國 立 交 通 大 學

統計學研究所

碩 士 論 文

利用高階概似方法對多個逆高斯分配的共同平均數建構信賴區間

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

研 究 生　：　吳益銘

指導教授　：　林淑惠　博士

洪慧念　博士

中 華 民 國 九 十 六 年 六 月

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

研 究 生: 吳益銘　　Student: I-Ming Wu
指導教授: 林淑惠 博士　　Advisors: Dr. Shu-Hui Lin
洪慧念 博士　　Dr. Hui-Nien Hong

國 立 交 通 大 學
統計學研究所
碩 士 論 文

A Thesis

Submitted to Institute of Statistics College of Science

National Chiao Tung University

in Partial fulfillment of the Requirements

for the Degree of

Master

in

Statistics

June 2007

Hsinchu, Taiwan, Republic of China

中華民國九十六年六月

利用高階概似方法對多個逆高斯分配的共同平均數建構信賴區間

研究生：吳益銘　　　　　　　　指導教授：林淑惠 博士

洪慧念 博士

國立交通大學統計學研究所

摘要

這篇文章裡我們討論了關於多個異質的逆高斯分配的共同平均數信賴區間的估計。我們所利用的方法是以概似函數為基礎的高階近似方法，並且我們利用模擬的方法去跟 signed log-likelihood ratio 及 simple t-test 兩種方法比較所建構出的信賴區間的覆蓋機率和平均寬度以及所對應檢定的型 Ⅰ 錯誤來檢視我們所利用方法的優劣。結果顯示我們所提出的方法即使在小樣本的情況下表現的很好，而且相對來說比前述的兩種方法可以提供更穩定、正確的結果。最後我們提供了兩個例子作為說明。

關鍵字 ： 覆蓋機率;期望長度;逆高斯分配;樞紐量;符號對數概似比;型Ⅰ誤差

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

Student: I-Ming Wu

Advisors: Dr. Shu-Hui Lin
Dr. Hui-Nien Hong

Institute of Statistics

National Chiao Tung University

Abstract

An interval estimation method for the common mean of several non-homogeneous inverse Gaussian (IG) populations is discussed. The proposed method is based on a higher order likelihood-based asymptotic procedure. The merits of the proposed method are numerically compared with the signed log-likelihood ratio statistic 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 even for very small samples. The methods are also applied to two examples.

Keywords: Coverage probability; Expected length; Inverse Gaussian; Pivotal quantity; Signed log-likelihood ratio statistic; Type I error

誌　謝

感覺時間過的很快，在交大統計所的兩年一下子就過去了。這兩年裡認識了很棒的同學，學到了很多的東西，覺得非常值得。首先我想先感謝的是李昭勝老師，在我推甄上財金和統計所茫然不知如何決定的時候，從美國打電話來給我提供了相當寶貴的意見。 進入統計所後，在學業以及其他方面李昭勝老師也都給予我相當多的指導照顧。 然而不幸的老師在這學期突然離我們而去，留給我們無比的傷悲。記取老師從一開始就給我們的指導，希望我們可以做出一篇可以發表的論文，因此我也很努力的把論文做好。 整篇論文得以完成，我要非常感謝學姊林淑惠老師，不斷的給予我意見、討論和鼓勵，讓我得以順利的完成這篇文章。同時，我要感謝論文的口試委員: 陳鄰安、洪慧念、黃景業、林淑惠老師所提供的寶貴意見，使得本論文得以更臻完整，我在此致上最深的謝意。

碩一的這年，很感謝同學們的互相鼓勵和支持，俊睿、建威、永在、阿淳、阿Q、花花、映伶、小米、侑俊等在課業和生活上給予我無比的支持，真的很感謝你們。 還有最特別的朋友，雅崢，謝謝你這段時間來的支持及鼓勵，給我信心和勇氣，我會永遠銘記在心。 妳們的支持是我前進的力量，並帶給我歡笑和溫馨的回憶，在此我致上我最真誠的感謝。

最後，僅將這篇論文獻給我所有的家人，謝謝你們在這兩年來一直在我背後默默的支持我和鼓勵我，讓我可以專心於課業和研究，為自己的理想努力和打拼，我願將我的成就與你們一同分享。

再次感謝支持我的家人，以及所有師長和朋友們，謝謝你們!

吳益銘　　謹誌于

國立交通大學統計研究所

中華民國九十六年六月

Contents

1. Introduction

In many application areas, such as demography, management science, hydrology, finance, etc., data are frequently positive and right-skewed. Recently, the inverse Gaussian (IG) distribution has drawn many attentions and the inferences concerned with the IG distribution have also grown rapidly because the IG distribution is an ideal candidate for modeling and analyzing the right-skewed and positive data. For instance, Wise (1971, 1975) and Wise et al. (1968) developed the IG population as a possible model to describe cycle time distribution for particles in the blood; Lancaster (1972) made use of the IG distribution in describing strike duration data, etc.

The history of the inverse Gaussian distribution can be traced back to 1915 when Schrödinger and Smoluchowski presented independent derivations of the density of the first passage time distribution of Bownian motion with positive drift. The modern day statistical community became aware of this distribution through the pioneering work of Tweedie (1941, 1945, 1946, 1947, 1956, 1957). The name “inverse Gaussian” was also given by Tweedie based on his discovers that the cumulant function of IG distribution is the inverse of the cumulant function of the normal distribution. For further details about the IG distribution and for applications, the reader is referred to the books by Chhikara and Folks (1989) and Seshadri (1999), respectively.

The probability density function (pdf) of IG distribution, $\mathrm{IG}(\mu,\lambda)$, is defined as

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

where μ is the mean parameter and λ is the scale parameter. We present some density curves in Fig.1.1 and Fig. 1.2 for various values of μ at $\lambda=1$ and λ at $\mu=1$.

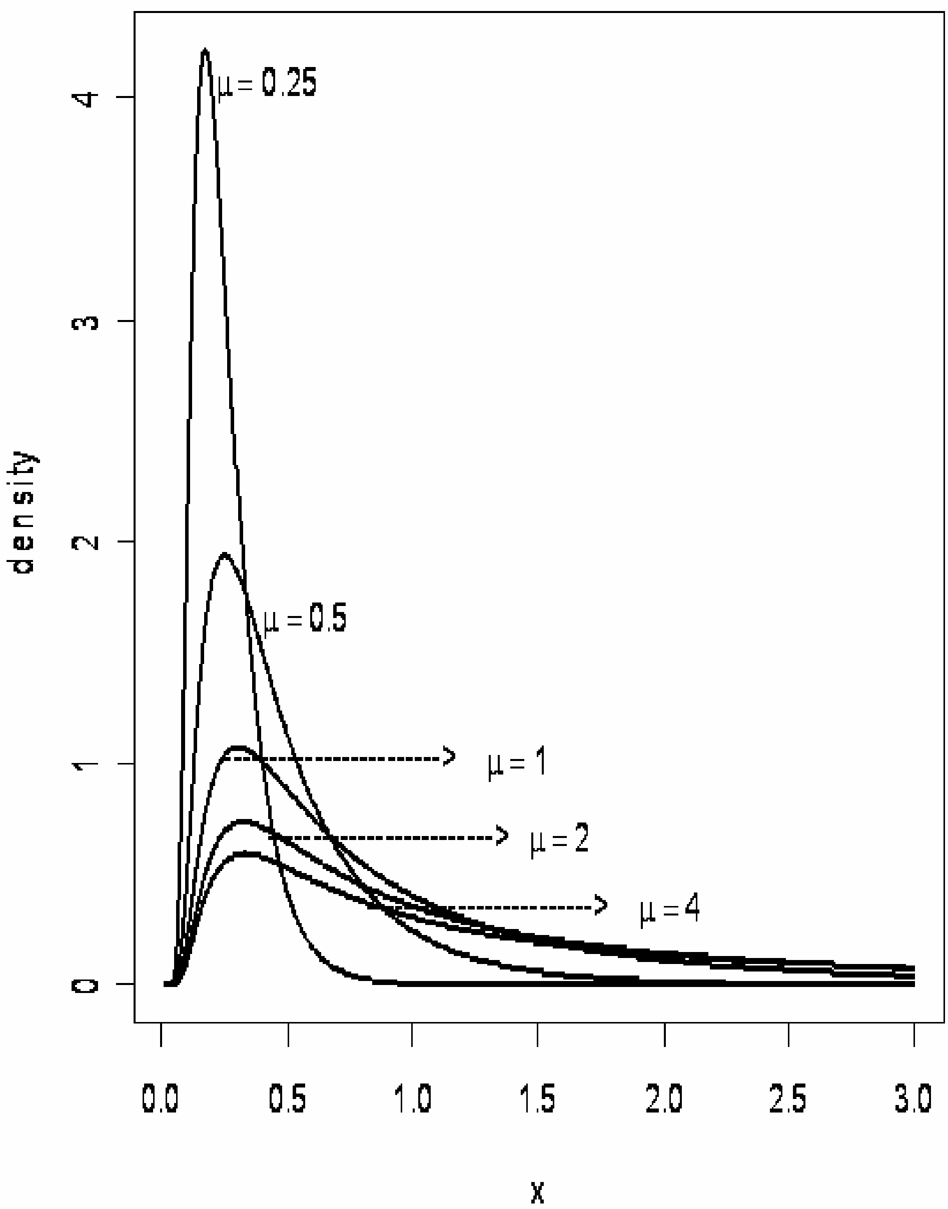

Fig. 1.1 *Density functions of* $IG(\mu,\lambda)$ *for fixed* $\lambda=1$ *and five values of* μ

