Applied Mathematics and Computation 217 (2011) 5480–5490

Contents lists available at ScienceDirect

Applied Mathematics and Computation

journal homepage: www.elsevier.com/locate/amc

On the common mean of several inverse Gaussian distributions based on a higher order likelihood method

S.H. Lin [a,*], I.M. Wu [b]

[a] Department of Insurance and Finance, National Taichung Institute of Technology, 129, Sec. 3, Sanmin road, Taichung 404, Taiwan
[b] Institute of Statistics, National ChiaoTung University, Hsinchu, Taiwan

ARTICLE INFO

Keywords:
Coverage probability
Expected length
Inverse Gaussian
Signed log-likelihood ratio statistic
Type I errors

ABSTRACT

An interval estimation method for the common mean of several heterogeneous inverse Gaussian (IG) populations is discussed. The proposed method is based on a higher order likelihood-based procedure. The merits of the proposed method are numerically compared with the signed log-likelihood ratio statistic, two generalized pivot quantities and the simple t-test method with respect to their expected lengths, coverage probabilities and type I errors. Numerical studies show that the coverage probabilities of the proposed method are very accurate and type I errors are close to the nominal level.05 even for very small samples. The methods are also illustrated with two examples.

1. Introduction

In many application areas, such as demography, management science, hydrology, finance, etc., data are frequently positive and right-skewed. In the past four decades, the inverse Gaussian (IG) distribution has drawn much attention and the inferences concerned with the IG distribution have also grown rapidly because IG is an ideal candidate for modeling and analyzing the right-skewed and positive data. For instance, Wise [1,2] and Wise et al. [3] developed the IG population as a possible model to describe cycle time distribution for particles in the blood and Lancaster [4] made use of the IG distribution in describing strike duration data. Furthermore, IG distribution can also serve as a convenient prior for scale in Bayesian approaches to estimation with assumed Gaussian data [5]. The IG distribution can not only accommodate a variety of shapes, from highly skewed to almost normal, but also shares many elegant and convenient properties with Gaussian models; e.g., the associated inference methods are based on the well-known t, χ^2, and F distributions as for the normal case. See Chhikara and Folks [6], Seshadri [7,8] and Mudholkar and Tian [9] more details of Gaussian and IG analogies.

The probability density function (pdf) of IG distribution, IG (μ,λ), is defined as

$$f(x;\mu,\lambda)=\left(\frac{\lambda}{2\pi x^3}\right)^{1/2}\exp\left\{-\frac{\lambda}{2\mu^2 x}(x-\mu)^2\right\},\quad x>0,\ \mu>0,\ \lambda>0, \tag{1.1}$$

where μ is the mean parameter and λ is the scale parameter. The inference methods of the IG model are closely analogous to those of the Gaussian model; for example, a very common problem in applied fields is to compare the means of several Gaussian populations, i.e.

$$H_0:\mu_1=\mu_2=\cdots=\mu_I \quad \text{vs.} \quad H_1:\text{not all } \mu_i\text{'s are equal.}$$

* Corresponding author. Address: 129 Sanmin Road Sec. 3, Taichung 404, Taiwan.
E-mail address: suelin@ntit.edu.tw (S.H. Lin).

0096-3003/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.amc.2010.12.019

If the variance of each population is homogeneous, the analysis of variance (ANOVA) can be used to perform the test. Similarly, the analysis of reciprocals (ANORE) can be used to test the equality of means of several IG samples if all scale parameters among groups are assumed to be equal [6]. When the scale parameters are non-homogeneous, the ANORE fails to solve the problem. Tian [10] proposed a method to test the equality of IG means under heterogeneity, based on a generalized test variable. However, when the null hypothesis is not rejected, the inferences for the common mean remain unsolved. Recently, Ye et al. [11] proposed a mixture method for the common mean problem based on generalized inference and the large sample theory. However, as the author has mentioned, if the sample sizes n_i are not large and/or the scale parameter λ_i is not large compared to μ_i, the approximate distributions don't fit well. Therefore, an alternate method which can be applied to general cases deserves further research.

In this paper, we will estimate and construct the $100(1-\alpha)\%$ confidence interval for the common mean of several non-homogeneous IG populations based on a higher order likelihood-based method. This method, in theory, has a higher order accuracy, $O(n^{-3/2})$, even when the sample size is small. Reid [12] provided a review and annotated the development of his method. The method has been applied to solve many practical problems involving interval estimation for a skewed distribution, e.g., Wu et al. [13] presented a confidence interval for a log-normal mean based on this method; Wu and Wong [14] used the method to improve the interval estimation for the two-parameter Birnbaum-Saunders distribution; and Tian and Wilding [15] used the method to construct confidence interval for the ratio of means of two independent IG distributions. In our case, the likelihood-based method also gives a satisfactory result for the problem of interval estimation for the common mean of several IG distributions.

The remainder of this article is organized as follows. In Section 2, we will briefly introduce the properties of IG distribution and the concepts of the signed log-likelihood ratio statistic and a higher order asymptotic method. The method is then applied to construct a confidence interval for the common mean of several independent IG populations in Section 3. The generalized inference approach and the classical procedure under the assumption of identical scale are also described in Section 3. We present several simulation studies and two numerical examples in Section 4 to illustrate the merits of our proposed method. Some concluding remarks are given in Section 5.

2. A general review

2.1. Some properties of IG distribution

For a random sample of n observations $x_1, x_2, \ldots, x_n$ from IG (μ,λ), the uniformly minimum variance unbiased estimators (UMVUEs) of μ and λ^{-1} are $\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i$ and $W=\frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{1}{xi}-\frac{1}{\bar{x}}\right)$, respectively, and a minimum sufficient statistic of (μ,λ) is $\left(\sum_{i=1}^{n}x_i,\sum_{i=1}^{n}\frac{1}{x_i}\right)$. It is easy to verify that

$$\bar{x}\sim \mathrm{IG}(\mu,n\lambda)\quad\text{and}\quad (n-1)W\sim\frac{1}{\lambda}\chi^2_{n-1}, \tag{2.1}$$

and that these two statistics are independently distributed.

Remark 1. Let $x\sim$ IG (μ,λ) and $\Lambda\sim\frac{1}{\lambda}\chi^2_n$ be two independent random variables, then $\frac{\lambda(x-\mu)^2}{\mu^2 x}\sim\chi^2_1$ and its distribution is independent of $\lambda\Lambda\sim\chi^2_n$. Let $\mathrm{M}=\frac{\sqrt{n}(x-\mu)}{\mu(x\Lambda)^{1/2}}$, then the distribution of $|\mathrm{M}|$ is the truncated Student's t variable with n degrees of freedom and M^2 has the F distribution with 1 and n degrees of freedom [6].

From (2.1) and Remark 1, we know that $\frac{n\lambda(\bar{x}-\mu)^2}{\mu^2\bar{x}}\sim\chi^2_1$ which is independent of $(n-1)W\sim\frac{1}{\lambda}\chi^2_{n-1}$. Let $U=\frac{\sqrt{n}(\bar{x}-\mu)}{\mu(\bar{x}W)^{1/2}}$, then the distribution of $|U|$ is the truncated Student's t with $n-1$ degrees of freedom and $U^2\sim F_{1,n-1}$.

