A NEW COMPUTATIONAL APPROACH FOR TESTING EQUALITY OF INVERSE GAUSSIAN MEANS UNDER HETEROGENEITY

Esra Yigit Gökpınar *, Esra Polat†, Fikri Gökpınar‡, Süleyman Günay§

Received 01:08:2011 : Accepted 19:07:2012

Abstract

In this article, a testing procedure based on computational approach test is proposed for the equality of several inverse Gaussian means under heterogeneity. Not requiring the knowledge of any sampling distribution, depending heavily on numerical computations and Monte Carlo simulation, moreover, figuring out the critical region automatically are the advantages of the computational approach test. We compare it with some of the existing tests; the parametric bootstrap and the generalized test variables in terms of type I risks and powers by using Monte Carlo simulation.

Keywords: Computational approach test, Generalized test variables, Hypothesis testing, Inverse gaussian distribution, Maximum likelihood estimate, Parametric bootstrap.

2000 AMS Classification: 62F03, 65C05, 65C60, 62E17.

1. Introduction

Inverse Gaussian (IG) distribution is given in quite a number of books on stochastic processes and probability. The probability distribution of the first passage time in Brownian motion is given by Schrödinger (1915). Since Tweedie (1945) has shown the inverse relationship between the cumulant generating function of the first passage time distribution and that of the normal distribution, it is called as an IG for the first passage time distribution. Further, as Wald (1947) has derived the limiting form of IG distribution, it is also called as Wald’s distribution, especially in the Russian literature [4].
The density function of the two-parameter IG distribution IG \((\mu, \lambda)\) is defined as in equation (1.1).

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

Here, \(\mu\) is the mean parameter and \(\lambda\) is the scale parameter. For the last three decades, the IG distribution has gained significant attention in describing and analyzing right-skewed data. For example, Doksum and Höyland (1992) developed a model for variable-stress-accelerated life testing experiments based on the IG distributions, Durham and Padgett (1997) used the IG models to develop a new general method based on cumulative damage for describing the failure of a system. Seshadri’s (1999) study is a good reference for other applications of IG distribution in life tests, remote sensing, etc. [11].

The main appeal of IG models lies in these facts: (1) they could accommodate a variety of shapes, from highly skewed to almost normal; (2) they are unique among the distributions for positively right-skewed data (e.g. Weibull, gamma, log normal) on account of its many elegant and convenient properties sharing with Gaussian models [10]. Hence, it could be mentioned that IG distribution has several properties analogous to normal distribution [5, 9, 10].

A very common problem in applied statistics is that of comparing the means of several populations. It is well known that there exists an analysis of variance (ANOVA) F-test for the problem of testing the equality of means from several independent samples under the assumptions of normality. Similarly, for testing equality of IG means when the scale parameters are the same, an analysis of reciprocals (ANORE) F-test is developed by Chhikara and Folks (1989) [7]. Since the number of populations having IG distributions is \(k\), the unknown mean of \(i\)th population is \(\mu_i\) and the general mean is \(\mu\), the null and alternative hypotheses could be written as shown in equation (1.2).

\[
(1.2) \quad H_0: \mu_1 = \mu_2 = \cdots = \mu_k = \mu \text{ and } H_1: \mu_i \neq \mu_j, \quad \exists i \neq j \quad (i, j = 1, \cdots, k)
\]

However, the disadvantage of ANORE is that it fails for testing equality of IG means, when the scale parameters are non-homogeneous. Therefore, to deal with this problem of comparing IG means in the case of unequal scale parameters, Tian (2006) developed an approach using the concepts of generalized test variables (GTV) and generalized p-values. The concept of GTV has been widely applied to a variety of practical settings, where standard inference methods do not exist. The GTV was proposed by Weerahandi (1995) in order to test the equality of normal means under heterogeneity and the generalized p-values approach was introduced by Tsui and Weerahandi (1989) [7]. Since the number of populations having IG distributions is \(k\), the unknown mean of \(i\)th population is \(\mu_i\) and the general mean is \(\mu\), the null and alternative hypotheses could be written as shown in equation (1.2).

\[
H_0: \mu_1 = \mu_2 = \cdots = \mu_k = \mu \text{ and } H_1: \mu_i \neq \mu_j, \quad \exists i \neq j \quad (i, j = 1, \cdots, k)
\]

However, the disadvantage of ANORE is that it fails for testing equality of IG means, when the scale parameters are non-homogeneous. Therefore, to deal with this problem of comparing IG means in the case of unequal scale parameters, Tian (2006) developed an approach using the concepts of generalized test variables (GTV) and generalized p-values. The concept of GTV has been widely applied to a variety of practical settings, where standard inference methods do not exist. The GTV was proposed by Weerahandi (1995) in order to test the equality of normal means under heterogeneity and the generalized p-values approach was introduced by Tsui and Weerahandi (1989) [7]. Since the number of populations having IG distributions is \(k\), the unknown mean of \(i\)th population is \(\mu_i\) and the general mean is \(\mu\), the null and alternative hypotheses could be written as shown in equation (1.2).

