 (0.4457)	0.7868	0.4069 (0.0114)	5.1089 (0.0119)	0.5112 (0.0446)	1.2837 (0.0070)	0.7104
80	0.4070 (0.0942)	5.0568 (0.0032)	0.7913 (0.2414)	1.3852 (0.0343)	0.8529	0.4030 (0.0106)	5.0931 (0.0087)	0.4997 (0.0399)	1.3365 (0.0186)	0.7101
90	0.4021 (0.0913)	5.0590 (0.0035)	0.7798 (0.2302)	1.3560 (0.0243)	0.8599	0.4008 (0.0102)	5.0894 (0.0080)	0.4914 (0.0366)	1.2820 (0.0067)	0.7122
100	0.3967 (0.0880)	5.0548 (0.0030)	0.7662 (0.2174)	1.2747 (0.0056)	0.8735	0.3988 (0.0098)	5.0735 (0.0054)	0.4866 (0.0348)	1.2752 (0.0057)	0.7128
	$p = 0.1$	$\mu = 5$	$\lambda_1 = 1$	$\lambda_2 = 1.2$	EPC	$p = 0.3$	$\mu = 5$	$\lambda_1 = 1$	$\lambda_2 = 1.2$	EPC
40	0.4904 (0.1524)	5.0619 (0.0038)	1.2390 (0.0571)	1.4952 (0.0871)	0.6958	0.4901 (0.0362)	5.0531 (0.0028)	1.1912 (0.0365)	1.6235 (0.1794)	0.5782
50	0.4920 (0.1536)	5.0302 (0.0009)	1.2220 (0.0493)	1.4021 (0.0408)	0.7099	0.4915 (0.0367)	5.0251 (0.0006)	1.1752 (0.0307)	1.3472 (0.0217)	0.5870
60	0.4932 (0.1546)	5.0248 (0.0006)	1.2131 (0.0454)	1.2975 (0.0095)	0.7575	0.4931 (0.0373)	5.0188 (0.0004)	1.1677 (0.0281)	1.2448 (0.0020)	0.5918
70	0.4942 (0.1554)	5.0267 (0.0007)	1.2075 (0.0431)	1.2828 (0.0069)	0.7541	0.4941 (0.0377)	5.0248 (0.0006)	1.1611 (0.0260)	1.2747 (0.0056)	0.5906
80	0.4954 (0.1563)	5.0148 (0.0002)	1.2042 (0.0417)	1.2696 (0.0048)	0.7473	0.4951 (0.0381)	5.0117 (0.0001)	1.1570 (0.0246)	1.2230 (0.0005)	0.5927
90	0.4951 (0.1561)	5.0158 (0.0003)	1.1999 (0.0400)	1.2444 (0.0020)	0.7877	0.4950 (0.0380)	5.0146 (0.0002)	1.1536 (0.0236)	1.2039 (0.0000)	0.5934
100	0.4965 (0.1572)	5.0119 (0.0001)	1.1991 (0.0396)	1.2477 (0.0023)	0.7492	0.4959 (0.0384)	5.0096 (0.0001)	1.1526 (0.0233)	1.1949 (0.0000)	0.5938
	$p = 0.1$	$\mu = 5$	$\lambda_1 = 0.3$	$\lambda_2 = 0.7$	EPC	$p = 0.3$	$\mu = 5$	$\lambda_1 = 0.3$	$\lambda_2 = 0.7$	EPC
40	0.4718 (0.1383)	5.1270 (0.0161)	0.6273 (0.1071)	2.2193 (2.3082)	0.5894	0.4641 (0.0269)	5.1535 (0.0236)	0.4926 (0.0371)	2.5032 (3.2515)	0.5798
50	0.4710 (0.1376)	5.1038 (0.0108)	0.6147 (0.0990)	3.3637 (7.0952)	0.5361	0.4629 (0.0265)	5.1172 (0.0137)	0.4839 (0.0338)	1.1128 (0.1704)	0.6310
60	0.4694 (0.1364)	5.0805 (0.0065)	0.6058 (0.0935)	0.9941 (0.0865)	0.7239	0.4595 (0.0254)	5.0899 (0.0081)	0.4741 (0.0303)	0.8884 (0.0355)	0.6437
70	0.4699 (0.1368)	5.0719 (0.0052)	0.6018 (0.0911)	0.8957 (0.0383)	0.7470	0.4597 (0.0255)	5.0793 (0.0063)	0.4700 (0.0289)	0.7796 (0.0063)	0.6526
80	0.4702 (0.1370)	5.0477 (0.0023)	0.5993 (0.0896)	0.7343 (0.0012)	0.8226	0.4581 (0.0250)	5.0619 (0.0038)	0.4644 (0.0270)	0.7145 (0.0002)	0.6563
90	0.4685 (0.1358)	5.0530 (0.0028)	0.5958 (0.0875)	0.7188 (0.0004)	0.8336	0.4565 (0.0245)	5.0536 (0.0029)	0.4618 (0.0262)	0.7193 (0.0004)	0.6560
100	0.4701 (0.1370)	5.0430 (0.0019)	0.5960 (0.0876)	0.7473 (0.0022)	0.8144	0.4575 (0.0248)	5.0450 (0.0020)	0.4615 (0.0261)	0.6650 (0.0012)	0.6599