2.2. The likelihood-based inference

Let $x=(x_1,x_2,\ldots,x_n)$ be an independent sample from some distribution and $l(\theta)=l(\theta;x)$ be the log-likelihood function based on the sample data. Suppose θ is the p-dimensional vector of parameters that can be partitioned into (μ,λ) with μ being the parameter of interest with dimension 1, and λ being the nuisance parameters with dimensions $p-1$. The signed log-likelihood ratio statistic $r(\mu)$ for inference on μ is defined as

$$r(\mu)=sgn(\hat{\mu}-\mu)\left\{2\left[l(\hat{\theta})-l(\hat{\theta}_\mu)\right]\right\}^{1/2}, \tag{2.2}$$

where $\hat{\theta}=(\hat{\mu},\hat{\lambda})$ is the overall maximum likelihood estimator (MLE) of θ and $\hat{\theta}_\mu=(\mu,\hat{\lambda}_\mu)$ is the constrained MLE of θ for a given μ. Cox and Hinkley [16] verified that $r(\mu)$ is asymptotically distributed as the standard normal distribution with first-order accuracy $O(n^{-1/2})$. A $100(1-\alpha)\%$ confidence interval for μ based on $r(\mu)$ can be obtained by

$$\{\mu:|r(\mu)|\leqslant z_{\alpha/2}\}, \tag{2.3}$$

where $z_{\alpha/2}$ is the $100(1-\alpha/2)$th percentile of a standard normal distribution. Since the log-likelihood ratio statistic is quite inaccurate when the sample size is small, Barndorff–Nielsen [17,18] proposed a higher order likelihood-based method which is known as the modified signed log-likelihood ratio,

$$r^*(\mu) = r(\mu) + r(\mu)^{-1}\log\left\{\frac{q(\mu)}{r(\mu)}\right\}, \tag{2.4}$$

where $r(\mu)$ is the sign log-likelihood ratio statistic and $q(\mu)$ is a statistic which can be expressed in various forms depending on the information available. A widely applicable formula for $q(\mu)$ that ensures the $O(n^{-3/2})$ accuracy provided by Fraser et al. [19] is defined as

$$q(\mu) = \frac{\left|l_{;V}(\hat\theta) - l_{;V}(\hat\theta_\mu) l_{\lambda;V}(\hat\theta_\mu)\right|}{\left|l_{\theta;V}(\hat\theta)\right|}\left\{\frac{\left|j_{\theta\theta}(\hat\theta)\right|}{\left|j_{\lambda\lambda}(\hat\theta_\mu)\right|}\right\}^{1/2}, \tag{2.5}$$

where $j_{\theta\theta}(\hat\theta)$ is the $p\times p$ observed information matrix and $j_{\lambda\lambda}(\hat\theta_\mu)$ is the $(p-1)\times(p-1)$ observed nuisance information matrix. The vector array $V = (v_1', \ldots, v_p')$ in (2.5) is obtained from a vector pivot quantity $R(x;\theta) = (R_1(x;\theta), \ldots, R_n(x;\theta))$ by

$$V = -\left(\frac{\partial R(x;\theta)}{\partial x}\right)^{-1}\left(\frac{\partial R(x;\theta)}{\partial \theta}\right)\Bigg|_{\hat\theta}, \tag{2.6}$$

where the distribution of $R_i(x;\theta)$ is free of the nuisance parameters λ. The quantity $l_{;V}(\theta)$ is the likelihood gradient with

$$l_{;V}(\theta) = \left\{\frac{d}{dv_1}l(\theta;x), \ldots, \frac{d}{dv_p}l(\theta;x)\right\} = \left\{\sum_{j=1}^{n} l_{;x_j}(\theta)v_{1,j}, \ldots, \sum_{j=1}^{n} l_{;x_j}(\theta)v_{p,j}\right\}, \tag{2.7}$$

where $\frac{d}{dv_i}l(\theta,x)$ is the directional derivative of the log-likelihood function along $v_i = \{v_{i,1}, \ldots, v_{i,n}\}$, $i = 1, \ldots, p$, and $l_{;x_j}(\theta) = \frac{\partial l(\theta)}{\partial x_j}$, $j = 1, \ldots, n$. Moreover,

$$l_{;V}(\hat\theta_\mu) = \frac{\partial l(\theta)}{\partial V}\bigg|_{\hat\theta_\mu}, \quad l_{\lambda;V}(\hat\theta_\mu) = \frac{\partial l_{;V}(\theta)}{\partial \lambda}\bigg|_{\hat\theta_\mu} \quad \text{and} \quad l_{\theta;V}(\hat\theta) = \frac{\partial l_{;V}(\theta)}{\partial \theta}\bigg|_{\hat\theta}.$$

Note that $r^*(\mu)$ achieves third-order accuracy to a standard normal distribution [19]. Hence a $100(1-\alpha)\%$ confidence interval for μ based on $r^*(\mu)$ is

$$\{\mu : |r^*(\mu)| \leqslant z_{\alpha/2}\}. \tag{2.8}$$

3. Inferences for the common mean of several independent IG populations

3.1. The likelihood-based confidence interval in the general case

Suppose $x_i = (x_{i1}, x_{i2}, \ldots, x_{in_i})$, $i = 1, 2, \ldots, I$, are I random samples from IG (μ, λ_i) populations. The parameters, $\theta = (\mu, \lambda_1, \ldots, \lambda_I)$, contain μ being the parameter of interest and $(\lambda_1, \ldots, \lambda_I)$ being the nuisance parameters. The log-likelihood function is

$$l(\theta; x_1, x_2, \ldots, x_I) = \frac{1}{2}\sum_{i=1}^{I} n_i \log\frac{\lambda_i}{2\pi} - \frac{3}{2}\sum_{i=1}^{I}\sum_{j=1}^{n_i}\log x_{ij} - \frac{1}{2\mu^2}\sum_{i=1}^{I}\sum_{j=1}^{n_i}\lambda_i x_{ij} + \frac{1}{\mu}\sum_{i=1}^{I} n_i\lambda_i - \frac{1}{2}\sum_{i=1}^{I}\sum_{j=1}^{n_i}\frac{\lambda_i}{x_{ij}}. \tag{3.1}$$

Differentiating the log-likelihood function (3.1) with respect to θ for the first order yields the following results:

$$\begin{aligned}\frac{\partial l(\theta)}{\partial \mu} &= \frac{-1}{\mu^2}\sum_{i=1}^{I} n_i\lambda_i + \frac{1}{\mu^3}\sum_{i=1}^{I}\sum_{j=1}^{n_i}\lambda_i x_{ij}\\ \frac{\partial l(\theta)}{\partial \lambda_i} &= \frac{n_i}{2\lambda_i} + \frac{n_i}{\mu} - \frac{1}{2}\sum_{j=1}^{n_i}\frac{1}{x_{ij}} - \frac{1}{2\mu^2}\sum_{j=1}^{n_i}x_{ij}, \quad i = 1, \ldots, I.\end{aligned} \tag{3.2}$$

The overall MLEs $\hat\theta = (\hat\mu, \hat\lambda_1, \ldots, \hat\lambda_I)$ can be uniquely obtained by solving the non-linear system (3.2) simultaneously. Furthermore, the constrained MLEs $\hat\theta_\mu = (\mu, \hat\lambda_{\mu,1}, \ldots, \hat\lambda_{\mu,I})$ for a given μ are

$$\hat\lambda_{\mu,i} = \frac{-n_i\mu^2}{(2n_i\mu - \sum_{j=1}^{n_i}x_{ij} - \mu^2\sum_{j=1}^{n_i}\frac{1}{x_{ij}})}, \quad i = 1, \ldots, I. \tag{3.3}$$

Choosing a vector pivot quantity $R = \{R_{11}, \ldots, R_{In_I}\}$ with $R_{ij} = \frac{\lambda_i(x_{ij}-\mu)^2}{\mu^2 x_{ij}}$, $i = 1, \ldots, I$; $j = 1, \ldots, n_I$, then $R_{ij} \sim \chi_1^2$ with the distribution free of any unknown parameters. Differentiating R_{ij} with respect to x and θ, we have

$$\frac{\partial R_{ij}}{\partial x_k} = \left(\frac{\mu^2 x_{ij}^2}{\lambda_i(x_{ij}^2-\mu^2)}\right)^{-1}, \quad \text{if } j=k; \quad \text{else } \frac{\partial R_{ij}}{\partial x_k}=0;$$
$$\frac{\partial R_{ij}}{\partial \mu} = \frac{-2\lambda_i(x_{ij}-\mu)}{\mu^3};$$
$$\frac{\partial R_{ij}}{\partial \lambda_k} = \frac{(x_{ij}-\mu)^2}{\mu^2 x_{ij}}, \quad \text{if } j=k; \quad \text{else } \frac{\partial R_{ij}}{\partial \lambda_k}=0. \tag{3.4}$$