\[
H_0: \mu_1 = \mu_2 = \cdots = \mu_k = \mu \text{ and } H_1: \mu_i \neq \mu_j, \quad \exists i \neq j \quad (i, j = 1, \cdots, k)
\]

However, the disadvantage of ANORE is that it fails for testing equality of IG means, when the scale parameters are non-homogeneous. Therefore, to deal with this problem of comparing IG means in the case of unequal scale parameters, Tian (2006) developed an approach using the concepts of generalized test variables (GTV) and generalized p-values. The concept of GTV has been widely applied to a variety of practical settings, where standard inference methods do not exist. The GTV was proposed by Weerahandi (1995) in order to test the equality of normal means under heterogeneity and the generalized p-values approach was introduced by Tsui and Weerahandi (1989) [7]. Since the number of populations having IG distributions is \(k\), the unknown mean of \(i\)th population is \(\mu_i\) and the general mean is \(\mu\), the null and alternative hypotheses could be written as shown in equation (1.2).

\[
H_0: \mu_1 = \mu_2 = \cdots = \mu_k = \mu \text{ and } H_1: \mu_i \neq \mu_j, \quad \exists i \neq j \quad (i, j = 1, \cdots, k)
\]

However, the disadvantage of ANORE is that it fails for testing equality of IG means, when the scale parameters are non-homogeneous. Therefore, to deal with this problem of comparing IG means in the case of unequal scale parameters, Tian (2006) developed an approach using the concepts of generalized test variables (GTV) and generalized p-values. The concept of GTV has been widely applied to a variety of practical settings, where standard inference methods do not exist. The GTV was proposed by Weerahandi (1995) in order to test the equality of normal means under heterogeneity and the generalized p-values approach was introduced by Tsui and Weerahandi (1989) [7]. Since the number of populations having IG distributions is \(k\), the unknown mean of \(i\)th population is \(\mu_i\) and the general mean is \(\mu\), the null and alternative hypotheses could be written as shown in equation (1.2).

\[
H_0: \mu_1 = \mu_2 = \cdots = \mu_k = \mu \text{ and } H_1: \mu_i \neq \mu_j, \quad \exists i \neq j \quad (i, j = 1, \cdots, k)
\]

However, the disadvantage of ANORE is that it fails for testing equality of IG means, when the scale parameters are non-homogeneous. Therefore, to deal with this problem of comparing IG means in the case of unequal scale parameters, Tian (2006) developed an approach using the concepts of generalized test variables (GTV) and generalized p-values. The concept of GTV has been widely applied to a variety of practical settings, where standard inference methods do not exist. The GTV was proposed by Weerahandi (1995) in order to test the equality of normal means under heterogeneity and the generalized p-values approach was introduced by Tsui and Weerahandi (1989) [7]. Since the number of populations having IG distributions is \(k\), the unknown mean of \(i\)th population is \(\mu_i\) and the general mean is \(\mu\), the null and alternative hypotheses could be written as shown in equation (1.2).

\[
H_0: \mu_1 = \mu_2 = \cdots = \mu_k = \mu \text{ and } H_1: \mu_i \neq \mu_j, \quad \exists i \neq j \quad (i, j = 1, \cdots, k)
\]
method for the common mean of several heterogeneous IG populations. The proposed method is based on a higher order likelihood-based procedure.

It is important to develop a test procedure for equality of IG means with satisfactory type I risk regardless of number of groups and the sample sizes. In this paper, a new computational approach test (CAT) is proposed for equality of IG means under heterogeneity. CAT method based on simulation and numerical computations uses the maximum likelihood estimates (MLEs), but does not require any asymptotic distribution. This approach provides an algorithmic framework based on the Monte-Carlo simulation and numerical computations, which can be implemented mechanically by applied researchers to draw statistical inferences, when a suitable parametric model is assumed for a given data set [8]. In this article, extensive simulation results are presented to evaluate the type I risk and power of the proposed CAT approach in comparison to that of the, PB and GTV approaches.

This article is organized as follows. In Section 2, the tests based on the PB and the GTV, used for testing the equality of several IG means, are presented. In Section 3, a new CAT based on simulation and numerical computations, used for the equality of IG means, is presented. In Section 4, simulation results on type I risk control and power are presented. Concluding remarks are summarized in Section 5.

2. Tests for Testing Equality of Inverse Gaussian Means under Heterogeneity

In this section, PB approach and GTV developed for testing equality of several IG means, without the assumption of equal scale parameters, are presented.