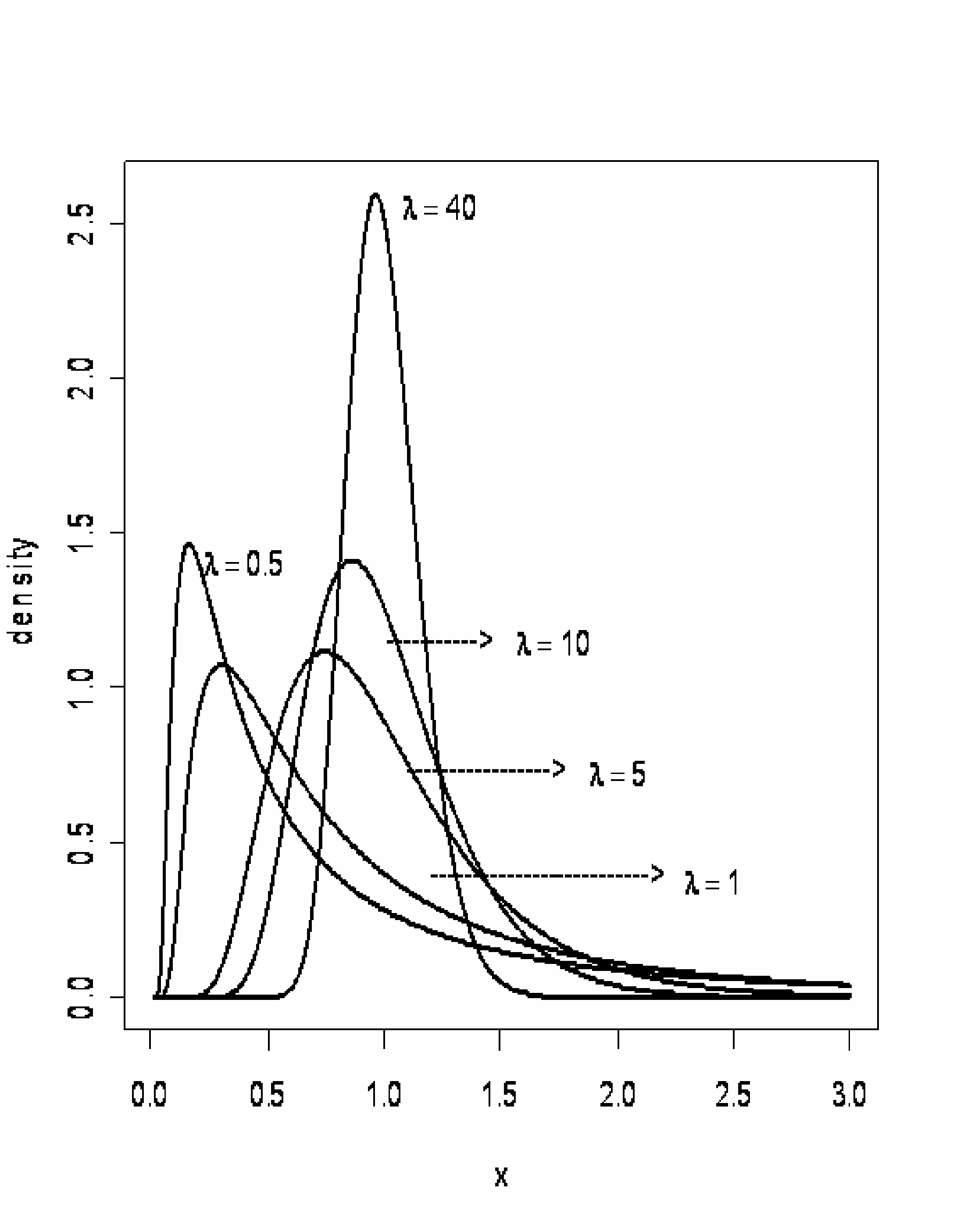

Fig. 1.2 *Density functions of* $IG(\mu,\lambda)$ *for fixed* $\mu = 1$ *and five values of* λ

In Fig. 1.1 and Fig. 1.2, the density of IG distribution varies from the highly right-skewed to almost symmetric for different parameter configurations. Hence, the inverse Gaussian distribution is very flexible in describing various data.

The inference methods of the IG model are very analogous to those of the Gaussian model; for example, a very common problem in applied field is to compare the means of several Gaussian populations, i.e.

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_I \quad \text{vs.} \quad H_A : \text{ not all } \mu_i\text{'}s \text{ are equal}. \tag{1.2}$$

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 also be used to test the equality of means of several IG samples if all scale parameters among groups are assumed to be equal (Tweedie, 1956 ; Chhikara and Folks, 1989). The testing procedure will be briefly introduced in Remark 1. But when the scale parameters are non-homogeneous, the ANORE fails to solve the problem as ANOVA fails to test (1.2) when these populations are not homogeneous. Tian (2005) proposed a method to test the equality of IG means under heterogeneity based on generalized test variable method. However, if the null hypothesis is not been rejected, the inferences for the common mean remain unsolved. Therefore, in this paper, we would like to estimate and construct the $100(1-\alpha)\%$ confidence interval for the common mean of several non-homogeneous IG populations. Our method is based on a higher order asymptotic likelihood based method. This method, in theory, has a higher order accuracy, $O(n^{-3/2})$, and is very accurate even when the sample size is small. Reid (1996) gave some review and annotation of the development. The method has also been applied to solve many practical problems involving interval estimation for a skewed distribution, e.g. Wu et al. (2002) applied this procedure to make the confidence interval estimation of the ratio of two independent lognormal distribution, Wu et al. (2003) presented a confidence interval for a log-normal mean based on this method; Wu and Wong (2004) used the method to improve the interval estimation for the two-parameter Birnbaum-Saunders distribution; Tian and Wilding (2005) used the method to construct confidence interval for the ratio of means of two independent IG distributions, etc. In our case, the likelihood-based method gives a satisfactory result as well.

Remark 1. One of the resemblances between the IG distribution and the ordinary Gaussian distribution is the method for the analysis of residuals. We are familiar with the analysis of variance (ANOVA) in normal inference theory, Tweedie (1956) introduce the so called analysis of reciprocals (ANORE) for the inverse Gaussian distribution. The procedure is as follows:

Assume that there are n_i components in the i^{th} populations and each population is distributed as $\text{IG}(\mu_i, \lambda)$, where μ_i, $i = 1, \ldots, I$ and λ are unknown. Further assume the random samples X_{ij}, $i = 1, \ldots, I$, $j = 1, \ldots, n_i$ are independent. We are interested in the problem of testing (1.2)

The likelihood function, denoted by $L(\mu_1, \ldots, \mu_I, \lambda; X_{ij} = x_{ij},\ i = 1, \ldots, I,\ j = 1, \ldots, n_I)$ is proportion to

$$\lambda^{\frac{N}{2}} \exp\left(-\frac{\lambda}{2} \sum_{i=1}^{I} \sum_{j=1}^{n_i} \frac{(x_{ij} - \mu_i)^2}{\mu_i^2 x_{ij}}\right) \tag{1.3}$$

where $N = \sum_{i=1}^{I} n_i$. Differentiation with respect to μ_i and λ yields the following maximum likelihood estimates (MLEs), $\hat{\theta} = (\hat{\mu}_1, \ldots, \hat{\mu}_I, \hat{\lambda})$, where

$$\begin{aligned} \hat{\mu}_i &= \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij} = \bar{x}_i, \\ N \hat{\lambda}^{-1} &= \sum_{i=1}^{I} \sum_{j=1}^{n_i} (x_{ij}^{-1} - \bar{x}_i^{-1}). \end{aligned} \tag{1.4}$$

While H_o is true, the estimates are

$$
\begin{aligned}
\hat{\mu} &= \frac{1}{n}\sum_{i=1}^{I}\sum_{j=1}^{n_i} x_{ij} = \bar{x}, \\
N\hat{\lambda}_0 &= \sum_{i=1}^{I}\sum_{j=1}^{n_i} (x_{ij}^{-1} - \bar{x}^{-1}).
\end{aligned}
\tag{1.5}
$$

The likelihood ratio can be reduced to

$$
\Lambda^{\frac{2}{N}} = \frac{N\hat{\lambda}^{-1}}{N\hat{\lambda}_0^{-1}} \equiv \frac{Q_o}{Q}
\tag{1.6}
$$

Note that Q can be decomposed into

$$
Q = Q_0 + \sum_{i=1}^{I} n_i \bar{x}_i^{-1} - N\bar{x}^{-1} \equiv Q_0 + Q_1
\tag{1.7}
$$

It is easy to verify that λQ_0 follows a chi-squared distribution with degrees of freedom $N-I$ while λQ_1 is a chi-squared distribution with degrees of freedom of ($I-1$) and Q_o and Q_1 are independent. It follows that the likelihood ratio test statistic

$$
\frac{(N-I)Q_1}{(I-1)Q_0} \sim F_{I-1,N-I},
\tag{1.8}
$$

where $F_{I-1,N-I}$ is the F-distribution with degrees of freedom $I-1$ and $N-I$ and the α level rejection region is given by solving the following inequality

$$
\frac{(N-I)Q_1}{(I-1)Q_0} > F_{I-1,N-I,1-\alpha}.
\tag{1.9}
$$

For illustration examples of ANORE, see Chhikara and Folks (1989) and Seshadri (1999) for further details.