Furthermore, $V=(v'_1,\ldots,v'_{I+1}) = -\left(\frac{\partial R}{\partial x}\right)^{-1}\left(\frac{\partial R}{\partial \theta}\right)\Big|_{\hat\theta}$ with

$$v_1 = \left(\frac{2x_{11}^2}{\hat\mu(x_{11}+\hat\mu)},\ldots,\frac{2x_{1n_1}^2}{\hat\mu(x_{1n_1}+\hat\mu)},\ldots,\frac{2x_{I1}^2}{\hat\mu(x_{I1}+\hat\mu)},\ldots,\frac{2x_{In_I}^2}{\hat\mu(x_{In_I}+\hat\mu)}\right),$$
$$v_{i+1} = (\underbrace{0,\ldots,0}_{n_1},\underbrace{0,\ldots,0}_{n_2},-\frac{x_{i1}(x_{i1}-\hat\mu)}{\hat\lambda_i(x_{i1}+\hat\mu)},\ldots,-\frac{x_{in_i}(x_{in_i}-\hat\mu)}{\hat\lambda_i(x_{in_i}+\hat\mu)},\underbrace{0,\ldots,0}_{n_{i+1}},\ldots,\underbrace{0,\ldots,0}_{n_I}), \quad i=1,\ldots,I.$$

The likelihood gradients, $l_{;V}(\theta) = \left\{\frac{d}{dv_1}l(\theta;x),\ldots,\frac{d}{dv_{I+1}}l(\theta;x)\right\}$, $l_{\lambda;V}(\theta)$ and $l_{\theta;V}(\theta)$ are

$$l_{;V}(\theta) = \begin{bmatrix} \sum_{i=1}^{I}\sum_{j=1}^{n_i} l_{;x_{ij}}(\theta)\, v_{1,j+(i-1)\times n_{i-1}} \\ \vdots \\ \sum_{i=1}^{I}\sum_{j=1}^{n_i} l_{;x_{ij}}(\theta)\, v_{I+1,j+(i-1)\times n_{i-1}} \end{bmatrix},$$

$$l_{\lambda;V}(\theta) = \begin{bmatrix} \sum_{i=1}^{I}\sum_{j=1}^{n_i} l_{\lambda_1;x_{ij}}(\theta)\, v_{1,j+(i-1)\times n_{i-1}} & \cdots & \sum_{i=1}^{I}\sum_{j=1}^{n_i} l_{\lambda_I;x_{ij}}(\theta)\, v_{I+1,j+(i-1)\times n_{i-1}} \\ \vdots & & \vdots \\ \sum_{i=1}^{I}\sum_{j=1}^{n_i} l_{\lambda_1;x_{ij}}(\theta)\, v_{I+1,j+(i-1)\times n_{i-1}} & \cdots & \sum_{i=1}^{I}\sum_{j=1}^{n_i} l_{\lambda_I;x_{ij}}(\theta)\, v_{I+1,j+(i-1)\times n_{i-1}} \end{bmatrix},$$

$$l_{\theta;V}(\theta) = \begin{bmatrix} \sum_{i=1}^{I}\sum_{j=1}^{n_i} l_{\mu;x_{ij}}(\theta)\, v_{1,j+(i-1)\times n_{i-1}} & \sum_{i=1}^{I}\sum_{j=1}^{n_i} l_{\lambda_1;x_{ij}}(\theta)\, v_{1,j+(i-1)\times n_{i-1}} & \cdots & \sum_{i=1}^{I}\sum_{j=1}^{n_i} l_{\lambda_I;x_{ij}}(\theta)\, v_{1,j+(i-1)\times n_{i-1}} \\ \vdots & \vdots & \vdots & \vdots \\ \sum_{i=1}^{I}\sum_{j=1}^{n_i} l_{\mu;x_{ij}}(\theta)\, v_{I+1,j+(i-1)\times n_{i-1}} & \sum_{i=1}^{I}\sum_{j=1}^{n_i} l_{\lambda_1;x_{ij}}(\theta)\, v_{I+1,j+(i-1)\times n_{i-1}} & \cdots & \sum_{i=1}^{I}\sum_{j=1}^{n_i} l_{\lambda_1;x_{ij}}(\theta)\, v_{I+1,j+(i-1)\times n_{i-1}} \end{bmatrix},$$

respectively. The observed Fisher information matrix and the observed nuisance information matrix are

$$j_{\theta\theta}(\theta) = \begin{bmatrix} \sum_{i=1}^{I}\sum_{j=1}^{n_i}\frac{3\lambda_i x_{ij}}{\mu^4} - \sum_{i=1}^{I}\frac{2\lambda_i n_i}{\mu^3} & \frac{n_1}{\mu^2}-\sum_{i=1}^{n_1}\frac{x_{1i}}{\mu^2} & \frac{n_2}{\mu^2}-\sum_{i=1}^{n_2}\frac{x_{2i}}{\mu^3} & \cdots & \cdots & \frac{n_I}{\mu^2}-\sum_{i=1}^{n_I}\frac{x_{Ii}}{\mu^3} \\ \frac{n_1}{\mu^2}-\sum_{i=1}^{n_1}\frac{x_{1i}}{\mu^3} & \frac{n_1}{2\lambda_1^2} & 0 & 0 & \cdots & 0 \\ \frac{n_2}{\mu^2}-\sum_{i=1}^{n_2}\frac{x_{2i}}{\mu^3} & 0 & \frac{n_2}{2\lambda_2^2} & 0 & \vdots & \vdots \\ \vdots & \vdots & 0 & \ddots & 0 & \vdots \\ \vdots & \vdots & \vdots & 0 & \ddots & 0 \\ \frac{n_I}{\mu^2}-\sum_{i=1}^{n_I}\frac{x_{Ii}}{\mu^3} & 0 & 0 & \cdots & 0 & \frac{n_I}{2\lambda_I^2} \end{bmatrix}$$

and

$$j_{\lambda\lambda}(\theta) = \begin{bmatrix} \frac{n_1}{2\lambda_1^2} & & 0 \\ & \ddots & \\ 0 & & \frac{n_I}{2\lambda_I^2} \end{bmatrix}.$$ Apply the above quantities to (2.5) and (2.4), $q(\mu)$ and then $r^*(\mu)$ can be obtained.

Although the confidence intervals based on $r(\mu)$ and $r^*(\mu)$ can be obtained here, in general, some simple numerical iteration procedure is needed to solve the upper bound limit and lower bound limit. In this paper we use the so-called secant method; the algorithm is summarized as follows:

Step 1: Give the tolerance ε for the purpose of accuracy;
Step 2: Select δ for the purpose of numerical differentiation;
Step 3: Give the initial estimate μ_0 to start the iteration;
Step 4: Compute

$$\mu_1 = \mu_0 + \frac{[Z_{\alpha/2} - r(\mu_0)]}{[r(\mu_0+\delta) - r(\mu_0-\delta)]/2\delta}; \tag{3.5}$$

Step 5: If $|\mu_1 - \mu_0| > \varepsilon$, replace μ_0 with μ_1 and return to Step 4 again, otherwise take the latest μ_1 as the lower bound limit of the $100(1-\alpha)\%$ confidence interval.