2.1. The Parametric Bootstrap Approach. In this section, a PB approach is presented that is used for testing equality of several IG means [7].

Let $X_{i1}, X_{i2}, \ldots, X_{in_i}$ be a random sample from an IG$(\mu_i, \lambda_i), i = 1, \ldots, k$ and the terms in equation (2.1) are defined. It is also well known that equation (2.2) could be written for $i = 1, \ldots, k$.

\begin{align}
(2.1) \quad & \bar{X}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij} \quad \text{and} \quad V_i = \sum_{j=1}^{n_i} \left( \frac{1}{X_{ij}} - \frac{1}{\bar{X}_i} \right) \\
(2.2) \quad & \tilde{X}_i \sim IG(\mu_i, n_i \lambda_i) \quad \text{and} \quad \lambda_i V_i \sim \chi^2_{n_i-1}
\end{align}

The hypotheses of interest are given before as shown in equation (1.2). Then, equation (2.3) is defined with $\hat{\mu} = \frac{\sum_{i=1}^{k} n_i \lambda_i \bar{X}_i}{\sum_{i=1}^{k} n_i \lambda_i}$. Under null hypothesis $H_0$, $Q(\lambda_1, \lambda_2, \ldots, \lambda_k) \sim \chi^2_{k-1}$ [7, 10].

\begin{align}
(2.3) \quad & Q(\lambda_1, \lambda_2, \ldots, \lambda_k) = \sum_{i=1}^{k} n_i \lambda_i \left( \frac{\bar{X}_i}{\hat{\mu}} - 1 \right)
\end{align}

The PB approach involves sampling from the estimated models. Based on the test statistic $Q$ in equation (2.3), the PB pivot variable could be developed as follows. Firstly, equation (2.4) is written with $\hat{\lambda}_i = n_i / V_i$ and $\lambda_{Bi} \sim \frac{\chi^2_{n_i-1}}{V_i}, i = 1, \ldots, k$. Then the PB pivot variable, based on the statistic in equation (2.3), is given by equation (2.5). In equation (2.5), $\hat{\mu} B = \frac{\sum_{i=1}^{k} n_i \lambda_i_{Bi} \bar{X}_{Bi}}{\sum_{i=1}^{k} n_i \lambda_{Bi}}$ [7].

\begin{align}
(2.4) \quad & \bar{X}_{Bi} \sim IG \left( \frac{\sum_{i=1}^{k} n_i \lambda_i \bar{X}_i}{\sum_{i=1}^{k} n_i \lambda_i}, n_i \hat{\lambda}_i \right) \\
(2.5) \quad & Q_B \left( \bar{X}_{B1}, \ldots, \bar{X}_{Bk}, \lambda_{B1}, \ldots, \lambda_{Bk} \mid \bar{X}_1, \ldots, \bar{X}_k, V_1, \ldots, V_k \right)
\end{align}
For a given data set with \((\bar{x}_1, \bar{x}_2, \cdots, \bar{x}_k)\) and \((v_1, v_2, \cdots, v_k)\), \(m\) bootstrap samples, \(Q_B^{(i)}, i = 1, \cdots, m\), are drawn. The PB approach rejects \(H_0\) at significance level \(\alpha\), as shown in equation (2.6). Here, \(Q_{BO}\) is an observed value of \(Q_B\) in equation (2.5), as given in equation (2.7). Here, \(\bar{\mu}\) and \(\bar{\lambda}_i, i = 1, \cdots, k\) are the MLEs under \(H_0\) from the observed samples [7].

\[
P = \frac{\#(Q_B^{(i)} > Q_{BO})}{m} \leq \alpha
\]

(2.6)

\[
Q_{BO} = \sum_{i=1}^{k} n_i \bar{\lambda}_i \left( \frac{1}{\bar{x}_i} - \frac{1}{\bar{\mu}} \right)
\]

(2.7)

The EM-algorithm for calculating \(\bar{\mu}\) and \(\bar{\lambda}_i\) is described as below. The log-likelihood function under null-hypotheses \(H_0\) is as shown in equation (2.8).

\[
(2.8) \quad \log \left( L(\mu, \lambda_1, \cdots, \lambda_k) \right) = \sum_{i=1}^{k} \frac{n_i}{2} \log(\lambda_i) - \frac{1}{2} \sum_{i=1}^{k} \sum_{j=1}^{n_i} \frac{\lambda_i}{2\pi x_{ij}^2} \left( x_{ij} - \mu \right)^2 + const
\]