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. Then the method is applied to construct a confidence interval for the common mean of several independent IG populations in section 3. The classical procedure under the assumption of identical scale is also described in section 3. We will present two numerical examples and two simulation studies in section 4 to illustrate the merits of our proposed method. Some concluding remarks are given in section 5.

2. A general review

In section 2.1 we give some basic characteristics of the IG distribution and some useful sampling distributions which will be used in later analysis. The main appeal of this article, the likelihood based inference technique, will be introduced in section 2.2.

2.1 Some properties of IG distribution

Let X be an inverse Gaussian distributed variate with parameters μ and λ. The probability density function of X is given in (1.1), and the distribution function $F(x)$ can be written in terms of the distribution function of the standard normal variate, $\Phi(x)$, as

$$F(x)=\Phi[\sqrt{\frac{\lambda}{x}}(\frac{x}{\mu}-1)]+\exp(\frac{2\lambda}{\mu})\Phi[-\sqrt{\frac{\lambda}{\mu}}(\frac{x}{\mu}+1)], \quad x>0 . \tag{2.1}$$

The characteristic function of X is given by

$$C_X(t)=\exp[\frac{\lambda}{\mu}(1-(1-\frac{2it\mu^2}{\lambda})^{\frac{1}{2}})], \tag{2.2}$$

and the moment generating function can be obtained from $C_X(t)$, i.e.,

$$M_X(t)=\mu^t\sum_{j=0}^{t-1}\frac{(t-1-j)!}{j!(t-1-j)!}(\frac{2\lambda}{\mu})^{-t}. \tag{2.3}$$

Differentiating the moment generating function with respect to t for the first and second order at $t=0$, we can obtain the mean and the variance of X as $E(X)=\mu$ and $\mathrm{var}(X)=\frac{\mu^3}{\lambda}$, respectively.

For a random sample $X_1, X_2, ..., X_n$ 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

$W = \frac{1}{n-1}\sum_{i=1}^{n}(\frac{1}{X_i} - \frac{1}{\bar{X}})$, respectively, and the minimum sufficient statistics of (μ, λ) are $(T_1,\ T_2)$, where $T_1 = \sum_{i=1}^{n} X_i$ and $T_2 = \sum_{i=1}^{n} \frac{1}{X_i}$. It is worthy to notice that

$$\bar{X} \sim \mathrm{IG}(\mu, n\lambda) \quad \text{and} \quad (n-1)W \sim \frac{1}{\lambda}\chi^2_{n-1}, \tag{2.4}$$

and that these two statistics are independently distributed. The proof can be found in Chhikara and Folks (1989).

Remark 2. Let $X \sim \mathrm{IG}(\mu, \lambda)$ 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. (Chhikara and Folks, 1989)

From (2.4) and Remark 2, 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

$$U = \frac{\sqrt{n}(\bar{X}-\mu)}{\mu(\bar{X}W)^{1/2}} \tag{2.5}$$

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 $\underset{\sim}{X} = (X_1, \ldots, X_n)$ be an independent sample from some distribution and $l(\theta) = l(\theta; \underset{\sim}{X} = \underset{\sim}{x})$ be the log-likelihood function based on the sample data. Suppose θ is the p-dimensional vector 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 $r(\mu)$ for inference on μ is defined as

$$r(\mu)=\text{sgn}(\hat{\mu}-\mu)\{2[l(\hat{\theta})-l(\hat{\theta}_{\mu})]\}^{1/2}, \tag{2.6}$$

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 (1974) 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

$$\{\mu:|r(\mu)|\le z_{\alpha/2}\}, \tag{2.7}$$

where $z_{\alpha/2}$ is the $100(1-\alpha/2)^{\text{th}}$ percentile of the standard normal distribution. Since the signed log-likelihood ratio statistic is quite inaccurate when the sample size is small, Barndorff-Nielsen (1986, 1991) proposed a higher order likelihood-based method which is known as the modified signed log-likelihood ratio,

$$r^{*}(\mu)=r(\mu)+r(\mu)^{-1}\log\{\frac{q(\mu)}{r(\mu)}\}, \tag{2.8}$$

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. For most of the conditions, $q(\mu)$ is not easy to obtain. Thomas (1999) presented an approximation to $q(\mu)$ with error of $O(n^{-1})$ and hence $r^{*}(\mu)$. A widely applicable formula for $q(\mu)$ that ensures the $O(n^{-3/2})$ accuracy provided by Fraser et al. (1999) is defined as

$$q(\mu)=\frac{\left|l_{;V}(\hat{\theta})-l_{;V}(\hat{\theta}_{\mu})\quad l_{\lambda;V}(\hat{\theta}_{\mu})\right|}{\left|l_{\theta;V}(\hat{\theta})\right|}\left\{\frac{\left|j_{\theta\theta}(\hat{\theta})\right|}{\left|j_{\lambda\lambda}(\hat{\theta}_{\mu})\right|}\right\}^{1/2}, \tag{2.9}$$

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 and

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

The vector array $V = (v_1', \dots, v_p')$ in (2.9) where $v_i = \{v_{i,1}, \dots, v_{i,n}\}$, $i = 1, \dots, p$, is obtained from a vector pivotal quantity $R(\underset{\sim}{x};\theta) = (R_1(x_1;\theta), \dots, R_n(x_n;\theta))$ by

$$V = -\left(\frac{\partial R(x;\theta)}{\partial x}\right)^{-1}\left(\frac{\partial R(x;\theta)}{\partial \theta}\right)\bigg|_{\hat{\theta}}, \tag{2.10}$$

where the distribution of $R_i(x_i;\theta)$ is free of the parameters. The choice of the pivotal quantity will be briefly discussed in Remark 4. The quantity $l_{;V}(\theta)$ is the likelihood gradient with

$$l_{;V}(\theta)' = \{\frac{d}{dv_1} l(\theta;\underset{\sim}{x}), \dots, \frac{d}{dv_p} l(\theta;\underset{\sim}{x})\}, \tag{2.11}$$

where $\frac{d}{dv_i} l(\theta, x) = \sum_{j=1}^{n} l_{;x_j}(\theta) \cdot v_{i,j} = \sum_{j=1}^{n} \frac{\partial l(\theta)}{\partial x_j} \cdot v_{ij}$, $i = 1, \dots, p$, $j = 1, \dots, n$.

Note that r^* achieves third-order accuracy to a standard normal distribution (Fraser et al. 1999). Therefore, a $100(1-\alpha)\%$ confidence interval for μ based on $r^*(\mu)$ is given by

$$\{\mu : |r^*(\mu)| \le z_{\alpha/2}\}. \tag{2.12}$$

Remark 3. Notice that the procedure we mention above is performed for one population. If the inference problem involves I independent populations, some modifications are needed in applying this method. First, we put all the observations from the I distinct random samples together. Denote the set of observations by $\underset{\sim}{X}$, where $\underset{\sim}{X} = (X_{11}, \dots, X_{1n_1}, X_{21}, \dots, X_{2n_2}, \dots, X_{I1}, \dots, X_{In_I})$. For each component in $\underset{\sim}{X}$, we construct a corresponding pivotal quantity, $R_{ij}(X_{ij}, \theta)$, $i = 1, \dots, I$, $j = 1, \dots. n_i$, then $q(\mu)$ and $r^*(\mu)$ can be constructed in similar manner as we mentioned above.

Remark 4. Explicit specification of the pivotal quantity $R_{ij}(x_{ij}, \theta)$ is not needed for the computation of $q(\mu)$ (Fraser et al. 1999). A simple and easy choice is given by the distribution function which is uniformly distributed. In fact, the choice of the pivotal quantity has crucial impact on the computation of the modified signed log-likelihood ratio algorithmatically. A choice which contains more information about the sample is preferred.