Table 4 (continued)

	$p = 0.1$	$\mu = 5$	$\lambda_1 = 1.5$	$\lambda_2 = 2$	EPC	$p = 0.3$	$\mu = 5$	$\lambda_1 = 1.5$	$\lambda_2 = 2$	EPC
40	0.4899 (0.1520)	5.0394 (0.0016)	2.0416 (0.2933)	2.3944 (0.1555)	0.7154	0.4905 (0.0363)	5.0379 (0.0014)	1.9197 (0.1762)	2.5468 (0.2990)	0.5896
50	0.4922 (0.1538)	5.0187 (0.0003)	2.0147 (0.2649)	2.3798 (0.1443)	0.7004	0.4914 (0.0366)	5.0161 (0.0003)	1.8912 (0.1530)	2.3525 (0.1242)	0.5934
60	0.4932 (0.1546)	5.0199 (0.0004)	2.0002 (0.2502)	2.1480 (0.0219)	0.7552	0.4928 (0.0372)	5.0172 (0.0003)	1.8784 (0.1432)	2.0029 (0.0000)	0.6018
70	0.4939 (0.1551)	5.0212 (0.0004)	1.9884 (0.2385)	2.1311 (0.0172)	0.7508	0.4936 (0.0375)	5.0219 (0.0005)	1.8670 (0.1347)	2.0144 (0.0002)	0.6020
80	0.4947 (0.1558)	5.0133 (0.0002)	1.9817 (0.2321)	2.0996 (0.0099)	0.7559	0.4939 (0.0376)	5.0135 (0.0002)	1.8578 (0.1281)	1.9731 (0.0007)	0.6033
90	0.4943 (0.1555)	5.0139 (0.0002)	1.9754 (0.2260)	2.0463 (0.0021)	0.8084	0.4941 (0.0377)	5.0155 (0.0002)	1.8536 (0.1250)	1.9380 (0.0038)	0.6033
100	0.4959 (0.1567)	5.0097 (0.0001)	1.9745 (0.2251)	2.0427 (0.0018)	0.7824	0.4952 (0.0381)	5.0112 (0.0001)	1.8519 (0.1238)	1.9211 (0.0062)	0.6042
	$p = 0.1$	$\mu = 5$	$\lambda_1 = 1$	$\lambda_2 = 3$	EPC	$p = 0.3$	$\mu = 5$	$\lambda_1 = 1$	$\lambda_2 = 3$	EPC
40	0.4614 (0.1306)	5.0566 (0.0032)	2.5363 (2.3601)	3.6750 (0.4557)	0.7844	0.4543 (0.0238)	5.0527 (0.0028)	1.8529 (0.7274)	4.1701 (1.3692)	0.6589
50	0.4600 (0.1296)	5.0355 (0.0013)	2.4768 (2.1811)	3.6343 (0.4023)	0.7886	0.4510 (0.0228)	5.0305 (0.0009)	1.8045 (0.6472)	2.8273 (0.0298)	0.6834
60	0.4542 (0.1255)	5.0320 (0.0010)	2.4158 (2.0044)	3.2113 (0.0447)	0.8371	0.4491 (0.0222)	5.0278 (0.0008)	1.7686 (0.5907)	3.0172 (0.0003)	0.6811
70	0.4542 (0.1255)	5.0312 (0.0010)	2.3929 (1.9402)	3.2394 (0.0573)	0.8322	0.4481 (0.0219)	5.0351 (0.0012)	1.7423 (0.5511)	2.9878 (0.0001)	0.6826
80	0.4515 (0.1236)	5.0278 (0.0008)	2.3671 (1.8690)	3.3692 (0.1363)	0.8232	0.4449 (0.0210)	5.0331 (0.0011)	1.7174 (0.5146)	2.6843 (0.0997)	0.6899
90	0.4514 (0.1235)	5.0253 (0.0006)	2.3565 (1.8401)	2.9633 (0.0014)	0.8643	0.4454 (0.0211)	5.0304 (0.0009)	1.7074 (0.5004)	2.8542 (0.0212)	0.6867
100	0.4488 (0.1216)	5.0208 (0.0004)	2.3375 (1.7889)	3.2506 (0.0628)	0.8355	0.4455 (0.0212)	5.0243 (0.0006)	1.7050 (0.4971)	2.5436 (0.2083)	0.6918