Replacing $Z_{\alpha/2}$ with $Z_{1-\alpha/2}$ in (3.5), we can obtain the upper bound limit for the $100(1-\alpha)\%$ confidence interval of the common mean μ. Similarly, the confidence interval based on $r^*(\mu)$ can be obtained by substituting $r^*(\mu)$ for $r(\mu)$ in (3.5).

3.1.1. The likelihood-based confidence interval when I = 2

In order to express the proposed method in details, we present the derivation of the confidence interval for the common mean of two independent IG populations. The log-likelihood function based on the observations is:

$$l(\theta;x_1,x_2) = \frac{n_1}{2}\log\frac{\lambda_1}{2\pi} - \frac{3}{2}\sum_{i=1}^{n_1}\log x_{1i} - \frac{\lambda_1}{2\mu^2}\sum_{i=1}^{n_1}x_{1i} + \frac{\lambda_1 n_1}{\mu} - \frac{\lambda_1}{2}\sum_{i=1}^{n_1}\frac{1}{x_{1i}} + \frac{n_2}{2}\log\frac{\lambda_2}{2\pi} - \frac{3}{2}\sum_{i=1}^{n_2}\log x_{2i} - \frac{\lambda_2}{2\mu^2}\sum_{i=1}^{n_2}x_{2i} + \frac{\lambda_2 n_2}{\mu} - \frac{\lambda_2}{2}\sum_{i=1}^{n_2}\frac{1}{x_{2i}}. \tag{3.6}$$

Take $R_{ij} = \frac{\lambda_i(x_{ij}-\mu)^2}{\mu^2 x_{ij}}$ to be the pivot quantity as we mentioned earlier and differentiate R_{ij} with respect to x and θ, we then have $\left(\frac{\partial R}{\partial x}\right)^{-1} = diag\left(\frac{\mu^2x_{11}^2}{\lambda_1(x_{11}^2-\mu^2)},\ldots,\frac{\mu^2x_{1n_1}^2}{\lambda_1(x_{1n_1}^2-\mu^2)},\frac{\mu^2x_{21}^2}{\lambda_2(x_{21}^2-\mu^2)},\ldots,\frac{\mu^2x_{2n_2}^2}{\lambda_2(x_{2n_2}^2-\mu^2)}\right)$ and $\left(\frac{\partial R}{\partial \theta}\right) = (r_1', r_2', r_3')$ with $diag(\cdot)$ being a diagonal matrix and

$$r_1 = \left(\frac{-2\lambda_1(x_{11}-\mu)}{\mu^3},\ldots,\frac{-2\lambda_1(x_{1n_1}-\mu)}{\mu^3},\frac{-2\lambda_2(x_{21}-\mu)}{\mu^3},\ldots,\frac{-2\lambda_2(x_{2n_2}-\mu)}{\mu^3}\right),$$
$$r_2 = \left(\frac{(x_{11}-\mu)^2}{\mu^2x_{11}},\ldots,\frac{(x_{1n_1}-\mu)^2}{\mu^2x_{1n_1}},0,\ldots,0\right),$$
$$r_3 = \left(0,\ldots,0,\frac{(x_{21}-\mu)^2}{\mu^2x_{21}},\ldots,\frac{(x_{2n_2}-\mu)^2}{\mu^2x_{2n_2}}\right).$$

So $V = (v_1',\ldots,v_3') = -\left(\frac{\partial R}{\partial x}\right)^{-1}\left(\frac{\partial R}{\partial \theta}\right)\Big|_{\hat\theta}$ is obtained with

$$\begin{aligned} v_1 &= \left(\frac{2x_{11}^2}{\hat\mu(x_{11}+\hat\mu)},\ldots,\frac{2x_{1n_1}^2}{\hat\mu(x_{1n_1}+\hat\mu)},\frac{2x_{21}^2}{\hat\mu(x_{21}+\hat\mu)},\ldots,\frac{2x_{2n_2}^2}{\hat\mu(x_{2n_2}+\hat\mu)}\right),\\ v_2 &= \left(-\frac{x_{11}(x_{11}-\hat\mu)}{\hat\lambda_1(x_{11}+\hat\mu)},\ldots,-\frac{x_{1n_1}(x_{1n_1}-\hat\mu)}{\hat\lambda_1(x_{1n_1}+\hat\mu)},0,\ldots,0\right),\\ v_3 &= \left(0,\ldots,0,-\frac{x_{21}(x_{21}-\hat\mu)}{\hat\lambda_2(x_{21}+\hat\mu)},\ldots,-\frac{x_{2n_2}(x_{2n_2}-\hat\mu)}{\hat\lambda_2(x_{2n_2}+\hat\mu)}\right). \end{aligned} \tag{3.7}$$

Moreover, the likelihood gradients are

$$l_{;V}(\theta) = \begin{bmatrix} \sum_{i=1}^{n_1}\frac{2x_{1i}^2}{\hat\mu(x_{1i}+\hat\mu)}\cdot\left(\frac{\hat\lambda_1}{2x_{1i}^2} - \frac{\hat\lambda_1}{2\hat\mu^2} - \frac{3}{2x_{1i}}\right) + \sum_{i=1}^{n_2}\frac{2x_{2i}^2}{\hat\mu(x_{2i}+\hat\mu)}\cdot\left(\frac{\hat\lambda_2}{2x_{2i}^2} - \frac{\hat\lambda_2}{2\hat\mu^2} - \frac{3}{2x_{2i}}\right)\\ \sum_{i=1}^{n_1}\frac{x_{1i}(\hat\mu-x_{1i})}{\hat\lambda_1(x_{1i}+\hat\mu)}\cdot\left(\frac{\hat\lambda_1}{2x_{1i}^2} - \frac{\hat\lambda_1}{2\hat\mu^2} - \frac{3}{2x_{1i}}\right)\\ \sum_{i=1}^{n_2}\frac{x_{2i}(\hat\mu-x_{2i})}{\hat\lambda_2(x_{2i}+\hat\mu)}\cdot\left(\frac{\hat\lambda_2}{2x_{2i}^2} - \frac{\hat\lambda_2}{2\hat\mu^2} - \frac{3}{2x_{2i}}\right) \end{bmatrix},$$

$$l_{\lambda;V}(\theta)=\begin{bmatrix}\sum_{i=1}^{n_1}\frac{2x_{1i}^2}{\hat{\mu}(x_{1i}+\hat{\mu})}\cdot\left(\frac{1}{2x_{1i}^2}-\frac{1}{2\hat{\mu}^2}\right) & \sum_{i=1}^{n_2}\frac{2x_{2i}^2}{\hat{\mu}(x_{2i}+\hat{\mu})}\cdot\left(\frac{1}{2x_{2i}^2}-\frac{1}{2\hat{\mu}^2}\right)\\ \sum_{i=1}^{n_1}\frac{x_{1i}(\hat{\mu}-x_{1i})}{\hat{\lambda}_1(x_{1i}+\hat{\mu})}\cdot\left(\frac{1}{2x_{1i}^2}-\frac{1}{2\hat{\mu}^2}\right) & 0\\ 0 & \sum_{i=1}^{n_2}\frac{x_{2i}(\hat{\mu}-x_{2i})}{\hat{\lambda}_2(x_{2i}+\hat{\mu})}\cdot\left(\frac{1}{2x_{2i}^2}-\frac{1}{2\hat{\mu}^2}\right)\end{bmatrix} \text{ and}$$