Remark 5. $q(\mu)$ in (2.9) can be expressed in various forms. It is different from the quantity

$$q(\mu) = \frac{\left| l_{;\hat{\theta}}(\hat{\theta}) - l_{;\hat{\theta}}(\hat{\theta}_\mu) \quad l_{\lambda;\hat{\theta}}(\hat{\theta}_\mu) \right|}{\left| l_{\theta;\hat{\theta}}(\hat{\theta}) \right|} \left\{ \frac{\left| j_{\theta\theta}(\hat{\theta}) \right|}{\left| j_{\lambda\lambda}(\hat{\theta}_\mu) \right|} \right\}^{1/2}, \tag{2.13}$$

where

$$l_{;\hat{\theta}}(\hat{\theta}) = \left. \frac{\partial l(\theta)}{\partial \hat{\theta}} \right|_{\theta=\hat{\theta}}, \quad l_{\lambda;\hat{\theta}}(\hat{\theta}_\mu) = \left. \frac{\partial l_{;\hat{\theta}}(\theta)}{\partial \lambda} \right|_{\theta=\hat{\theta}_\mu}, \quad l_{\theta;\hat{\theta}}(\hat{\theta}) = \left. \frac{\partial l_{;\hat{\theta}}(\theta)}{\partial \theta} \right|_{\theta=\hat{\theta}}$$

given by Barndorff-Nielsen (1991) and Barndorff-Nielsen & Cox (1994) for computing $r^*(\mu)$. Both variants of $r^*(\mu)$ can reach third-order accuracy, although $r^*(\mu)$ can be obtained through different forms. The expression in (2.13) involves differentiation with respect to $\hat{\theta}$, if there is no analytic form of MLEs, then sometimes such $q(\mu)$ is difficult or impossible to obtain. For example, as for our problem, the inference for the common mean of I independent IG populations, there are $I+1$ parameters and $2I$ minimal sufficient statistics which is the so-called $(2I, I+1)$ curved exponential model, $q(\mu)$ in (2.13) is not easy to apply since the MLEs do not have closed form, so the computation of the derivative with respect to $\hat{\theta}$ is not available. On the other hand, $q(\mu)$ in (2.9) is easy to implement algorithmically. Such a $q(\mu)$ is quite flexible; hence we will use it to perform the higher order and likelihood-based inferences in the following section.

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

In this section we apply the procedure we present in section 2 to construct the confidence interval. The derivation for the general case will be presented in section 3.1. The two independent IG populations case will be discussed in section 3.2. At last, we derive the t-like confidence interval under the assumption of homogeneity for comparison purpose. All the results will be utilized in the simulation in the next section.

3.1 The likelihood-based confidence interval in the general case

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

$$l(\theta; X_1 = x_1,..., X_I = x_I)$$
$$= \frac{1}{2}\sum_{i=1}^{I} n_i \log\frac{\lambda_i}{2\pi} - \frac{3}{2}\sum_{i=1}^{I}\sum_{j=1}^{n_i}\log x_{ij} - \frac{1}{2\mu^2}\sum_{i=1}^{I}\sum_{j=1}^{n_i}\lambda_i x_{ij} + \frac{1}{\mu}\sum_{i=1}^{I} n_i\lambda_i - \frac{1}{2}\sum_{i=1}^{I}\sum_{j=1}^{n_i}\frac{\lambda_i}{x_{ij}}. \qquad (3.1)$$

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

$$\frac{\partial l(\theta)}{\partial \mu} = \frac{-1}{\mu^2}\sum_{i=1}^{I} n_i\lambda_i + \frac{1}{\mu^3}\sum_{i=1}^{I}\sum_{j=1}^{n_i}\frac{\lambda_i}{x_{ij}^2}$$
$$\frac{\partial l(\theta)}{\partial \lambda_i} = \frac{n_i}{2\lambda_i} + \frac{n_i}{\mu} - \frac{1}{2}\sum_{j=1}^{n_i} x_{ij}^2 - \frac{1}{2\mu^2}\sum_{j=1}^{n_i}\frac{1}{x_{ij}^2}\ ,\ i = 1,..., I. \qquad (3.2)$$

The overall MLEs $\hat{\theta} = (\hat{\mu}, \hat{\lambda}_1,..., \hat{\lambda}_I)$ can be uniquely obtained by solving the non-linear system (3.2) simultaneously. It seems that in our problem, there is no analytic form for the overall MLEs. Therefore, some numerical facility is needed to get the numerical solutions. Furthermore, the constrained MLEs $\hat{\theta}_\mu = (\mu, \hat{\lambda}_{\mu,1}, ..., \hat{\lambda}_{\mu,I})$ for a given μ are

$$\hat{\lambda}_{\mu,i} = \frac{-n_i \mu^2}{(2n_1\mu - \sum_{i=1}^{n_i} x_i - \mu^2 \sum_{i=1}^{n_i} \frac{1}{x_i})}, \quad i = 1, ..., I \, . \tag{3.3}$$

Choosing a vector of pivotal quantity $R = \{R_{11}, ..., R_{In_I}\}$ with $R_{ij} = \frac{\lambda_i (x_{ij} - \mu)^2}{\mu^2 x_{ij}}$, $i = 1,..,I; j = 1,..,n_i$, 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

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

$$\frac{\partial R_{ij}}{\partial \mu} = \frac{-2\lambda_i (x_{ij} - \mu)}{\mu^3} \, ;$$

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

Thus

$$\left(\frac{\partial R}{\partial x}\right)^{-1} = diag\left[\frac{\mu^2 x_{11}^2}{\lambda_1 (x_{11}^2 - \mu^2)}, ..., \frac{\mu^2 x_{1n_1}^2}{\lambda_1 (x_{1n_1}^2 - \mu^2)}, ..., \frac{\mu^2 x_{I1}^2}{\lambda_I (x_{I1}^2 - \mu^2)}, ..., \frac{\mu^2 x_{In_I}^2}{\lambda_I (x_{In_I}^2 - \mu^2)}\right] \, , \tag{3.5}$$

where diag[.] is the abbreviation of the diagonal matrix.

Furthermore, $\left(\frac{\partial R}{\partial \theta}\right) = \left[\left(\frac{\partial R}{\partial \mu}\right)', \left(\frac{\partial R}{\partial \lambda_1}\right)', ..., \left(\frac{\partial R}{\partial \lambda_I}\right)'\right] = [h_1', ..., h_{I+1}']$ with

$$h_1 = \left(\frac{-2\lambda_1 (x_{11} - \mu)}{\mu^3}, ..., \frac{-2\lambda_1 (x_{1n_1} - \mu)}{\mu^3}, ..., \frac{-2\lambda_I (x_{I1} - \mu)}{\mu^3}, ..., \frac{-2\lambda_I (x_{In_I} - \mu)}{\mu^3}\right)$$

$$h_{j+1} = (\underbrace{0, ..., 0}_{\sum_{k=1}^{j-1} n_k}, \frac{(x_{j1} - \mu)^2}{\mu^2 x_{j1}}, ..., \frac{(x_{jn_j} - \mu)^2}{\mu^2 x_{jn_j}}, 0...., 0) \, , j = 1, ..., I \, . \tag{3.6}$$

Note that V is a vector array and use (3.5) and (3.6), V can be expressed as follows:

$$V=(v_1^{'},\ldots,v_{I+1}^{'})=-\left(\frac{\partial R}{\partial x}\right)^{-1}\left(\frac{\partial R}{\partial \theta}\right)\Bigg|_{\hat\theta}$$

$$=\begin{bmatrix}
\frac{2x_{11}^2}{\hat\mu(x_{11}+\hat\mu)} & -\frac{x_{11}(x_{11}-\hat\mu)}{\hat\lambda_1(x_{11}+\hat\mu)} & 0 & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\frac{2x_{1n_1}^2}{\hat\mu(x_{1n_1}+\hat\mu)} & -\frac{x_{1n_1}(x_{1n_1}-\hat\mu)}{\hat\lambda_1(x_{1n_1}+\hat\mu)} & 0 & \cdots & 0 & 0 \\
\vdots & 0 & -\frac{x_{21}(x_{21}-\hat\mu)}{\hat\lambda_2(x_{21}+\hat\mu)} & \ddots & \vdots & \vdots \\
\vdots & \vdots & \vdots & \ddots & 0 & 0 \\
\vdots & 0 & -\frac{x_{2n_2}(x_{2n_2}-\hat\mu)}{\hat\lambda_2(x_{2n_2}+\hat\mu)} & \vdots & -\frac{x_{I-1,1}(x_{I-1,1}-\hat\mu)}{\hat\lambda_{I-1}(x_{I-1,1}+\hat\mu)} & 0 \\
\vdots & 0 & 0 & \ddots & \vdots & \vdots \\
\vdots & \vdots & \vdots & \ddots & -\frac{x_{I-1,n_{I-1}}(x_{I-1,n_{I-1}}-\hat\mu)}{\hat\lambda_{I-1}(x_{I-1,n_{I-1}}+\hat\mu)} & \vdots \\
\vdots & 0 & 0 & \vdots & 0 & 0 \\
\frac{2x_{I1}^2}{\hat\mu(x_{I1}+\hat\mu)} & 0 & 0 & \cdots & \vdots & -\frac{x_{I1}(x_{I1}-\hat\mu)}{\hat\lambda_I(x_{I1}+\hat\mu)} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\frac{2x_{In_I}^2}{\hat\mu(x_{In_I}+\hat\mu)} & 0 & 0 & \cdots & 0 & -\frac{x_{In_I}(x_{In_I}-\hat\mu)}{\hat\lambda_I(x_{In_I}+\hat\mu)}
\end{bmatrix}.$$