5.3 Illustrative examples

We consider the following examples as applications of the considered model. We use *goft* package (González-Estrada and Villaseñor 2017) in R programming language to fit the IG distribution to the data. To check the equality of the means, we use the algorithm proposed by Shi and Lv (2012). The significance of the IG distribution can be illustrated by its application to real-world data with positive skewness. While the Gamma distribution is commonly used for many data appearing in real-world situations, the IG distribution is an alternative for fitting such data. For example, Watson and Smith (1985) studied the breaking strengths of single carbon fibers with varying lengths. Consider the data set pertaining to the breaking strength of 10 mm long single carbon fibers. The R^2 value is 98.68% when the data is fitted with the Gamma distribution, and the R^2 value is 98.75% when using the IG distribution. This demonstrates that both the gamma and IG distribution can be used to fit such a data set. (see Fig. 6).

Example 1 Feigl and Zelen (1965) observed the white blood cell (WBC) counts for 33 patients. These patients were divided into two groups, AG positive and AG negative. They were formed based on whether Auer rods and/or significant granulature of leukemic cells were present in the bone marrow at the time of diagnosis. They find that the survival probability depends on the WBC count for AG positive group, whereas it does not depend on the WBC count for AG negative group. The WBC data for both groups are presented below after dividing by 1000.

AG positive: 2.3,.75, 4.3, 2.6, 6, 10.5, 10, 17, 5.4, 7, 9.4, 32, 35, 100, 100, 52, 100; AG negative: 4.4, 3, 4, 1.5, 9, 5.3, 10, 19, 27, 28, 31, 26, 21, 79, 100, 100.

For testing the IG distribution fit to the data, the p values for the groups are 0.6058 and 0.5962, respectively. This implies that the null hypothesis is not rejected based on the data and the data for each patient group follows IG distribution. The p-value to check the equality of mean parameters for the two groups is 0.995. Thus, we do not reject the null hypothesis and conclude that the group means are equal. Fig. 7 represents the density and CDF plots of the empirical distribution and IG distribution. Within the IG framework, the scale-like parameter (λ) reflects the variability of WBC counts. A higher λ value indicates a wider range of WBC counts within a group. This ratio of λ_is between two groups indicates the relative variability in