Here, \(x_{ij}, i = 1, \cdots, k, j = 1, \cdots, n_i\) are observed values of \(X_{ij}\). The MLEs of \(\mu, \lambda_1, \cdots, \lambda_k\) under \(H_0\) have no closed forms. The estimates can be obtained through the iterations as below: updating the estimates from \(L\)-step estimates \((\lambda_1^{(0)}, \cdots, \lambda_k^{(0)}, \mu^{(0)})\) by equation (2.9). Here, initial value \(\mu^{(0)}\) could set as the grand mean \(\frac{1}{\sum_{i=1}^{k} \bar{x}_i, \lambda_1^{(0)}}, \cdots, \lambda_k^{(0)}, \mu^{(0)}\) converge to the MLEs under \(H_0\) denoted as \((\hat{\lambda}_1, \cdots, \hat{\lambda}_k, \hat{\mu})\) [7].

\[
(2.9) \quad \lambda_i^{(l+1)} = \frac{n_i}{\sum_{j=1}^{n_i} \left( \frac{1}{x_{ij}} - \frac{1}{\mu^{(l)}} \right)^2}, i = 1, \cdots, k, \mu^{(l+1)} = \frac{\sum_{i=1}^{k} n_i \lambda_i^{(l+1)} \bar{x}_i}{\sum_{i=1}^{k} n_i \lambda_i^{(l+1)}}
\]

Moreover, the algorithm for calculating p-value using the PB approach could be given as shown below [7]. Algorithm:

1. For a given data \((x_{11}, x_{12}, \cdots, x_{1n_1}), i = 1, \cdots, k\), calculate observed statistics \((\bar{x}_1, \bar{x}_2, \cdots, \bar{x}_k)\) and \((v_1, v_2, \cdots, v_k)\). Then calculate the MLEs \(\hat{\mu}, \hat{\lambda}_1, \cdots, \hat{\lambda}_k\) under null hypotheses. Compute \(Q_{BO}\) as defined in equation (2.7).

2. Independently generate \(X_{Bi} \sim IG\left(\frac{\sum_{i=1}^{k} n_i \lambda_i \bar{x}_i}{\sum_{i=1}^{k} n_i \lambda_i}, n_i \hat{\lambda}_i\right)\) and \(\lambda_{Bi} \sim \frac{\lambda_i^{(n_i-1)} \bar{x}_i}{v_i}, i = 1, \cdots, k\). Then compute \(Q_B\) in equation (2.5).

3. Repeat step 2 a total \(m\) times and obtain \(m\) values of \(Q_B\), denoted as \(Q_B^{(i)}, i = 1, 2, \cdots, m\).

4. The percentage that \(Q_B^{(i)}\) \((i = 1, 2, \cdots, m)\) are greater than or equal to \(Q_{BO}\) is a Monte Carlo estimate of the p-value for testing \(H_0\) vs. \(H_1\).

**2.2. A Generalized Test Variable Approach.** Weserahandi (1995) presented a GTV for the equality of several Gaussian means under heterogeneity. A parallel test for IG distribution is developed as follows [11]. Let \(X_{1i}, X_{2i}, \cdots, X_{ni}\) be a random sample from an IG \((\mu_i, \lambda_i)\) population, with \(\bar{X}_i\) as the ith sample mean for \(i = 1, \cdots, k\). Let \(\bar{x}_i\) denotes the ith observed sample mean for \(i = 1, \cdots, k\). It is well known that both of the equation (2.1) and equation (2.2) could be written as it is done before in PB approach section. The generalized pivot for \(\lambda_i\) could
be written as shown in equation (2.10). Here, $v_i$ is the observed value of $V_i$. Obviously, $R_{\lambda_i}$ coincides with the traditional pivot for $\lambda_i$ [7, 11].

\begin{equation}
R_{\lambda_i} = \frac{\lambda_i V_i}{v_i} \sim \frac{\chi^2_{i-1}}{v_i}
\end{equation}

Under the null hypothesis that is given in equation (1.2), equation (2.1) could be written. Here, $\hat{\mu} = \frac{\sum_{i=1}^{k} n_i x_i \lambda_i}{\sum_{i=1}^{k} n_i \lambda_i}$ is as given before in Section 2.1. In the same vein of the GTV for the equality of normal means, a potential GTV is defined for the equality of IG means as in equation (2.12). Here, $q \left( R_{\lambda_1}, R_{\lambda_2}, \ldots, R_{\lambda_k} \right) = \sum_{i=1}^{k} n_i R_{\lambda_i} \left( \frac{1}{n_i} - \frac{1}{R_{\lambda_i}} \right)$ with $R_{\mu} = \frac{\sum_{i=1}^{k} n_i \bar{x}_i}{\sum_{i=1}^{k} n_i}$. That is, $q \left( R_{\lambda_1}, R_{\lambda_2}, \ldots, R_{\lambda_k} \right)$ is $Q \left( \lambda_1, \lambda_2, \ldots, \lambda_k \right)$ with $\bar{x}_i$, replaced by the observed value $\bar{X}_i$ and $\lambda_i$, replaced by the corresponding pivot $R_{\lambda_i}$ for $i = 1, \ldots, k$ [7, 10, 11].