$$l_{\theta;V}(\theta)=\begin{bmatrix}\sum_{i=1}^{2}\sum_{j=1}^{n_i}\frac{\hat{\lambda}_i}{\hat{\mu}^4}\cdot\frac{2x_{ij}^2}{x_{ij}+\hat{\mu}} & \sum_{i=1}^{n_1}\frac{2x_{1i}^2}{\hat{\mu}(x_{1i}+\hat{\mu})}\cdot\left(\frac{1}{2x_{1i}^2}-\frac{1}{2\hat{\mu}^2}\right) & \sum_{i=1}^{n_2}\frac{2x_{2i}^2}{\hat{\mu}(x_{2i}+\hat{\mu})}\cdot\left(\frac{1}{2x_{2i}^2}-\frac{1}{2\hat{\mu}^2}\right)\\ -\frac{1}{\hat{\mu}^3}\sum_{i=1}^{n_1}\frac{x_{1i}(x_{1i}-\hat{\mu})}{(x_{1i}+\hat{\mu})} & \sum_{i=1}^{n_1}\frac{x_{1i}(\hat{\mu}-x_{1i})}{\hat{\lambda}_1(x_{1i}+\hat{\mu})}\cdot\left(\frac{1}{2x_{1i}^2}-\frac{1}{2\hat{\mu}^2}\right) & 0\\ -\frac{1}{\hat{\mu}^3}\sum_{i=1}^{n_2}\frac{x_{2i}(x_{2i}-\hat{\mu})}{(x_{2i}+\hat{\mu})} & 0 & \sum_{i=1}^{n_2}\frac{x_{2i}(\hat{\mu}-x_{2i})}{\hat{\lambda}_2(x_{2i}+\hat{\mu})}\cdot\left(\frac{1}{2x_{2i}^2}-\frac{1}{2\hat{\mu}^2}\right)\end{bmatrix}.$$

Furthermore, the observed Fisher information matrix and the observed nuisance information matrix are $j_{\theta\theta}(\theta)=\begin{bmatrix}\frac{3\lambda_1 s_1}{\mu^4}-\frac{2\lambda_1 n_1}{\mu^3}+\frac{3\lambda_2 s_2}{\mu^4}-\frac{2\lambda_2 n_2}{\mu^3} & \frac{n_1}{\mu^2}-\frac{s_1}{\mu^3} & \frac{n_2}{\mu^2}-\frac{s_2}{\mu^3}\\ \frac{n_1}{\mu^2}-\frac{s_1}{\mu^3} & \frac{n_1}{2\lambda_1^2} & 0\\ \frac{n_2}{\mu^2}-\frac{s_2}{\mu^3} & 0 & \frac{n_2}{2\lambda_2^2}\end{bmatrix}$ and $j_{\lambda\lambda}(\theta)=\begin{bmatrix}\frac{n_1}{2\lambda_1^2} & 0\\ 0 & \frac{n_2}{2\lambda_2^2}\end{bmatrix}$, respectively, where $s_1=\sum_{i=1}^{n_1}X_{1i}$ and $s_2=\sum_{i=1}^{n_2}X_{2i}$. Finally, $r^*(\mu)$ is then obtained by applying these quantities to (2.5) and (2.4).

3.2. The generalized inference method [11]

Ye et al. [11] proposed a mixture of generalized inference method and the large sample theory. Their procedures for deriving two generalized pivot quantities, $\widetilde{T}_1$ and $\widetilde{T}_2$, are briefly introduced below.

Suppose $x_i=(x_{i1},x_{i2},\ldots,x_{in_i}),\ i=1,2,..,I$, are I independent populations with parameters (μ,λ_i) for each population, from (2.1) we know that $\bar{x}_i=\frac{1}{n_i}\sum_{j=1}^{n_i}x_{ij}\sim IG(\mu,n_i\lambda_i)$ and $n_i\lambda_i V_i\sim\chi^2_{n_i-1}$, where $V_i=\frac{1}{n_i}\sum_{j=1}^{n_i}\left(\frac{1}{x_{ij}}-\frac{1}{\bar{x}_i}\right)$. Let $\bar{x}_i^o$ and v_i^o denote the observed values of $\bar{x}_i$ and V_i, respectively. The author defined the first generalized pivot quantity (GP1) for common mean μ as

$$\widetilde{T}_1=\frac{\sum_{i=1}^{I}n_iR_iT_i}{\sum_{i=1}^{I}n_iR_i}, \tag{3.8}$$

where

$$R_i=\frac{n_i\lambda_iV_i}{n_iv_i^o}\sim\frac{\chi^2_{n_i-1}}{n_iv_i^o}\quad\text{and}\quad T_i=\frac{\bar{x}_i^o}{\left|1+\frac{\sqrt{n_i\lambda_i}(\bar{x}_i-\mu)}{\mu\sqrt{\bar{x}_i}}\sqrt{\frac{\bar{x}_i^o}{n_iR_i}}\right|}\overset{d}{\sim}\frac{\bar{x}_i^o}{\left|1+Z_i\sqrt{\frac{\bar{x}_i^o}{n_iR_i}}\right|} \tag{3.9}$$

are the generalized pivot quantities for λ_i and μ, respectively, based on the ith sample. It is noted that $\overset{d}{\sim}$ denotes "approximately distributed" and $\frac{\sqrt{n_i\lambda_i}(\bar{x}_i-\mu)}{\mu\sqrt{\bar{x}_i}}\overset{d}{\sim}Z_i$ when the sample sizes n_i are large and/or the scale parameter λ_i is large compared to μ_i. Let $\widetilde{T}_1(\alpha)$ be the 100αth percentile $\widetilde{T}_1$, the $100(1-\alpha)$%generalized confidence interval for μ based on $\widetilde{T}_1$ is

$$\left\{\widetilde{T}_1(\alpha/2),\widetilde{T}_1(1-\alpha/2)\right\}. \tag{3.10}$$

The generalized p-value for testing $H_0:\mu=\mu_0$ vs. $H_1:\mu\neq\mu_0$ can be computed as

$$p_1=2\min\left\{\Pr(\widetilde{T}_1\leqslant\mu_0),\Pr(\widetilde{T}_1\geqslant\mu_0)\right\}. \tag{3.11}$$

Furthermore, if the scale parameters λ_i's are known, then $\tilde{\mu}=\frac{\sum_{i=1}^{I}n_i\lambda_i\bar{x}_i}{\sum_{i=1}^{I}n_i\lambda_i}$ is the MLE of μ based on I independent IG populations and $\tilde{\mu}\sim IG(\mu,\sum_{i=1}^{I}n_i\lambda_i)$ [6]. The author provided a second generalized pivot quantity (GP2) for μ as

$$\widetilde{T}_2=\frac{R}{\left|1+\frac{\sqrt{\sum_{i=1}^{I}n_i\lambda_i}(\tilde{\mu}-\mu)}{\mu\sqrt{\tilde{\mu}}}\sqrt{\frac{R}{\sum_{i=1}^{I}n_iR_i}}\right|}\overset{d}{\sim}\frac{R}{\left|1+Z\sqrt{\frac{R}{\sum_{i=1}^{I}n_iR_i}}\right|}, \tag{3.12}$$

where $R=\frac{\sum_{i=1}^{I}n_iR_i\bar{x}_i^o}{\sum_{i=1}^{I}n_iR_i}$, $Z\sim N(0,1)$ and R_i is defined in (3.9). The $100(1-\alpha)$% generalized confidence interval for μ and the generalized p-value for testing $H_0:\mu=\mu_0$ vs. $H_1:\mu\neq\mu_0$ based on $\widetilde{T}_2$ can be respectively obtained by substituting $\widetilde{T}_2$ in place of $\widetilde{T}_1$ in (3.10) and (3.11).