It remains to obtain the likelihood gradients, $l_{;V}(\theta)$, $l_{\lambda;V}(\theta)$, $l_{\theta;V}(\theta)$, and the fisher information matrices $j_{\theta\theta}(\theta)$ and $j_{\lambda\lambda}(\theta)$. Notice again that

$$l_{;V}(\theta)^{'}=\frac{\partial l(\theta)}{\partial V}=\left[\frac{\partial l(\theta)}{\partial v_1},\ldots,\frac{\partial l(\theta)}{\partial v_{I+1}}\right],$$

$$l_{\lambda;V}(\theta)^{'}=\frac{\partial l_{;V}(\theta)}{\partial \lambda}=\left[\frac{\partial l_{;V}(\theta)}{\partial \lambda_1},\ldots,\frac{\partial l_{;V}(\theta)}{\partial \lambda_I}\right],$$

$$l_{\theta;V}(\hat\theta)^{'}=\frac{\partial l_{;V}(\theta)}{\partial \theta}=\left[\frac{\partial l_{;V}(\theta)}{\partial \mu},\frac{\partial l_{;V}(\theta)}{\partial \lambda_1},\ldots,\frac{\partial l_{;V}(\theta)}{\partial \lambda_I}\right],$$

where

$$\frac{\partial l(\theta)}{\partial v_k}=\sum_{i=1}^{I}\sum_{j=1}^{n_i} l_{;x_{ij}}(\theta)v_{k,j+(i-1)\times n_{i-1}},\ k=1,\ldots,p,$$

$$\frac{\partial l_{;V}(\theta)}{\partial \mu}=\left[\sum_{i=1}^{I}\sum_{j=1}^{n_i} l_{\mu;x_{ij}}(\theta)v_{1,j+(i-1)\times n_{i-1}},\ldots,\sum_{i=1}^{I}\sum_{j=1}^{n_i} l_{\mu;x_{ij}}(\theta)v_{I+1,j+(i-1)\times n_{i-1}}\right],$$

$$\frac{\partial l_{;V}(\theta)}{\partial \lambda_m}=\left[\sum_{i=1}^{I}\sum_{j=1}^{n_i} l_{\lambda_m;x_{ij}}(\theta)v_{1,j+(i-1)\times n_{i-1}},\ldots,\sum_{i=1}^{I}\sum_{j=1}^{n_i} l_{\lambda_m;x_{ij}}(\theta)v_{I+1,j+(i-1)\times n_{i-1}}\right],\ m=1,\ldots,I.$$

Therefore the complete results are

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

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

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

The observed information matrix and the observed nuisance information matrix can be calculated by multiplying the Hessian matrix for the log-likelihood function by (-1) . The results are given below.

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

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 (2.8) and (2.9),

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

and then $r^*(\mu)=r(\mu)+r(\mu)^{-1}\log\{\frac{q(\mu)}{r(\mu)}\}$ can be obtained.

Although the values of r and r^* 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 thesis we use the so-called secant method (or the modified Newton-Raphson method) to obtain the confidence limit; 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

$$\mu_1=\mu_0+\frac{[Z_{\alpha/2}-r(\mu_0)]}{[r(\mu_0+\delta)-r(\mu_0-\delta)]/2\delta} \tag{3.7}$$

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 (3.7), 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^* can be obtained by substituting r^* for r in (3.7).

3.2 The likelihood-based confidence interval when *I=2*

For the purpose of illustration, we present the derivation of the confidence interval for the common mean of two independent IG populations. Let $\underset{\sim}{X}_1 = X_{11},\ldots,X_{1n_1} \sim \text{IG}(\mu,\lambda_1)$ and $\underset{\sim}{X}_2 = X_{21},\ldots,X_{2n_2} \sim \text{IG}(\mu,\lambda_2)$ be two independent inverse Gaussian samples. The log-likelihood function based on the observations is

$$l(\theta;\underset{\sim}{X}_1=\underset{\sim}{x}_1,\underset{\sim}{X}_2=\underset{\sim}{x}_2)=\frac{n_1}{2}\log\frac{\lambda_1}{2\pi}-\frac{3}{2}\sum_{i=1}^{n_1}\log x_{1i}-\frac{\lambda_1}{2\mu^2}\sum_{i=1}^{n_1}x_{1i}+\frac{\lambda_1 n_1}{\mu}-\frac{\lambda_1}{2}\sum_{i=1}^{n_1}\frac{1}{x_{1i}}+\frac{n_2}{2}\log\frac{\lambda_2}{2\pi}$$
$$-\frac{3}{2}\sum_{i=1}^{n_2}\log x_{2i}-\frac{\lambda_2}{2\mu^2}\sum_{i=1}^{n_2}x_{2i}+\frac{\lambda_2 n_2}{\mu}-\frac{\lambda_2}{2}\sum_{i=1}^{n_2}\frac{1}{x_{2i}}. \tag{3.8}$$

The MLEs and the constrained MLEs are $\hat{\theta}=(\hat{\mu},\hat{\lambda}_1,\hat{\lambda}_2)$ and $\hat{\theta}_\mu=(\mu,\hat{\lambda}_{1\mu},\hat{\lambda}_{2\mu})$, respectively .

Take $R_{ij}=\dfrac{\lambda_i(x_{ij}-\mu)^2}{\mu^2 x_{ij}}$ to be the pivotal quantity as we mentioned earlier and differentiate R_{ij} with respect to x and θ, we then have

$$\left(\frac{\partial R}{\partial x}\right)^{-1}=\begin{bmatrix}\frac{\mu^2x_{11}^2}{\lambda_1(x_{11}^2-\mu^2)} & & & & & 0\\ & \ddots & & & & \\ & & \frac{\mu^2x_{1n_1}^2}{\lambda_1(x_{1n_1}^2-\mu^2)} & & & \\ & & & \frac{\mu^2x_{21}^2}{\lambda_2(x_{21}^2-\mu^2)} & & \\ & 0 & & & \ddots & \\ & & & & & \frac{\mu^2x_{2n_2}^2}{\lambda_2(x_{2n_2}^2-\mu^2)}\end{bmatrix}$$

and

$$
\left(\frac{\partial z}{\partial \theta}\right)=\begin{bmatrix}
\frac{-2\lambda_1(x_{11}-\mu)}{\mu^3} & \frac{(x_{11}-\mu)^2}{\mu^2 x_{11}} & 0 \\
\vdots & \vdots & \vdots \\
\frac{-2\lambda_1(x_{1n_1}-\mu)}{\mu^3} & \frac{(x_{1n_1}-\mu)^2}{\mu^2 x_{1n_1}} & 0 \\
\frac{-2\lambda_2(x_{21}-\mu)}{\mu^3} & 0 & \frac{(x_{21}-\mu)^2}{\mu^2 x_{21}} \\
\vdots & \vdots & \vdots \\
\frac{-2\lambda_2(x_{2n_2}-\mu)}{\mu^3} & 0 & \frac{(x_{2n_2}-\mu)^2}{\mu^2 x_{2n_2}}
\end{bmatrix}.
$$

So $V=(v_1^{'},v_2^{'},v_3^{'})=-\left(\frac{\partial R}{\partial x}\right)^{-1}\left(\frac{\partial R}{\partial \theta}\right)\Big|_{\hat{\theta}}$

$$
=\begin{bmatrix}
\frac{2x_{11}^2}{\hat{\mu}(x_{11}+\hat{\mu})} & -\frac{x_{11}(x_{11}-\hat{\mu})}{\hat{\lambda}_1(x_{11}+\hat{\mu})} & 0 \\
\vdots & \vdots & \vdots \\
\frac{2x_{1n_1}^2}{\hat{\mu}(x_{1n_1}+\hat{\mu})} & -\frac{x_{1n_1}(x_{1n_1}-\hat{\mu})}{\hat{\lambda}_1(x_{1n_1}+\hat{\mu})} & 0 \\
\frac{2x_{21}^2}{\hat{\mu}(x_{21}+\hat{\mu})} & 0 & -\frac{x_{21}(x_{21}-\hat{\mu})}{\hat{\lambda}_2(x_{21}+\hat{\mu})} \\
\vdots & \vdots & \vdots \\
\frac{2x_{2n_2}^2}{\hat{\mu}(x_{2n_2}+\hat{\mu})} & 0 & -\frac{x_{2n_2}(x_{2n_2}-\hat{\mu})}{\hat{\lambda}_2(x_{2n_2}+\hat{\mu})}
\end{bmatrix}.
$$

Moreover, the likelihood gradients are

$$
l_{;V}(\theta)=\begin{bmatrix}
\sum_{i=1}^{n_1}\frac{2x_{1i}^2}{\hat{\mu}(x_{1i}+\hat{\mu})}\cdot\left(\frac{\hat{\lambda}_1}{2x_{1i}^2}-\frac{\hat{\lambda}_1}{2\hat{\mu}^2}-\frac{3}{2x_{1i}}\right)+\sum_{i=1}^{n_2}\frac{2x_{2i}^2}{\hat{\mu}(x_{2i}+\hat{\mu})}\cdot\left(\frac{\hat{\lambda}_2}{2x_{2i}^2}-\frac{\hat{\lambda}_2}{2\hat{\mu}^2}-\frac{3}{2x_{2i}}\right) \\
\sum_{i=1}^{n_1}\frac{x_{1i}(\hat{\mu}-x_{1i})}{\hat{\lambda}_1(x_{1i}+\hat{\mu})}\cdot\left(\frac{\hat{\lambda}_1}{2x_{1i}^2}-\frac{\hat{\lambda}_1}{2\hat{\mu}^2}-\frac{3}{2x_{1i}}\right) \\
\sum_{i=1}^{n_2}\frac{x_{2i}(\hat{\mu}-x_{2i})}{\hat{\lambda}_2(x_{2i}+\hat{\mu})}\cdot\left(\frac{\hat{\lambda}_2}{2x_{2i}^2}-\frac{\hat{\lambda}_2}{2\hat{\mu}^2}-\frac{3}{2x_{2i}}\right)
\end{bmatrix},
$$

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

and

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

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

$$j_{\lambda\lambda}(\theta)=\begin{bmatrix} \frac{n_1}{2\lambda_1^2} & 0 \\ 0 & \frac{n_2}{2\lambda_2^2} \end{bmatrix},$$

where $s_1=\sum_{i=1}^{n_1}X_{1i}$ and $s_2=\sum_{i=1}^{n_2}X_{2i}$. Finally, a $100(1-\alpha)\%$ confidence interval of μ based on $r^*(\mu)$ is then obtained by applying these quantities to (2.8) and (2.12).