\begin{equation}
Q \left( \lambda_1, \lambda_2, \ldots, \lambda_k \right) = \sum_{i=1}^{k} n_i \lambda_i \left( \frac{1}{X_i} - \frac{1}{\mu} \right) \sim \chi^2_{k-1}
\end{equation}

\begin{equation}
T = \frac{q \left( \lambda_1, \lambda_2, \ldots, \lambda_k \right)}{q \left( R_{\lambda_1}, R_{\lambda_2}, \ldots, R_{\lambda_k} \right)}
\end{equation}

The proposed test statistic $T$ satisfies the three conditions in order to be an actual GTV:

First of all, as $\bar{X}_i = \bar{x}_i$, $V_i = v_i$, $R_{\lambda_i} = \lambda_i$ ($i = 1, \ldots, k$) and $Q \left( \lambda_1, \ldots, \lambda_k \right) = q \left( R_{\lambda_1}, R_{\lambda_2}, \ldots, R_{\lambda_k} \right)$. Hence, the observed value of $T$ is $t_{obs} = 1$. Secondly, under null hypothesis, $Q \left( \lambda_1, \lambda_2, \ldots, \lambda_k \right) \sim \chi^2_{k-1}$ and $q \left( R_{\lambda_1}, R_{\lambda_2}, \ldots, R_{\lambda_k} \right)$ are functions of a set of independent random variables $\chi^2_{k-1}$ and observed values $\bar{x}_i$ and $V_i$ for $i = 1, \ldots, k$. Thus, $T$ does not depend on any unknown parameters. Thirdly, $T$ tends to take large values from null hypothesis. Consequently, $T$ becomes an actual GTV for testing the equality of IG means [11]. As $T$ tends to take larger than 1 for deviations from null hypothesis, the generalized p-value could be obtained as shown in equation (2.13). Here, $G_{k-1} \left[ \cdot \cdot \cdot \right]$ represents the cumulative distribution function of $\chi^2_{k-1}$ distribution and the expectation $E \left\{ \cdot \cdot \cdot \right\}$ is taken with respect to k independent $\chi^2_{n_i-1}$ ($i = 1, 2, \ldots, k$) random variables [7, 11].

\begin{equation}
p \text{- value} = \Pr ob \left( T \geq t_{obs} = 1 \mid H_0 \right) = \Pr ob \left( Q \left( \lambda_1, \lambda_2, \ldots, \lambda_k \right) \geq q \left( R_{\lambda_1}, R_{\lambda_2}, \ldots, R_{\lambda_k} \right) \mid H_0 \right) = 1 - E \left\{ G_{k-1} \left[ q \left( R_{\lambda_1}, R_{\lambda_2}, \ldots, R_{\lambda_k} \right) \right] \right\}
\end{equation}

Based on the simulation results in Tian (2006), the proposed test based on the concept of GTV was strongly recommended being used instead of ANORE test for testing equality of IG means, owing to the fact that it has excellent type I risk control under both homogeneity and heterogeneity.

### 3. The Computational Approach Test for Testing Equality of Inverse Gaussian Means under Heterogeneity

In this section, a test procedure is given based on CAT for testing equality of several IG means under the unequal scale parameters. Firstly, before applying developed CAT procedure for testing the null hypothesis given in equation (1.2), the $H_0$ is expressed in terms of suitable scalar $\eta$. $\eta$ is defined as shown in equation (3.1). Here, $\bar{\mu} = \sum_{i=1}^{k} n_i \mu_i / n$ and $n = \sum_{i=1}^{k} n_i$.
\[ \eta = \eta (\mu_1, \ldots, \mu_k) = \sum_{i=1}^{k} n_i \left( \frac{1}{\mu_i} - \frac{1}{\bar{\mu}} \right) \]

It is clear that testing \( H_0 \) against \( H_1 \) is equivalent to testing \( H_0^* : \eta = 0 \) against \( H_1^* : \eta > 0 \). Thus, MLE of \( \eta \) can be used as a test statistic. If \( H_0 \) is true then \( \eta = 0 \), otherwise \( \eta \) is getting greater than 0. In this study, \( \eta = \sum_{i=1}^{k} n_i \left( \frac{1}{\mu_i} - \frac{1}{\bar{\mu}} \right) \) is chosen. However, one can choose another suitable expression. The choice of \( \eta \) is the most important point for the success of the CAT method.