3.3. Simple t-test confidence interval

For the purpose of comparison, we calculate a simple t-test confidence interval that is inspired from the analysis of reciprocals (ANORE). This method can provide an exact confidence interval when the scale parameters are homogeneous. Suppose $x_i = (x_{i1}, x_{i2}, \ldots, x_{in_i}),\ i = 1, 2, .., I$,are$I$independent populations with parameters (μ, λ_i) for each population. It can be shown that $\bar{x} = \frac{1}{N}\sum_{i=1}^{I}\sum_{j=1}^{n_i} x_{ij} \sim IG(\mu, N\lambda)$ and $(N-I)W \sim \frac{1}{\lambda}\chi^2_{N-I}$ are independent distributed, where $N = \sum_{i=1}^{I} n_i,\ \bar{x}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} x_{ij}$ and $W = \frac{1}{N-I}\sum_{i=1}^{I}\sum_{j=1}^{n_i}(x_{ij}^{-1} - \bar{x}_i^{-1})$. Moreover, from Remark 1, we know $\left|\frac{\sqrt{N}(\bar{x}-\mu)}{\mu(\bar{x}W)^{1/2}}\right|$ is the truncated student's t distribution with $N-I$ degrees of freedom. Therefore, the two-sided 100$(1-\alpha)$% for μ is

$$\begin{cases} \left[\bar{x}\left(1 + t_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{x}W}{N}}\right)^{-1}, \bar{x}\left(1 - t_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{x}W}{N}}\right)^{-1}\right], & \text{if } 1 - t_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{x}W}{N}} > 0 \\ \left[\bar{x}\left(1 + t_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{x}W}{N}}\right)^{-1}, \infty\right], & \text{otherwise.} \end{cases} \tag{3.13}$$

4. Simulation studies and numerical examples

4.1. Simulation studies

To illustrate the merits of the proposed method, we present simulation studies of the confidence intervals and type I errors applied to a variety of scale parameter configurations and different combinations of small sample sizes for two and three populations. In the simulation, we exhibit the coverage probabilities, the average lengths of the.95 confidence intervals and also evaluate the type I errors at the nominal significance level.05 based on $r(\mu)$, $r^*(\mu)$, two generalized pivot quantities (GP1 and GP2), and the simple t-test method (S.T.). The results given in Tables 1–4 below are based on 10,000 simulation runs for each combination.

From Tables 1 and 2, we see that although the confidence intervals based on $r(\mu)$ and GP2 have shorter average lengths comparing to the other three methods, the coverage probabilities are too liberal to attain the proposed coverage probabilities of.95 for each combination. The confidence intervals based on the simple t-test method have good coverage probabilities but these coverage probabilities decrease when the heterogeneity increases. Moreover, when the scale parameter is small relative to μ, the interval lengths constructed by the simple t-test are unbounded (i.e., a one-sided interval). In these cases, the simple t-test method gives less information about the target value than those based on the other methods. The GP1 method performs well in the coverage probabilities when the scale parameters are large compared to mu, but the performance grows worse when the scale parameters decrease. On the other hand, the confidence intervals based on $r^*(\mu)$ not only have almost exact coverage probabilities in each combination (except for few exceptions), but the average lengths are also quite decent and acceptable. Therefore, in terms of the overall comparisons, the higher order likelihood-based method outperforms the other four methods.

Table 1
Simulation results of 95% confidence interval of μ = 1 for two populations.

(n_1, n_2)	λ_1	λ_2	$r^*(\mu)$		GP1		GP2		$r(\mu)$		S.T.	
			CP	Length	CP	Length	CP	Length	CP	Length	CP	Length
(5,10)	0.2	1	0.951	9.387	0.963	5.451	0.933	1.810	0.923	1.637	0.954	∞
	0.5	1	0.948	14.340	0.961	1.335	0.935	1.000	0.917	1.554	0.945	∞
	1	3	0.952	1.054	0.953	1.475	0.935	1.064	0.926	0.786	0.943	1.024
	3	10	0.947	0.456	0.943	0.444	0.937	0.384	0.920	0.387	0.939	0.486
	1	10	0.948	0.477	0.950	0.456	0.940	0.400	0.923	0.409	0.933	0.791
(10,5)	0.2	1	0.931	23.093	0.969	5.489	0.939	2.261	0.847	1.368	0.955	∞
	0.5	1	0.944	13.672	0.959	12.439	0.930	6.529	0.897	1.534	0.949	∞
	1	3	0.949	1.958	0.955	2.633	0.935	1.481	0.906	0.967	0.950	1.506
	3	10	0.948	0.598	0.945	0.746	0.938	0.603	0.903	0.464	0.949	0.623
	1	10	0.947	0.840	0.945	0.845	0.934	0.679	0.925	0.573	0.951	1.361
(10,10)	0.2	1	0.951	7.914	0.968	2.803	0.944	1.309	0.929	1.659	0.958	∞
	0.5	1	0.955	2.320	0.955	4.332	0.933	1.561	0.933	1.363	0.948	∞
	1	3	0.945	0.873	0.954	0.896	0.939	0.655	0.924	0.708	0.946	0.918
	3	10	0.945	0.410	0.943	0.358	0.938	0.331	0.921	0.360	0.947	0.456
	1	10	0.949	0.459	0.950	0.484	0.943	0.375	0.926	0.399	0.947	0.795

CP: coverage probability; length: average length.

Table 2
Simulation results of 95% confidence interval of $\mu = 1$ for three populations.

(n_1,n_2,n_3)	λ_1	λ_2	λ_3	$r^*(\mu)$		GP1		GP2		$r(\mu)$		S.T.	
				CP	Length	CP	Length	CP	Length	CP	Length	CP	Length
(5,8,10)	0.1	0.1	1	0.949	4.865	0.972	6.162	0.937	2.856	0.923	1.611	0.965	∞
	0.1	0.5	1	0.949	3.382	0.966	21.096	0.934	7.197	0.922	1.456	0.959	∞
	1	1	5	0.948	0.671	0.952	0.807	0.934	0.580	0.923	0.542	0.948	0.807
	1	1	10	0.949	0.467	0.958	0.606	0.939	0.429	0.925	0.396	0.943	0.766
	1	5	10	0.948	0.393	0.944	0.497	0.931	0.402	0.921	0.337	0.938	0.511
(5,10,8)	0.1	0.1	1	0.952	5.964	0.972	8.408	0.937	6.293	0.917	1.589	0.963	∞
	0.1	0.5	1	0.948	3.784	0.970	7.983	0.930	2.515	0.918	1.492	0.952	∞
	1	1	5	0.945	0.754	0.958	0.618	0.938	0.433	0.919	0.586	0.951	0.873
	1	1	10	0.950	0.537	0.957	0.559	0.939	0.380	0.919	0.435	0.947	0.844
	1	5	10	0.945	0.413	0.947	0.591	0.929	0.471	0.917	0.350	0.938	0.524
(10,8,5)	0.1	0.1	1	0.930	7.016	0.975	7.526	0.932	2.687	0.839	1.306	0.967	∞
	0.1	0.5	1	0.946	5.952	0.970	11.160	0.934	5.868	0.889	1.467	0.965	∞
	1	1	5	0.943	0.957	0.950	0.977	0.930	0.500	0.901	0.668	0.950	0.974
	1	1	10	0.945	0.720	0.945	1.063	0.935	0.544	0.898	0.518	0.953	0.956
	1	5	10	0.946	0.521	0.946	0.507	0.933	0.441	0.902	0.406	0.948	0.711

CP: coverage probability; length: average length.

Table 3
Type I errors for $H_0 : \mu = \mu_0$ vs. $H_1 : \mu \neq \mu_0$ at $I = 2$ and $\alpha = 0.05$.