3.3 Simple t-test confidence interval

For the purpose of comparison, we present 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 $X_i = (X_{i1}, \ldots, X_{in_i}),\ i = 1, 2, .., I,$ are I independent populations with parameters (μ, λ_i) for each population, from (2.4) we know 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 2, 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, a two-sided $100(1-\alpha)\%$ for μ can be obtained by solving the following inequality

$$\left\{\mu : \left|\frac{\sqrt{N}(\bar{X}-\mu)}{\mu(\bar{X}W)^{1/2}}\right| < t_{1-\frac{\alpha}{2}, N-I}\right\}$$

$$= \left\{\mu : \left|\frac{\sqrt{N}\left(\frac{\bar{X}}{\mu} - 1\right)}{(\bar{X}W)^{1/2}}\right| < t_{1-\frac{\alpha}{2}, N-I}\right\}.$$

The confidence interval is summarized as

$$\begin{cases} \left[\bar{X}\left(1 + t_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{X}W}{N}}\right)^{-1},\ \bar{X}\left(1 - t_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{X}W}{N}}\right)^{-1}\right] & \text{, if } 1 - t_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{X}W}{N}} > 0 \\ \left[\bar{X}\left(1 + t_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{X}W}{N}}\right)^{-1},\ \infty\right] & \text{, otherwise} \end{cases}. \qquad (3.9)$$

4 . Simulation studies and numerical examples

In this section, we first present a simulation to show the normal approximation of r and r^*. In section 4.2, we conduct a simulation study of the proposed procedure for different parameter configurations. In section 4.3, a real-life data and a simulated data are given as illustrations.

4.1 Normal approximation for r and r^*

To show that the normal approximations for r and r^* are adequate, we conduct a simulation to show the validity and also to compare the extent of asymptotic of r and r^*. We present the Q-Q plots for both r^* and r under the same parameter settings. For two populations, the sample sizes $(n_1, n_2) = (5,10)$, (10, 5) and (10,10) and $(\mu, \lambda_1, \lambda_2) =$ (1,0.2,1) are demonstrated. And the sample sizes $(n_1, n_2, n_3) = (5,8,10)$, (5,10,8) and (10,8,5) and $(\mu, \lambda_1, \lambda_2, \lambda_3) = (1, 0.1, 0.5, 1)$ are chosen for three populations. All the simulations are based on 5,000 repetitions. For two populations, the Q-Q plots are given in Fig. 4.1 and for three populations, the Q-Q plots are given in Fig. 4.2. Both Fig. 4.1 and Fig. 4.2 show that the asymptotic of r^* is better than r which is consistent as we expect. In these figures, we can see, most of the tail of the Q-Q plots of r deviate off the straight line which is the indicator of normality whereas those of r^* seem to overlap the straight line for most of the part. In short, r^* is more accurate in approximating to the standard normal.

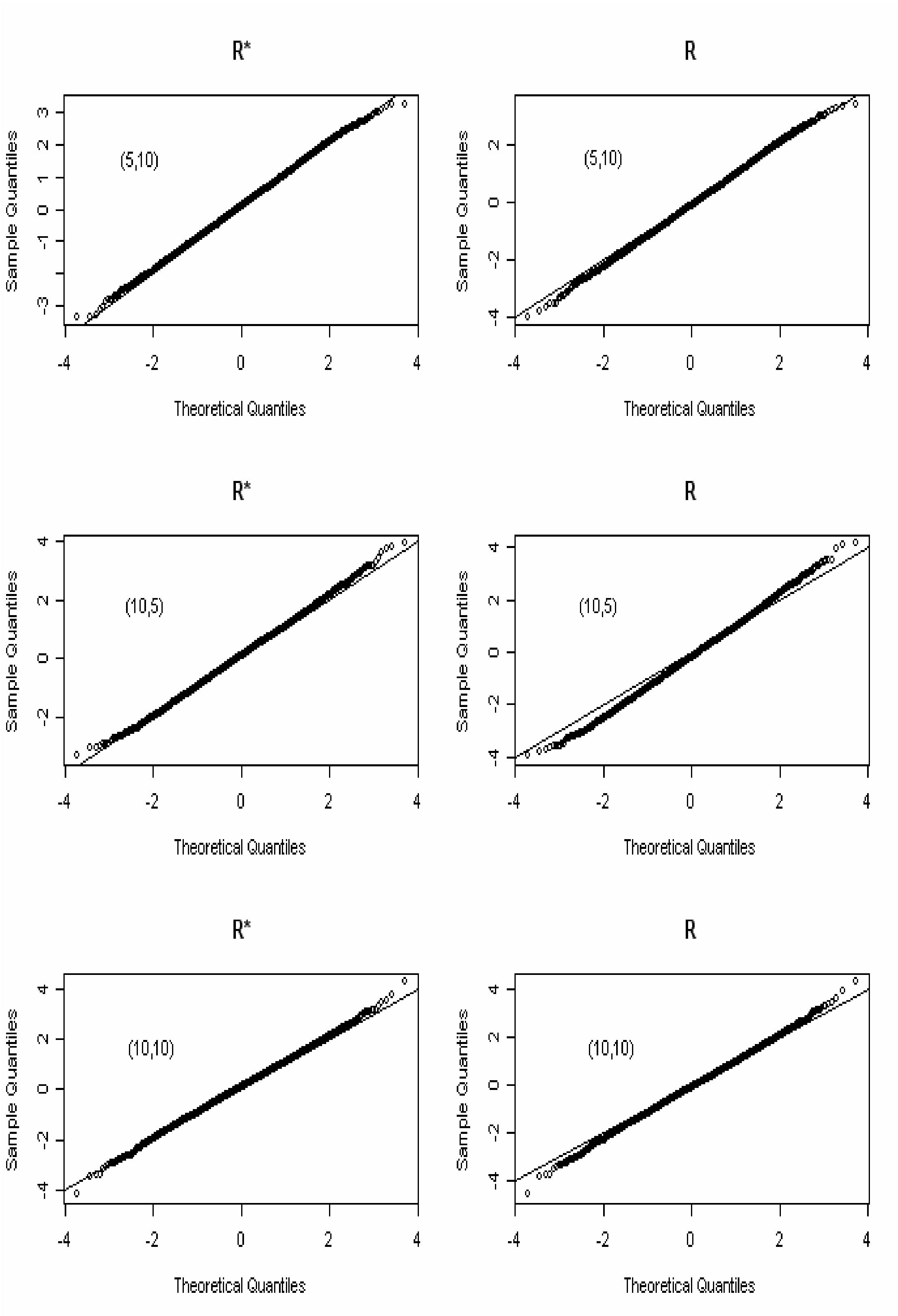

Fig. 4.1 *Q-Q plots for two populations at* $(\mu, \lambda_1, \lambda_2) = (1,\ 0.2,\ 1)$