The test procedure of our proposed CAT could be given as shown below:

i) The MLE of the parameters are obtained as 
\[ \hat{\mu}_{(ML)} = \overline{X}_i \text{ and } \hat{\lambda}_{(ML)}^{-1} = \frac{1}{n_i} \sum_{j=1}^{n_i} (X_{ij}^{-1} - \overline{X}_{-k}^{-1}) \]

Therefore, the test statistic is rewritten as \( \hat{\eta}_{ML} = \sum_{i=1}^{k} n_i \left( \frac{1}{\overline{X}_{-k}} - \frac{1}{\bar{\mu}} \right) \). The observed value of \( \hat{\eta}_{ML} \) is \( \hat{\eta}_{ML0} \).

ii) If \( H_0^* \) is true, then \( X_{ij} \sim IG (\mu, \lambda_i) \) \( (1 \leq j \leq n_i, 1 \leq i \leq k) \). Here, \( \mu \) and \( (\lambda_1, \ldots, \lambda_k) \) are nuisance parameters. Hence, MLEs of \( \mu \) and \( \lambda_i \) based on the assumption under null hypothesis \( (\mu_1 = \mu_2 = \cdots = \mu_k = \mu) \) give the restricted maximum likelihood estimates (RMLEs) of these parameters. The procedure of obtaining RMLEs of the \( \mu \) and \( \lambda_i \) parameters could be given as follows:

Firstly, under the restricted model, log - likelihood function of the sample \( x_{11}, x_{12}, \ldots, x_{in} \) is given as in equation (3.2).

\[ L = \frac{1}{2} \sum_{i=1}^{k} \sum_{j=1}^{n_i} \log \frac{1}{\nu} - \frac{3}{2} \sum_{i=1}^{k} n_i \log x_{ij} - \frac{1}{2} \sum_{i=1}^{k} n_i \lambda_i x_{ij} - \frac{1}{2} \sum_{i=1}^{k} n_i \lambda_i - \frac{1}{2} \sum_{i=1}^{k} \sum_{j=1}^{n_i} \frac{\lambda_i}{x_{ij}} \]

Differentiating the equation (3.2) with respect to \( \mu \) and \( \lambda_i \) yields the following results:

\[ \frac{\partial L}{\partial \mu} = \frac{1}{\mu} \sum_{i=1}^{k} \sum_{j=1}^{n_i} \lambda_i x_{ij} - \frac{1}{2} \sum_{i=1}^{k} n_i \lambda_i \]

\[ \frac{\partial L}{\partial \lambda_i} = \frac{n_i}{\nu} + \frac{n_i}{\mu} - \frac{1}{2} \sum_{j=1}^{n_i} x_{ij} - \frac{1}{2} \sum_{j=1}^{n_i} x_{ij} \lambda_i, \quad i = 1, \ldots, l \]

As can be seen from equation (3.3), the RMLEs of the \( \mu \) and \( \lambda \) parameters have no closed forms. Therefore, the RMLEs of these parameters could be obtained using the iterations given in equation (2.9). Here \( \mu^{(l)}, \lambda^{(l)}_1, \ldots, \lambda^{(l)}_k \) converge to the RMLEs denoted as \( \hat{\mu}_{RML}, \hat{\lambda}_{RML} \).

iii) Generate artificial sample \( X_{i1}, \ldots, X_{in_i}, i = 1, \ldots, k \) i.i.d. from \( IG (\hat{\mu}_{RML}, \hat{\lambda}_{RML}) \) a large number of times (say, \( m \) times). For each of these replicated samples, recalculate the values of \( \hat{\eta}_{ML}^{(l)} \) \( (j = 1, \ldots, m) \).

iv) Calculate the p-value as \( p = \frac{\# \left( \hat{\eta}_{ML}^{(l)} > \hat{\eta}_{ML0} \right)}{m} \). In the case of \( p < \alpha \), \( H_0 \) is rejected.

4. A Simulation Study

In this section for testing equality of IG means under heterogeneity, the GTV, PB approaches and CAT are compared according to type I risks and powers for different combinations of parameters \((\mu, \lambda)\) and sample sizes. For this purpose, we consider some cases from smaller to larger sample sizes with different number of groups as \(k=3\), \(k=4\) and \(k=5\). For specified nominal level of \( \alpha=0.05 \), we used \( m=10000 \) times to calculate the
simulated type I risks and powers of these tests. Firstly, we calculated simulated type I risks of tests under null hypothesis. The numerical results for estimated type I risks are given as in Table 1 to Table 3.