(λ_1,λ_2)	Tests	$n_1 = 5$, $n_2 = 10$					$n_1 = 10$, $n_2 = 5$				
		μ_0					μ_0				
		0.2	0.8	1.2	2.0	5.0	0.2	0.8	1.2	2.0	5.0
(0.2,1)	$r^*(\mu)$	0.0522	0.0528	0.0529	0.0493	0.0506	0.0550	0.0495	0.0567	0.0551	0.0524
	GP1	0.0439	0.0432	0.0382	0.0310	0.0344	0.0514	0.0345	0.0275	0.0317	0.0259
	GP2	0.0575	0.0700	0.0643	0.0610	0.0607	0.0658	0.0650	0.0605	0.0677	0.0714
	$r(\mu)$	0.0774	0.0746	0.0750	0.0686	0.0702	0.1027	0.0903	0.0989	0.0912	0.0912
	S.T.	0.0615	0.0430	0.0489	0.0426	0.0324	0.0502	0.0378	0.0410	0.0381	0.0313
(0.5,1)	$r^*(\mu)$	0.0485	0.0562	0.0569	0.0553	0.0575	0.0528	0.0560	0.0522	0.0552	0.0530
	GP1	0.0590	0.0363	0.0382	0.0337	0.0260	0.0495	0.0394	0.0340	0.0302	0.0244
	GP2	0.0685	0.0632	0.0655	0.0662	0.0657	0.0624	0.0642	0.0687	0.0672	0.0714
	$r(\mu)$	0.0832	0.0862	0.0844	0.0813	0.0808	0.0901	0.0950	0.0922	0.0922	0.0883
	S.T.	0.0555	0.0507	0.0514	0.0504	0.0509	0.0469	0.0515	0.0461	0.0434	0.0484
(1,3)	$r^*(\mu)$	0.0526	0.0528	0.0500	0.0578	0.0564	0.0542	0.0548	0.0573	0.0570	0.0553
	GP1	0.0574	0.0514	0.0440	0.0402	0.0352	0.0609	0.0554	0.0437	0.0417	0.0294
	GP2	0.0619	0.0660	0.0627	0.0645	0.0679	0.0655	0.0655	0.0660	0.0670	0.0662
	$r(\mu)$	0.0783	0.0783	0.0744	0.0810	0.0790	0.0966	0.0985	0.0941	0.0939	0.0939
	S.T.	0.0583	0.0545	0.0503	0.0539	0.0524	0.0514	0.0471	0.0464	0.0535	0.0424
(1,5)	$r^*(\mu)$	0.0507	0.0566	0.0517	0.0551	0.0570	0.0585	0.0537	0.0558	0.0563	0.0575
	GP1	0.0587	0.0467	0.0472	0.0404	0.0355	0.0602	0.0562	0.0502	0.0395	0.0347
	GP2	0.0600	0.0595	0.0647	0.0575	0.0614	0.0609	0.0670	0.0674	0.0644	0.0684
	$r(\mu)$	0.0788	0.0803	0.0768	0.0779	0.0790	0.1009	0.0936	0.0990	0.0980	0.0989
	S.T.	0.0634	0.0616	0.0610	0.0573	0.0499	0.0502	0.0471	0.0465	0.0460	0.0371
(1,10)	$r^*(\mu)$	0.0528	0.0550	0.0484	0.0559	0.0487	0.0559	0.0538	0.0517	0.0582	0.0571
	GP1	0.0559	0.0497	0.0497	0.0444	0.0357	0.0579	0.0452	0.0512	0.0444	0.0357
	GP2	0.0622	0.0610	0.0609	0.0590	0.0540	0.0617	0.0565	0.0619	0.0595	0.0565
	$r(\mu)$	0.0751	0.0759	0.0727	0.0777	0.0697	0.1040	0.0977	0.0967	0.0997	0.0981
	S.T.	0.0775	0.074	0.0682	0.0589	0.0478	0.0532	0.0468	0.0507	0.0431	0.0344

Furthermore, from Tables 3 and 4, we can see the type I errors based on the simple t-test method are not stable since the type I errors decrease as the mean parameter under the null hypothesis increases. Similarly, the type I errors obtained by the GP1 methods do not perform well under those small sample sizes and small scale parameters configurations. The type I errors based on the GP1 method are getting worse compared to the nominal level.05 as the mean parameters increase. The type I errors based on GP2 and $r(\mu)$ are respectively around 0.6 to.07 and.07 to.10 for each combination which are too large compared to the nominal level.05. By contrast, the type I errors based on $r^*(\mu)$ are not only stable, but the values are also very close to the nominal level .05. Thus, we can say that the proposed procedure can well tolerate heterogeneity among populations and give robust and reliable results under different scenarios.

Table 4
Type I errors for $H_0 : \mu = \mu_0$ vs. $H_1 : \mu \neq \mu_0$ at $I = 3$ and $\alpha = 0.05$.

$(\lambda_1,\lambda_2,\lambda_3)$	Tests	$(n_1,n_2,n_3)=(5,8,10)$					$(n_1,n_2,n_3)=(5,10,8)$				
		μ_0					μ_0				
		0.2	0.8	1.2	2.0	5.0	0.2	0.8	1.2	2.0	5.0
(0.1,0.1,1)	$r^*(\mu)$	0.0481	0.0564	0.0542	0.0563	0.0551	0.0550	0.0539	0.0516	0.0576	0.0549
	GP1	0.0424	0.0274	0.0268	0.0270	0.0260	0.0388	0.0274	0.0224	0.0200	0.0234
	GP2	0.0656	0.0600	0.0620	0.0648	0.0638	0.0640	0.0654	0.0608	0.0558	0.0668
	$r(\mu)$	0.0716	0.0774	0.0748	0.0796	0.0729	0.0845	0.0789	0.0769	0.0814	0.0787
	S.T.	0.0483	0.0391	0.0336	0.0258	0.0187	0.0472	0.0355	0.0290	0.0247	0.0194
(0.1,0.5,1)	$r^*(\mu)$	0.0539	0.0572	0.0528	0.0548	0.0554	0.0516	0.0563	0.0534	0.0570	0.0525
	GP1	0.0558	0.0314	0.0282	0.0254	0.0146	0.0452	0.0364	0.0300	0.0240	0.0156
	GP2	0.0684	0.0666	0.0642	0.0682	0.0666	0.0614	0.0684	0.0658	0.0656	0.0672
	$r(\mu)$	0.0812	0.0832	0.0749	0.0761	0.0769	0.0778	0.0841	0.0804	0.0804	0.0791
	S.T.	0.0594	0.0462	0.0395	0.0377	0.0283	0.0558	0.0450	0.0422	0.0378	0.0271
(1,1,5)	$r^*(\mu)$	0.0520	0.0505	0.0560	0.0567	0.0558	0.0517	0.0525	0.0542	0.0548	0.0577
	GP1	0.0610	0.0470	0.0464	0.0424	0.0284	0.0582	0.0504	0.0412	0.0372	0.0316
	GP2	0.0646	0.0670	0.0682	0.0724	0.0634	0.0650	0.0708	0.0668	0.0658	0.0740
	$r(\mu)$	0.0782	0.0755	0.0797	0.0824	0.0768	0.0843	0.0840	0.0839	0.0800	0.0864
	S.T.	0.0572	0.0594	0.0557	0.0510	0.0436	0.0544	0.0516	0.0524	0.0472	0.0409
(1,1,10)	$r^*(\mu)$	0.0485	0.0516	0.0494	0.0564	0.0557	0.0549	0.0541	0.0528	0.0543	0.0581
	GP1	0.0570	0.0478	0.0482	0.0400	0.0294	0.0584	0.0536	0.0436	0.0408	0.0304
	GP2	0.0620	0.0640	0.0662	0.0648	0.0614	0.0640	0.0682	0.0608	0.0668	0.0648
	$r(\mu)$	0.0746	0.0724	0.0730	0.0792	0.0767	0.0853	0.0852	0.0797	0.0820	0.0864
	S.T.	0.0579	0.0534	0.0537	0.0522	0.0394	0.0588	0.0496	0.0484	0.0458	0.0399
(1,5,10)	$r^*(\mu)$	0.0517	0.0514	0.0514	0.0537	0.0524	0.0518	0.0534	0.0540	0.0557	0.0525
	GP1	0.0642	0.0540	0.0556	0.0448	0.0410	0.0566	0.0520	0.0476	0.0490	0.0398
	GP2	0.0688	0.0682	0.0680	0.0658	0.0656	0.0624	0.0658	0.0646	0.0684	0.0666
	$r(\mu)$	0.0779	0.0777	0.0775	0.0815	0.0766	0.0800	0.0828	0.0827	0.0865	0.0806
	S.T.	0.0638	0.0659	0.0660	0.0553	0.0544	0.0634	0.0625	0.0611	0.0603	0.0477

Table 5
Data for Example 1.