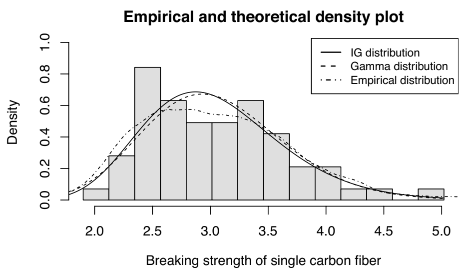

Fig. 6 Density plot for breaking strength of single carbon fiber

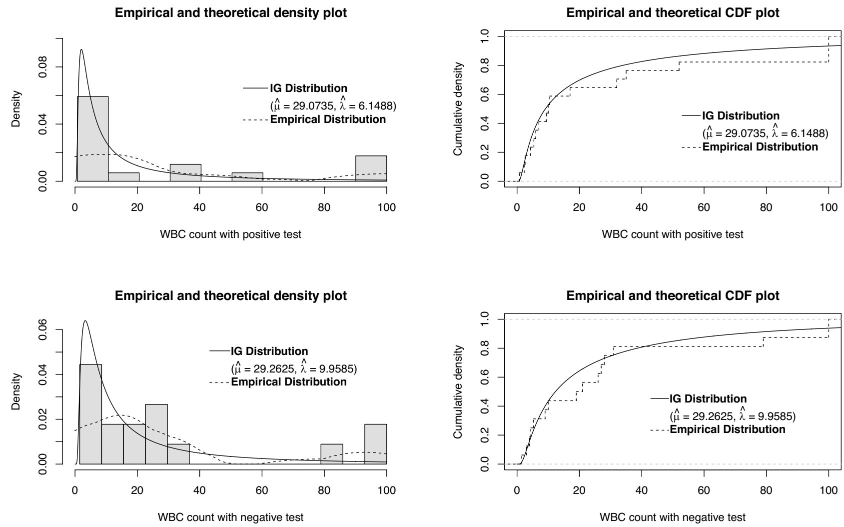

Fig. 7 Density and CDF plots for the data sets considered in Example 1

Table 5 95% CIs of conditional error rate for WBC counts data

Method		Method		Method[a]	
Sym	(0.3195, 0.5271)	Per	(0.3043, 0.4950)	Bcp	(0.3274, 0.4974)
Abcp	(0.3271, 0.4974)	Jack	(0.3291, 0.5472)	GV	(0.3203, 0.5677)

[a]GV: Generalized variable based method

WBC counts between the groups. In Table 5, we have presented CIs of conditional ERC into either of two groups using the proposed methods.

Example 2 Shapiro et al. (1987) recorded numbers of T_4 cells (per mm^3) in the blood of 40 patients, where 20 people were affected with Hodgkin's disease, and the rest were not diagnosed with the disease. The number of T_4 cells for both the groups are available in Krishnamoorthy and Tian (2008). Chhikara and Folks (1989) showed that the data sets for both groups follow the IG distribution. We consider the data sets of two groups for studying the two-class classification problem where each class density is IG. First, the model assumptions need to be checked. To fit the IG distribution for both groups, the *p*-values are 0.6494 and 0.4099, respectively. The *p*-value to test the equality of means of both the groups is 0.058. Thus both groups follow IG distributions with an equal mean. The sample mean for the groups are 0.8232 and 0.5221, respectively and $(s_1, s_2) = (0.7105, 0.8663)$. Figure 8 represents the density and CDF plots of the empirical distribution and IG distribution for the

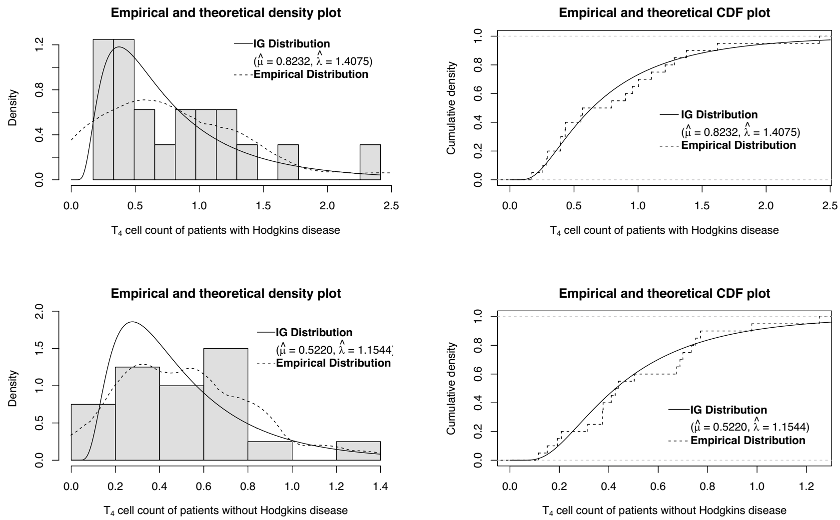

Fig. 8 Density and CDF plots for the data sets considered in Example 2

Table 6 95% CIs for Hodgkins disease data

Method		Method		Method	
Sym	(0.3807, 0.5226)	Per	(0.3672, 0.4985)	Bcp	(0.4169, 0.4997)
Abcp	(0.4163, 0.4997)	Jack	(0.4097, 0.5639)	GV	(0.3679, 0.5858)