**Table 1. Estimated type I risks of three tests for k=3.**

<table>
<thead>
<tr>
<th>µ</th>
<th>n</th>
<th>λ₁, λ₂, λ₃</th>
<th>PB</th>
<th>GTV</th>
<th>CAT</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>30 30 30</td>
<td>0.059</td>
<td>0.046</td>
<td>0.048</td>
<td></td>
</tr>
<tr>
<td></td>
<td>30 40 50</td>
<td>0.058</td>
<td>0.047</td>
<td>0.047</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>30 30 30</td>
<td>0.058</td>
<td>0.049</td>
<td>0.048</td>
<td></td>
</tr>
<tr>
<td></td>
<td>30 35 40</td>
<td>0.056</td>
<td>0.046</td>
<td>0.047</td>
<td></td>
</tr>
<tr>
<td></td>
<td>30 40 50</td>
<td>0.059</td>
<td>0.051</td>
<td>0.050</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>30 30 30</td>
<td>0.054</td>
<td>0.046</td>
<td>0.048</td>
<td></td>
</tr>
<tr>
<td></td>
<td>30 35 40</td>
<td>0.049</td>
<td>0.045</td>
<td>0.047</td>
<td></td>
</tr>
<tr>
<td></td>
<td>30 40 50</td>
<td>0.054</td>
<td>0.048</td>
<td>0.050</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>30 30 30</td>
<td>0.062</td>
<td>0.048</td>
<td>0.047</td>
<td></td>
</tr>
<tr>
<td></td>
<td>30 35 40</td>
<td>0.056</td>
<td>0.043</td>
<td>0.044</td>
<td></td>
</tr>
<tr>
<td></td>
<td>30 40 50</td>
<td>0.060</td>
<td>0.049</td>
<td>0.046</td>
<td></td>
</tr>
</tbody>
</table>

**Table 2. Estimated type I risks of three tests for k=4.**

<table>
<thead>
<tr>
<th>µ</th>
<th>n</th>
<th>λ₁, λ₂, λ₃, λ₄</th>
<th>PB</th>
<th>GTV</th>
<th>CAT</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>30 30 30 30</td>
<td>0.055</td>
<td>0.053</td>
<td>0.042</td>
<td></td>
</tr>
<tr>
<td></td>
<td>30 35 40 45</td>
<td>0.054</td>
<td>0.051</td>
<td>0.042</td>
<td></td>
</tr>
<tr>
<td></td>
<td>30 40 50 60</td>
<td>0.053</td>
<td>0.049</td>
<td>0.044</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>30 30 30 30</td>
<td>0.054</td>
<td>0.050</td>
<td>0.049</td>
<td></td>
</tr>
<tr>
<td></td>
<td>30 35 40 45</td>
<td>0.051</td>
<td>0.049</td>
<td>0.044</td>
<td></td>
</tr>
<tr>
<td></td>
<td>30 40 50 60</td>
<td>0.052</td>
<td>0.050</td>
<td>0.045</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>30 30 30 30</td>
<td>0.052</td>
<td>0.050</td>
<td>0.046</td>
<td></td>
</tr>
<tr>
<td></td>
<td>30 35 40 45</td>
<td>0.048</td>
<td>0.046</td>
<td>0.045</td>
<td></td>
</tr>
<tr>
<td></td>
<td>30 40 50 60</td>
<td>0.049</td>
<td>0.048</td>
<td>0.047</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>30 30 30 30</td>
<td>0.053</td>
<td>0.048</td>
<td>0.043</td>
<td></td>
</tr>
<tr>
<td></td>
<td>30 35 40 45</td>
<td>0.055</td>
<td>0.052</td>
<td>0.045</td>
<td></td>
</tr>
<tr>
<td></td>
<td>30 40 50 60</td>
<td>0.055</td>
<td>0.052</td>
<td>0.044</td>
<td></td>
</tr>
</tbody>
</table>

As seen from Table 1-Table 3, while the type I risks of PB approach exceed the nominal level in the case of k=3, they close to the 0.05 for k=5. GTV performs contrast to the
Table 3. Estimated type I risks of three tests for \(k=5\).

<table>
<thead>
<tr>
<th>(\mu)</th>
<th>(n)</th>
<th>(\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5)</th>
<th>PB</th>
<th>GTV</th>
<th>CAT</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>30</td>
<td>30 30 30 30 30</td>
<td>0.055</td>
<td>0.059</td>
<td>0.044</td>
</tr>
<tr>
<td></td>
<td>35</td>
<td>35 40 45 50</td>
<td>0.053</td>
<td>0.057</td>
<td>0.043</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>50 60 70</td>
<td>0.051</td>
<td>0.055</td>
<td>0.044</td>
</tr>
<tr>
<td>10</td>
<td>30</td>
<td>30 30 30 30 30</td>
<td>0.055</td>
<td>0.059</td>
<td>0.049</td>
</tr>
<tr>
<td></td>
<td>35</td>
<td>35 40 45 50</td>
<td>0.052</td>
<td>0.055</td>
<td>0.046</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>50 60 70</td>
<td>0.049</td>
<td>0.054</td>
<td>0.043</td>
</tr>
<tr>
<td>20</td>
<td>30</td>
<td>30 30 30 30 30</td>
<td>0.053</td>
<td>0.054</td>
<td>0.050</td>
</tr>
<tr>
<td></td>
<td>35</td>
<td>35 40 45 50</td>
<td>0.052</td>
<td>0.053</td>
<td>0.045</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>50 60 70</td>
<td>0.049</td>
<td>0.051</td>
<td>0.044</td>
</tr>
</tbody>
</table>