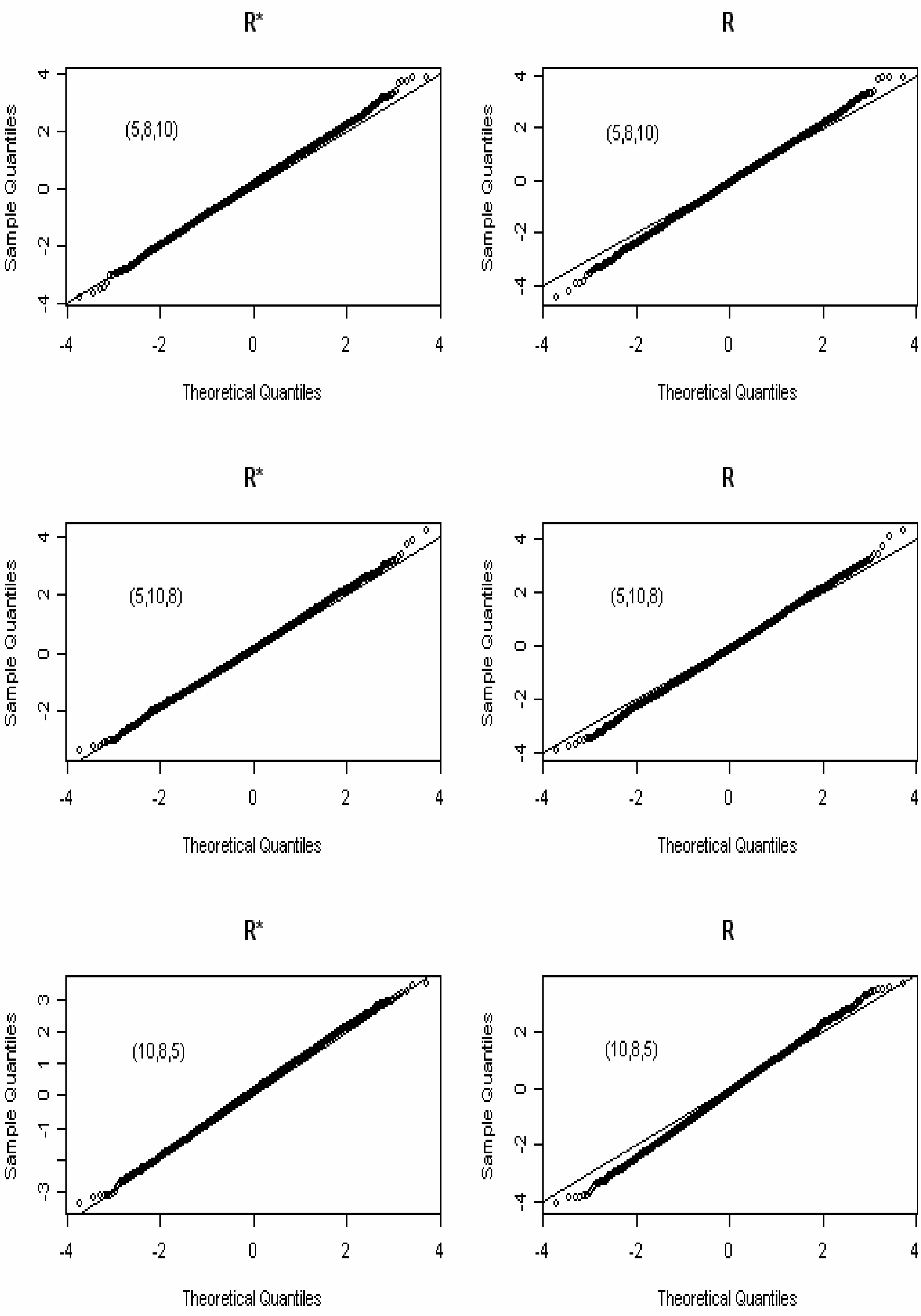

Fig. 4.2 *Q-Q plots for three populations at* $(\mu, \lambda_1, \lambda_2, \lambda_3) = (1,\ 0.1,\ 0.5,\ 1)$

4.2 Simulation studies

To evaluate the accuracy of the proposed method, we present simulation studies of the confidence interval and type I errors applied to a variety of scale parameter configurations and different settings of sample size for two and three populations. We exhibit the coverage probabilities, the average lengths of the 95% confidence intervals and calculate type I errors based on r, r^* and simple t-test method. The results given in tables 1-4 below are based on 10,000 simulation runs for each combination. From the table 1 and table 2, we see that although the confidence intervals based on the directed log-likelihood ratio method, r, have shortest average lengths comparing to the other two methods, the coverage probabilities are too short to attain the proposed coverage probabilities in each combination. The confidence intervals based on the simple t-test method also show a good performance on coverage probabilities, but these coverage probabilities decrease when the heterogeneity increases. Moreover, when the scale parameter is small related to μ, then the intervals constructed by the simple t-test are unbounded (i.e., a one-sided interval). In these cases, the method gives less information about the target value than those based on r^* and r. On the other hand, the confidence intervals based on higher order likelihood-based method, r^*, not only have almost exact coverage probabilities in each combination (except for few pairs having large sample sizes with relatively small scale parameters), but the average lengths are also quite acceptable. Therefore, for the overall comparisons, the higher order likelihood-based method outperforms the other two methods.

Table 1

Simulation results of 95% confidence interval of $\mu = 1$ for two populations

(n_1, n_2)	λ_1	λ_2	r^*		r		simple t-test	
			CP	Length	CP	Length	CP	Length
(5,10)	0.2	1	0.951	9.387	0.923	1.637	0.954	∞
	0.5	1	0.948	14.340	0.917	1.554	0.945	∞
	1	3	0.952	1.054	0.926	0.786	0.943	1.024
	3	10	0.947	0.456	0.920	0.387	0.939	0.486
	1	10	0.948	0.477	0.923	0.409	0.933	0.791
(10,5)	0.2	1	0.931	23.093	0.847	1.368	0.955	∞
	0.5	1	0.944	13.672	0.897	1.534	0.949	∞
	1	3	0.949	1.958	0.906	0.967	0.950	1.506
	3	10	0.948	0.598	0.903	0.464	0.949	0.623
	1	10	0.947	0.840	0.925	0.573	0.951	1.361
(10,10)	0.2	1	0.951	7.914	0.929	1.659	0.958	∞
	0.5	1	0.955	2.320	0.933	1.363	0.948	∞
	1	3	0.945	0.873	0.924	0.708	0.946	0.918
	3	10	0.945	0.410	0.921	0.360	0.947	0.456
	1	10	0.949	0.459	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^*		r		simple t-test	
				CP	Length	CP	Length	CP	Length
(5,8,10)	0.1	0.1	1	0.949	4.865	0.923	1.611	0.965	∞
	0.1	0.5	1	0.949	3.382	0.922	1.456	0.959	∞
	1	1	5	0.948	0.671	0.923	0.542	0.948	0.807
	1	1	10	0.949	0.467	0.925	0.396	0.943	0.766
	1	5	10	0.948	0.393	0.921	0.337	0.938	0.511
(5,10,8)	0.1	0.1	1	0.952	5.964	0.917	1.589	0.963	∞
	0.1	0.5	1	0.948	3.784	0.918	1.492	0.952	∞
	1	1	5	0.945	0.754	0.919	0.586	0.951	0.873
	1	1	10	0.950	0.537	0.919	0.435	0.947	0.844
	1	5	10	0.945	0.413	0.917	0.350	0.938	0.524
(10,8,5)	0.1	0.1	1	0.928	7.016	0.839	1.306	0.967	∞
	0.1	0.5	1	0.946	5.952	0.889	1.467	0.965	∞
	1	1	5	0.943	0.957	0.901	0.668	0.950	0.974
	1	1	10	0.945	0.720	0.898	0.518	0.953	0.956
	1	5	10	0.946	0.521	0.902	0.406	0.948	0.711

CP : Coverage probability; Length : Average length

Furthermore, from tables 3 and 4, we can see the type I errors based on r^* and r are quite stable under different parameter configurations since the type I errors based on the simple t-test method decrease as the mean parameter under the null hypothesis increases. From tables 3 and 4, the type I errors based on r^* are significantly better than that of r since the type I errors based on r are around 0.07 to 0.10 which are too large comparing to the nominal level 0.05. On the contrary, the type I errors based on r^* are not only stable, but the values also close to the nominal level 0.05. Thus, we can say that the proposed procedure can deal with heterogeneity among populations and give robust and reliable results under different scenarios.

Table3: Type I errors for $H_0: \mu = \mu_0$ vs. $H_0: \mu \neq \mu_0$ at I=2 and $\alpha = 0.05$

		$n_1 = 5,\ n_2 = 10$					$n_1 = 10,\ n_2 = 5$				
		μ_0					μ_0				
(λ_1, λ_2)		0.2	0.8	1.2	2.0	5.0	0.2	0.8	1.2	2.0	5.0
(0.2, 1)	(1)	0.0522	0.0528	0.0529	0.0493	0.0506	0.055	0.0495	0.0567	0.0551	0.0524
	(2)	0.0774	0.0746	0.075	0.0686	0.0702	0.1027	0.0903	0.0989	0.0912	0.0912
	(3)	0.0615	0.0430	0.0489	0.0426	0.0324	0.0502	0.0378	0.0410	0.0381	0.0313
(0.5,1)	(1)	0.0485	0.0562	0.0569	0.0533	0.0575	0.0528	0.056	0.0522	0.0552	0.053
	(2)	0.0832	0.0862	0.0844	0.0813	0.0808	0.0901	0.095	0.0922	0.0922	0.0883
	(3)	0.0555	0.0507	0.0514	0.0504	0.0509	0.0469	0.0515	0.0461	0.0434	0.0484
(1, 3)	(1)	0.0526	0.0528	0.0500	0.0578	0.0564	0.0542	0.0548	0.0573	0.0570	0.0553
	(2)	0.0783	0.0783	0.0744	0.081	0.079	0.0966	0.0985	0.0941	0.0939	0.0939
	(3)	0.0583	0.0545	0.0503	0.0539	0.0524	0.0514	0.0471	0.0464	0.0535	0.0424
(1, 5)	(1)	0.0507	0.0566	0.0517	0.0551	0.057	0.0585	0.0537	0.0558	0.0563	0.0575
	(2)	0.0788	0.0803	0.0768	0.0779	0.079	0.1009	0.0936	0.099	0.098	0.0989
	(3)	0.0634	0.0616	0.061	0.0573	0.0499	0.0502	0.0471	0.0465	0.046	0.0371
(1, 10)	(1)	0.0528	0.055	0.0484	0.0559	0.0487	0.0559	0.0538	0.0517	0.0582	0.0571
	(2)	0.0751	0.0759	0.0727	0.0777	0.0697	0.104	0.0977	0.0967	0.0997	0.0981
	(3)	0.0775	0.074	0.0682	0.0589	0.0478	0.0532	0.0468	0.0507	0.0431	0.0344

The above type I errors due to (1)- $r^*(\mu)$; (2)- $r(\mu)$; (3)- simple t-test.