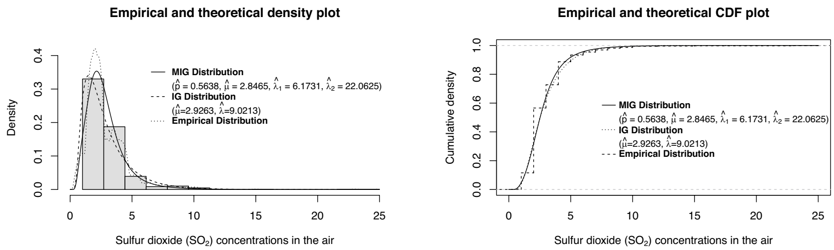

Fig. 9 Density and CDF plots for the data sets considered in Example 3

datasets. In Table 6, we compute the CIs of the conditional ERC using the proposed methods. The 90% and 95% credible intervals for the error rate are (0.3942, 0.5639) and (0.3804, 0.5775), respectively.

Example 3 Balakrishnan et al. (2009) discussed several aspects of MIG distribution for fitting positively skewed data. They analyzed different data sets from actuarial science, engineering, and toxicology. In a monitoring station in Santiago, hourly SO_2 concentrations (in ppm) are recorded as a part of an environmental air pollution data set. The frequency of each value is mentioned in the respective parentheses. The mean, median, and mode of the data set are 2.9261, 2, and 2, respectively. The figure and relation among mean, median, and mode suggest that the data follows a positively skewed distribution. The R^2 value for fitting an IG distribution to the data is 88.67%, whereas R^2 value for fitting MIG distribution with the common mean is 89.37%. The *ig_test* function from *goft* package in R indicates that an IG distribution is not a good fit since the p value is nearly zero. The estimates of the parameters for IG distribution are $\hat{\mu} = 2.9261, \hat{\lambda} = 9.0213$. For MIG distribution with an equal mean, estimates of the parameters using EM algorithm are $\hat{p} = 0.5638, \hat{\mu} = 2.8465, \hat{\lambda}_1 = 6.1731$ and $\hat{\lambda}_2 = 22.0625$. Figure 9 shows that a mixture of IG distributions having an equal mean provides a better fit than an IG distribution for the dataset.

6 Conclusions

We have studied the estimation of the function of parameters for two IG populations having an equal mean. We have derived CIs of the conditional ERC using the bootstrap, jackknife, and generalized variable approaches. The CI based on the generalized variable estimator performs better than other intervals in terms of CP. A noninformative probability matching prior is used to obtain HPD credible intervals for the conditional error rate. Opting for credible intervals is advised to estimate the error rate, as these intervals tend to have shorter lengths than other CIs with the same CP. Using the EM algorithm, we have derived estimators of the parameters for a mixture of IG distributions. The estimators are used to find the EPCs. For illustration purposes, two datasets are used to find the CIs of the conditional ERC. The third dataset is an example where a mixture of IG distributions with an equal mean fit better than a single IG distribution. Based on mixtures of Gaussian distributions, model-based classification methods are useful for various practical problems. As an extension, multivariate normal IG distribution can be used for model-based classification. A copy of R code will be shared with interested researchers upon request.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00180-024-01554-6.

Acknowledgements We thank the editor, associate editor, and reviewers for their constructive suggestions, which considerably enhanced the manuscript. We are thankful to the editor for suggesting to incorporate credible intervals. The second author gratefully acknowledges the financial support under MATRICS project (No. MTR/2022/000535) received from SERB, Department of Science and Technology, India.