PB approach that its type I risks close to 0.05 for \(k=3\), whereas they exceed the 0.05 for \(k=5\). However, the CAT method seems to have type I risks lower also almost close to the nominal level for \(k=3, k=4\) and \(k=5\).

After calculating the type I risks of three methods, we calculated the estimated power of the tests for different combinations of parameters and sample sizes.

Table 4. Estimated power of three tests for \(k=3\).

<table>
<thead>
<tr>
<th>(n)</th>
<th>(\lambda_1, \lambda_2, \lambda_3)</th>
<th>(\mu_1, \mu_2, \mu_3)</th>
<th>PB</th>
<th>GTV</th>
<th>CAT</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>10 11 12</td>
<td>10 12 14</td>
<td>0.108</td>
<td>0.091</td>
<td>0.092</td>
</tr>
<tr>
<td></td>
<td>10 13 16</td>
<td>0.385</td>
<td>0.343</td>
<td>0.356</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10 11 12</td>
<td>0.118</td>
<td>0.098</td>
<td>0.104</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10 12 14</td>
<td>0.263</td>
<td>0.229</td>
<td>0.247</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10 13 16</td>
<td>0.429</td>
<td>0.391</td>
<td>0.420</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10 11 12</td>
<td>0.128</td>
<td>0.107</td>
<td>0.112</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10 12 14</td>
<td>0.276</td>
<td>0.245</td>
<td>0.266</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10 13 16</td>
<td>0.458</td>
<td>0.419</td>
<td>0.450</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10 11 12</td>
<td>0.176</td>
<td>0.162</td>
<td>0.166</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10 12 14</td>
<td>0.431</td>
<td>0.409</td>
<td>0.422</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10 13 16</td>
<td>0.689</td>
<td>0.669</td>
<td>0.681</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>10 11 12</td>
<td>0.179</td>
<td>0.166</td>
<td>0.174</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10 12 14</td>
<td>0.484</td>
<td>0.464</td>
<td>0.476</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10 13 16</td>
<td>0.734</td>
<td>0.717</td>
<td>0.743</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10 11 12</td>
<td>0.197</td>
<td>0.181</td>
<td>0.188</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10 12 14</td>
<td>0.517</td>
<td>0.495</td>
<td>0.516</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10 13 16</td>
<td>0.790</td>
<td>0.776</td>
<td>0.795</td>
<td></td>
</tr>
</tbody>
</table>
The numerical results for estimated power of the three tests are presented as above in Table 4 to Table 6. Since the type I risks of PB approach exceed the nominal level for \( k=3 \), it appears to be more powerful than the other two tests. If both of the CAT and GTV are compared for \( k=3 \), the CAT appears to be more powerful than the GTV.

For \( k=4 \) and \( n=15 \), the PB approach appears to be more powerful than the other two tests. However, it could be emphasized that when the sample size increases i.e., \( n=30 \), the
CAT performs a bit more better than the other two tests in terms of power. Additionally, the CAT is positively affected from the increments in the differences between the scale parameters and the means of groups.

For $k=5$, the GTV appears to be more powerful than other tests because of its type I risks exceed the nominal level. When the differences between scale parameters and means of groups are increased, the CAT appears to be more powerful than the PB approach.

5. Conclusion

In this article, we have proposed the CAT for testing the equality of several inverse Gaussian means, especially, under heterogeneity. We have compared the CAT with some of the existing tests; the PB and GTV. For a different sample sizes and parameters, we have investigated the performance of these three tests using Monte Carlo simulation. It could be observed from the simulation results that for $k=3$ the PB approach seems to have type I risks exceeding the nominal level and for $k=5$ the GTV performs in a similar way. However, the type I risks of CAT are generally less than 0.05 for all the numbers of groups. Therefore, we could mention that the CAT is not affected from the changes in the number of groups. Furthermore, according to power comparison results, the CAT appears to be more powerful than the other tests, when the differences between scale parameters and means of groups are increased.

Consequently, in respect of our simulation study, even for comparing different number of groups (as $k=3, 4$ or $5$), CAT could be suggested as a good alternative for testing the equality of inverse Gaussian means under heterogeneity of scale parameters.

References