Population i	1	2	3
	0.7312	1.3932	1.6999
	1.7314	0.5934	1.2698
	0.7109	1.6046	0.7887
	0.0303	2.0649	1.0535
	0.7044	1.2238	0.7973
		0.0538	1.4988
			1.4685
$\bar{x}_i$	0.7816	1.1556	1.2252
w_i	31.3779	17.7229	0.4820

$w_i = \sum_{j=1}^{n_i}(x_{ij}^{-1} - \bar{x}_i^{-1})$.

Table 6
The 95% confidence intervals for the common mean.

Method	Point estimate $\hat{\mu}$	Interval estimate	Length
$r^*(\mu)$	1.221	(0.961, 1.728)	0.767
GP1	1.817	(0.972, 2.089)	1.117
GP2	1.258	(0.952, 1.691)	0.739
$r(\mu)$	1.221	(0.980, 1.605)	0.625
S.T.	1.078	(0.553, 20.711)	20.158

Table 7
Data for Example 2 ($n_i = 10, i = 1,2,3,4$).

$\bar{x}_i$	2.635	2.055	1.748	2.023
w_i	4.5147	107.7351	3.1558	67.4330

$w_i = \sum_{j=1}^{n_i}(x_{ij}^{-1} - \bar{x}_i^{-1})$.

Table 8
The 95% confidence intervals for the common mean.

Method	Point estimate $\hat{\mu}$	Interval estimate	Length
$r^*(\mu)$	2.110	(1.484, 4.384)	2.900
GP1	3.249	(1.511, 7.740)	6.229
GP2	2.232	(1.446, 3.736)	2.290
$r(\mu)$	2.110	(1.480, 3.653)	2.173
S.T.	2.116	$(1.031, \infty)$	∞

4.2. Two numerical examples

Example 1. We first present a three population IG simulated data with $(n_1, n_2, n_3) = (5, 6, 7)$ and $(\mu, \lambda_1, \lambda_2, \lambda_3) = (1, 0.2, 1, 10)$ as illustrative example. The original data and the summary data are depicted in Table 5. The interval estimations based on five methods are given in Table 6. Four confidence intervals based on $r^*(\mu)$, $r(\mu)$, GP1 and GP2 give satisfactory result under the heterogeneous data set when compared with that based on the simple t-test method. Although the one based on $r^*(\mu)$ is a little wider than those of GP2 and $r(\mu)$, in general, it gives better coverage probabilities compared with them.

Example 2. The data is available in p.462, Nelson [20]. In this data, there are 60 "times-to-breakdown" in minutes of an insulating fluid subjected to high voltage stress. Since IG distribution is widely applied as a lifetime model in reliability analysis, here we consider the failure time of the insulating fluid for each group as an IG distributed random variable. If the experiment was under control, the mean of each group should be the same. For illustrative purpose, we pick the first four groups as demonstration and apply the procedure induced by Tian [10] to test the equality of the means for the first four groups. The resulting p-value is.8693; we can follow up by constructing the confidence interval for the common mean parameter. The summary data and the results are given in Tables 7 and 8, respectively.

From Table 7, we see w_i, $i = 1, 2, 3, 4$ the estimators of the reciprocal of the scale parameters are quite different among groups implying the existence of heterogeneity.

In Table 8, all five intervals cover the corresponding point estimates and those based on $r^*(\mu)$, $r(\mu)$ and GP2 give satisfactory interval lengths compared to GP1 and the simple t-test. In this case, the simple t-test only provides a one-sided interval.

5. Conclusions

In this paper, we presented a higher order likelihood-based procedure to construct the confidence interval of the common mean of several independent IG populations. In our simulation, we compared this procedure with four alternative methods. The numerical examples showed that the proposed method gives nearly exact coverage probabilities and the type I errors calculated are close to the nominal level.05 even for small sample sizes and small scale parameters. The method is able to integrate the information of several heterogeneous IG populations, and therefore is useful for a variety of practical applications.

Acknowledgement

The author cordially thanks the Editor-in-Chief and anonymous reviewers for their valuable comments which led to the improvement of this paper. This research has been supported by grant NSC 98-2118-M-025-001 from the National Science Council of Taiwan.

References

[1] M.E. Wise, Skew probability curves with negative powers of time and related to random walks in series, Statistica Neerlandica 25 (1971) 159–180.
[2] M.E. Wise, Skew Distributions in Biomedicine Including Some with Negative Powers of Time in Statistical Distributions in Scientific Work, vol. 2, D. Reidel, Dordrecht, Holland, 1975.
[3] M.E. Wise, S.B. Osborn, J. Anderson, R.W.S. Tomlinson, A stochastic model for turnover of radiocalcium based on the observed laws, Math. Biosci. 2 (1968) 199–224.
[4] T. Lancaster, A stochastic model for the duration of a strike, J. Roy. Stat. Soc. A 135-2 (1972) 257–271.
[5] S.K. Bhattacharya, Bayesian normal analysis with an inverse Gaussian prior, Ann. Inst. Statist. Math., Part A (1987) 623–626.
[6] R.S. Chhikara, J.L. Folks, The Inverse Gaussian Distribution, Marcel Dekker, New York, 1989.
[7] V. Seshadri, The Inverse Gaussian Distribution: A Case Study in Exponential Families, Oxford University Press, New York, 1993.
[8] V. Seshadri, The Inverse Gaussian Distribution: Statistical Theory and Applications, Springer, New York, 1999.
[9] G.S. Mudholkar, L. Tian, An entropy characterization of the inverse Gaussian distribution and related goodness-of-fit tests, J. Stat. Planning Inference 102 (2002) 211–221.
[10] L. Tian, Testing equality of inverse Gaussian means under heterogeneity based on generalized test variable, Comput. Stat. Data Anal. 51 (2006) 1156–1165.
[11] R.D. Ye, T.F. Ma, S.G. Wang, Inferences on the common mean of several inverse Gaussian populations, Comput. Stat. Data Anal. 54 (2010) 906–915.
[12] N. Reid, Likelihood and higher-order approximations to tail areas: a review and annotated bibliography, Can. J. Statist. 24 (1996) 141–166.
[13] J. Wu, A.C.M. Wong, G. Jiang, Likelihood-based confidence interval for a log-normal mean, Stat. Med. 22 (2003) 1849–1860.

[14] J. Wu, A.C.M. Wong, Improved interval estimation for the two-parameter Birnbaum–Saunders distribution, Comput. Stat. Data Anal. 47 (2004) 809–821.
[15] L. Tian, G.E. Wilding, Confidence interval of the ratio of means of two independent inverse Gaussian distributions, J. Stat. Planning Inference 133 (2005) 381–386.
[16] D.R. Cox, D.V. Hinkley, Theoretical Statistics, Chapman & Hall, London, 1974.
[17] O.E. Barndorff-Nielsen, Inference on full and partial parameters based on the standardized signed log likelihood ratio, Biometrika 73 (1986) 307–322.
[18] O.E. Barndorff-Nielsen, Modified signed log-likelihood ratio, Biometrika 78 (1991) 557–563.
[19] D.A.S. Fraser, N. Reid, J. Wu, A simple general formula for tail probabilities for frequentist and Bayesian inference, Biometrika 86 (1999) 249–264.
[20] W. Nelson, Applied Life Data Analysis, Addison-Wesley, 1982.