References

Ahmad M, Chaubey Y, Sinha B (1991) Estimation of a common mean of several univariate inverse Gaussian populations. Ann Inst Stat Math 43(2):357–367

Amoh R (1985) Estimation of a discriminant function from a mixture of two inverse Gaussian distributions when sample size is small. J Stat Comput Simul 20(4):275–286

Anderson T (2003) An introduction to multivariate statistical analysis, 3rd edn. Wiley, New York

Balakrishnan N, Leiva V, Sanhueza A, Cabrera E (2009) Mixture inverse Gaussian distributions and its transformations, moments and applications. Statistics 43(1):91–104

Batsidis A, Zografos K (2006) Discrimination of observations into one of two elliptic populations based on monotone training samples. Metrika 64(2):221–241

Berger JO, Bernardo JM (1989) Estimating a product of means: Bayesian analysis with reference priors. J Am Stat Assoc 84(405):200–207

Bernardo JM (1979) Reference posterior distributions for Bayesian inference. J R Stat Soc Ser B Stat Methodol 41(2):113–128

Chen M-H, Shao Q-M (1999) Monte Carlo estimation of Bayesian credible and HPD intervals. J Comput Graph Stat 8(1):69–92

Chhikara R, Folks J (1989) The inverse Gaussian distribution: theory, methodology and application. Marcel Dekker, New York

Chung H-C, Han C-P (2009) Conditional confidence intervals for classification error rate. Comput Stat Data Anal 53(12):4358–4369

Conde D, Fernández M, Salvador B (2005) A classification rule for ordered exponential populations. J Stat Plann Inference 135(2):339–356

Cox DR, Reid N (1987) Parameter orthogonality and approximate conditional inference. J R Stat Soc Ser B Stat Methodol 49(1):1–18

Datta GS, Ghosh M (1995) Some remarks on noninformative priors. J Am Stat Assoc 90(432):1357–1363

Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. J R Stat Soc Ser B Stat Methodol 39(1):1–22

Feigl P, Zelen M (1965) Estimation of exponential survival probabilities with concomitant information. Biometrics 21(4):826

González-Estrada E, Villaseñor JA (2017) An R package for testing goodness of fit: goft. J Stat Comput Simul 88(4):726–751

Gupta RC, Akman HO (1995) Bayes estimation in a mixture inverse Gaussian model. Ann Inst Stat Math 47(3):493–503

Jana N, Chakraborty A (2023) Estimating error rate of classification into several normal populations under equal mean restriction. Commun Stat Simul Comput. https://doi.org/10.1080/03610918.2023.2240549

Jana N, Kumar S (2019) Ordered classification rules for inverse Gaussian populations with unknown parameters. J Stat Comput Simul 89(14):2597–2620

Karlis D (2002) An EM type algorithm for maximum likelihood estimation of the normal–inverse Gaussian distribution. Stat Probab Lett 57(1):43–52

Kim DH, Kang SG, Lee WD (2006) Noninformative priors for linear combinations of the normal means. Stat Pap 47(2):249–262

Krishnamoorthy K, Tian L (2008) Inferences on the difference and ratio of the means of two inverse Gaussian distributions. J Stat Plann Inference 138(7):2082–2089

Lehmann E, Casella G (2006) Theory of point estimation. Springer, Berlin

Mukerjee R, Reid N (1999) On a property of probability matching priors: matching the alternative coverage probabilities. Biometrika 86(2):333–340

Shapiro CM, Beckmann E, Christiansen N, Bitran JD, Kozloff M, Billings AA, Telfer MC (1987) Immunologic status of patients in remission from Hodgkin's disease and disseminated malignancies. Am J Med Sci 293(6):366–370

Shi J-H, Lv J-L (2012) A new generalized -value for testing equality of inverse Gaussian means under heterogeneity. Stat Probab Lett 82(1):96–102

Sindhu TN, Khan HM, Hussain Z, Al-Zahrani B (2018) Bayesian inference from the mixture of half-normal distributions under censoring. J Natl Sci Found 46(4):587–600

Tian L, Wilding GE (2005) Confidence intervals of the ratio of means of two independent inverse Gaussian distributions. J Stat Plann Inference 133(2):381–386
Tibshirani R (1989) Noninformative priors for one parameter of many. Biometrika 76(3):604–608
Watson AS, Smith RL (1985) An examination of statistical theories for fibrous materials in the light of experimental data. J Mater Sci 2(9):3260–3270
Ye R-D, Ma T-F, Wang S-G (2010) Inferences on the common mean of several inverse Gaussian populations. Comput Stat Data Anal 54(4):906–915