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

		$n_1 = 5,\ n_2 = 8,\ n_3 = 10$					$n_1 = 5,\ n_2 = 10,\ n_3 = 8$				
		μ_0					μ_0				
$(\lambda_1, \lambda_2, \lambda_3)$		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)	(1)	0.0481	0.0564	0.0542	0.0563	0.0551	0.055	0.0539	0.0516	0.0576	0.0549
	(2)	0.0716	0.0774	0.0748	0.0796	0.0729	0.0845	0.0789	0.0769	0.0814	0.0787
	(3)	0.0483	0.0391	0.0336	0.0258	0.0187	0.0472	0.0355	0.029	0.0247	0.0194
(0.1,0.5,1)	(1)	0.0539	0.0572	0.0528	0.0548	0.0554	0.0516	0.0563	0.0534	0.0570	0.0525
	(2)	0.0812	0.0832	0.0749	0.0761	0.0769	0.0778	0.0841	0.0804	0.0804	0.0791
	(3)	0.0594	0.0462	0.0395	0.0377	0.0283	0.0558	0.0450	0.0422	0.0378	0.0271
(1,1,5)	(1)	0.0520	0.0505	0.0560	0.0567	0.0558	0.0517	0.0525	0.0542	0.0548	0.0577
	(2)	0.0782	0.0755	0.0797	0.0824	0.0768	0.0843	0.0840	0.0839	0.0800	0.0864
	(3)	0.0572	0.0594	0.0557	0.051	0.0436	0.0544	0.0516	0.0524	0.0472	0.0409
(1,1,10)	(1)	0.0485	0.0516	0.0494	0.0564	0.0557	0.0549	0.0541	0.0528	0.0543	0.0581
	(2)	0.0746	0.0724	0.073	0.0792	0.0767	0.0853	0.0852	0.0797	0.0820	0.0864
	(3)	0.0579	0.0534	0.0537	0.0522	0.0394	0.0588	0.0496	0.0484	0.0458	0.0399
(1,5,10)	(1)	0.0517	0.0514	0.0514	0.0537	0.0524	0.0518	0.0534	0.054	0.0557	0.0525
	(2)	0.0779	0.0777	0.0775	0.0815	0.0766	0.08	0.0828	0.0827	0.0865	0.0806
	(3)	0.0638	0.0659	0.0660	0.0553	0.0544	0.0634	0.0625	0.0611	0.0603	0.0477

The above type I errors due to (1)- $r^*(\mu)$; (2)- $r(\mu)$; (3)- simple t-test.

4.3 Two examples

Example 1

We first presented a 3 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 r^*, r and the simple t-test method are given in table 6. Both of the confidence interval based on r^* and r give satisfactory result under the heterogeneous data set when comparing with that based on the simple t-test method. Although the one based on r^* is a little wider than that of r, in general, it gives a better coverage comparing with r.

Table 5

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
r^*	1.221	(0.961, 1.728)
r	1.221	(0.980, 1.605)
simple t -test	1.078	(0.553, 20.711)

Example 2

The second data set is taken from Tweedie (1956) and has been used by Seshadri (1999, p.175) to perform the ANORE. The data set consists of four unbalanced populations and is modeled via the IG distribution. When λ_i's are assumed to be the same for all groups, the P-value is 0.1879 for testing equality of four IG means through the analysis of reciprocals method. Therefore, we can use the proposed method to construct the confidence interval for the common mean parameter. The data and the interval estimation are given in tables 7 and 8. Table 7 shows that when the data are homogeneous, the point estimators and confidence intervals for three methods are quite similar.

Table 7

Population i	1	2	3	4
	8.7	8.5	8.4	8.1
	9.0	8.6	9.0	8.4
	8.4	8.4	8.9	8.5
	8.6	8.3		
		8.8		
$\bar{x}_i$	8.675	8.520	8.767	8.333
w_i	3.049×10^{-4}	2.455×10^{-4}	3.077×10^{-4}	1.856×10^{-4}

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

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

method	Point estimate $\hat{\mu}$	Interval estimate
r^*	8.56	(8.353, 8.789)
r	8.56	(8.407, 8.718)
simple t -test	8.57	(8.440, 8.711)

5. Conclusions

In this thesis, we presented an accurate 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 two alternative methods with respect to their coverage probabilities, average lengths and type I errors. The numerical examples showed that the proposed method gives nearly exact coverage probability and the type I errors are close to the nominal level .05 even for small sample size. The method is able to integrate the information of several non-homogeneous IG populations, and therefore is useful for a variety of practical applications.

References

Barndorff-Nielsen, O.E., 1986. Inference on full and partial parameters based on the standardized signed log likelihood ratio. Biometrika 73, 307-322.

Barndorff-Nielsen, O.E., 1991. Modified signed log-likelihood ratio. Biometrika 78, 557-563.

Barndorff-Nielsen, O.E., Cox, D.R., 1994. Inference and asymptotics. Chapman & Hall, London.

Chhikara, R.S., Folks, J.L., 1989. The inverse Gaussian distribution. Marcel Dekker, New York.

Cox, D.R., Hinkley, D.V., 1974. Theoretical statistics. Chapman & Hall, London.

Fraser, D.A.S., Reid, N., Wu, J., 1999. A simple general formula for tail probabilities for frequentist and Bayesian inference. Biometrika 86, 249-264.

Lancaster, T., 1972. A stochastic model for the duration of a strike. Journal of the Royal Statistical Society A135(2), 257-271.

Mudholkar, G. S., Tian, L., 2002. An entropy characterization of the inverse Gaussian distribution and related goodness-of-fit tests. Journal of Statistical Planning and Inference 102, 211-221.

Reid, N., 1996. Likelihood and higher-order approximations to tail areas: A review and annotated bibliography. Can. J. Statist. 24, 141-166.

Seshadri, V., 1999. The Inverse Gaussian distribution: Statistical Theory and Applications. Springer, New York.

Thomas, A.S., 1999. An empirical adjustment to the likelihood ratio statistic Biometrika 86, 235-247.

Tian, L., Wilding, G.E., 2005. Confidence interval of the ratio of means of two independent inverse Gaussian distributions. Journal of Statistical Planning and Inference 133, 381-386.

Tian, L., 2005. Testing equality of inverse Gaussian means under heterogeneity based on generalized test variable. Computational Statistics & Data Analysis 51, 1156-1165.

Tweedie, M.C.K.,1941. A mathematical investigation of some electrophoretic measurements on colloids. Unpublished M.Sc. Thesis, University of Reading, England.

Tweedie, M.C.K.,1945. Inverse statistical variates. Nature, 155, 453.

Tweedie, M.C.K.,1946. The regression of the sample variance on the sample mean. Journal of London Mathematical Society. 21, 22-28

Tweedie, M.C.K.,1947. Functions of a statistical variate with given means, with special reference to Laplacian distributions. Proceedings of the Cambridge Philosophical Society 43, 41-49

Tweedie, M.C.K., 1956. Some statistical properties of inverse Gaussian distributions. Virginia Journal of Science 7, New Series, 160-165.

Tweedie, M.C.K, 1957. Statistical properties of inverse Gaussian distributions-Ⅰ. Annals of Mathematical Statistics 28, 362-377.

Tweedie, M.C.K, 1957. Statistical properties of inverse Gaussian distributions-Ⅱ. Annals of Mathematical Statistics 28, 696-705.

Wise, M. E., 1971. Skew probability curves with negative powers of time and related to random walks in series. Statistica Neerlandica 25, 159-180.

Wise, M. E., 1975. Skew distributions in biomedicine including some with negative powers of time in statistical distributions in scientific work. Vol. 2, D. Reidel, Dordrecht, Holland.

Wise, M. E., Osborn, S. B., Anderson, J., and Tomlinson, R. W. S., 1968. A stochastic model for turnover of radiocalcium based on the observed laws. Mathematical Biosciences 2, 199-224.

Wu, J., Jiang, G., Wong, A.C, Sun, X., 2002. Likelihood analysis for the ratio of means of two independent log-normal distributions. Biometrics 58(2), 463-469.

Wu, J., Wong, A.C.M., Jiang, G., 2003. Likelihood-based confidence interval for a log-normal mean. Statistics in Medicine 22, 1849-1860.

Wu, J., Wong, A.C.M., 2004. Improved interval estimation for the two-parameter Birnbaum-Saunders distribution. Computational Statistics & Data Analysis 47, 809-821