

Testing the equality of several gamma means: a parametric bootstrap method with applications

Ching-Hui Chang · Jyh-Jiuan Lin · Nabendu Pal

Received: 6 April 2009 / Accepted: 16 July 2010 / Published online: 5 August 2010
© Springer-Verlag 2010

Abstract We apply a recently developed ‘Computational Approach Test’ (CAT), a variant of the parametric bootstrap method, to test the equality of means of several gamma distributions. All parameters are assumed to be unknown, and we consider two cases-(i) the shape parameters, though unknown, are assumed to be equal; and (ii) the shape parameters are all unknown and possibly unequal. The CAT, as applied to the above two cases, doesn’t require the knowledge of any sampling distribution, depends heavily on numerical computations and Monte-Carlo simulation, and figures out the critical region automatically. The power and/or size of our proposed CAT is quite encouraging compared with the other tests reported in the literature. The proposed method can be used as a logical alternative approach to classical one-way ANOVA when one is not sure about normality, and positively skewed distribution is a possibility for the observed data. Though the proposed CAT has been used recently to compare normal means by the present authors, its usefulness for comparing gamma means hadn’t been studied before. This paper shows that the CAT can be as good as, if not better than, the other proposed methods discussed in the literature to test the equality of several gamma means. Real life datasets have been used to illustrate the applicability of this method. Also, our comprehensive numerical study reveals that some of the frequently cited methods are not as good as they are claimed to be.

C.-H. Chang
Department of Applied Statistics and Information Science,
Ming Chuan University, Taoyuan County, Taiwan

J.-J. Lin
Department of Statistics, Tamkang University, Tamsui, Taipei, Taiwan

N. Pal (✉)
Department of Mathematics, University of Louisiana at Lafayette,
Louisiana, LA 70504, USA
e-mail: nxp3695@louisiana.edu

Keywords Hypotheses · Parametric bootstrap · Power function · Size

1 Introduction

The standard two-parameter gamma probability distribution is found to have wide applications in industrial engineering and quality control. The distribution is used to model life-time data due to its flexibility in shape which can vary from extreme positively skewed to almost symmetric on the positive side of the real line. Theoretically, the gamma distribution is a natural generalization of the exponential distribution which is quite popular in reliability studies. Apart from the life-time data, gamma distribution has found applications in many areas including hydrology, seismology, and economics. For more on these applications see [Singh and Singh \(1985\)](#), [McDonald and Jensen \(1979\)](#), [Bougeault \(1982\)](#), [Coe and Stern \(1982\)](#), [Li et al. \(2007\)](#), [Aksoy \(2000\)](#), [Simpson \(1972\)](#), [Singh \(2004\)](#), [Ming Tan et al. \(2007\)](#) and [Coles et al. \(2008\)](#).

In this paper we assume that independent samples are available from k -distinct gamma populations. To be precise, let X_{i1}, \dots, X_{ini} are *iid* following a gamma distribution with scale parameter λ_i and shape parameter δ_i (henceforth denoted by $G(\delta_i, \lambda_i)$) with *pdf*

$$f_i(x|\delta_i, \lambda_i) = \left(\lambda_i^{\delta_i} / \Gamma(\delta_i)\right) x^{\delta_i-1} \exp(-\lambda_i x), \quad (1.1)$$

$\lambda_i > 0$, $\delta_i > 0$, $x > 0$. The mean (or expected value) of $G(\delta_i, \lambda_i)$ is $\mu_i = (\delta_i/\lambda_i)$, $1 \leq i \leq k$; and our goal is to test

$$\begin{aligned} H_0 : \mu_1 = \mu_2 = \dots = \mu_k = \mu \quad (\text{say}) \\ \text{i.e., } H_0 : \delta_i = \mu\lambda_i, \quad 1 \leq i \leq k, \quad \text{for some } \mu > 0, \end{aligned} \quad (1.2)$$

against the alternative (H_A) that at least two means differ (i.e., $\mu_i \neq \mu_j$ for some $i \neq j$). But before going into the details of hypothesis testing let us review the parameter estimation briefly.

The method of moments estimators of (δ_i, λ_i) are easy to obtain, but they lack efficiency since they are not functions of the sufficient statistics. Based on the sample from the i th population, the sufficient statistics for (δ_i, λ_i) is (\bar{X}_i, \tilde{X}_i) , where

$$\bar{X}_i = \sum_{j=1}^{n_i} X_{ij}/n_i, \quad \tilde{X}_i = \left(\prod_{j=1}^{n_i} X_{ij}\right)^{1/n_i}, \quad 1 \leq i \leq k; \quad (1.3)$$

are respectively the arithmetic mean (AM) and the geometric mean (GM) of the i th sample observations. The maximum likelihood estimators (MLEs) of δ_i and λ_i , denoted by $\hat{\delta}_{i(ML)}$ and $\hat{\lambda}_{i(ML)}$ respectively, are found by solving the following system of equations

$$\ln \delta_i - \psi(\delta_i) = \ln R_i, \quad \lambda_i = \delta_i/\tilde{X}_i; \quad (1.4)$$

where $R_i = \bar{X}_i/\tilde{X}_i$ (≥ 1) is the ratio of the AM and the GM, and $\psi(c) = (\partial/\partial c)\ln\Gamma(c)$ represents the di-gamma function. Due to the non-explicit nature of the above MLEs, the exact sampling distributions of $\hat{\delta}_{i(ML)}$ and $\hat{\lambda}_{i(ML)}$ are impossible to get. However, [Greenwood and Durand \(1960\)](#) provided a convenient rational approximations of $\hat{\delta}_{i(ML)}$ and $\hat{\lambda}_{i(ML)}$.

The asymptotic distribution of $(\hat{\delta}_{i(ML)}, \hat{\lambda}_{i(ML)})$, as $n_i \rightarrow \infty$, is a bivariate normal with mean (δ_i, λ_i) with the following covariance structure:

$$\begin{aligned} Var\left(\sqrt{n_i}\hat{\delta}_{i(ML)}\right) &\approx \delta_i/\{\delta_i\psi'(\delta_i) - 1\}; \\ Var\left(\sqrt{n_i}\hat{\lambda}_{i(ML)}\right) &\approx \lambda_i^2\psi'(\delta_i)/\{\delta_i\psi'(\delta_i) - 1\}; \\ Cov\left(\sqrt{n_i}\hat{\lambda}_{i(ML)}, \sqrt{n_i}\hat{\delta}_{i(ML)}\right) &\approx \lambda_i/\{\delta_i\psi'(\delta_i) - 1\}. \end{aligned} \tag{1.5}$$

For testing H_0 (in (1.2)) the existing results fall under two broad cases—**Case(I)**: Testing H_0 under the assumption $\delta_1 = \delta_2 = \dots = \delta_k = \delta$ (*say*), where δ is unknown; and **Case(II)**: Testing H_0 when δ_i 's are all unknown, and possibly unequal.

The rest of the paper is organized as follows. In the next section (Sect. 2) we briefly review the existing main results relevant for testing H_0 . In Sect. 3, we present our proposed CAT and how it is implemented for testing the equality of several gamma means. Section 4, provides the results of our comprehensive simulation study comparing several test procedures in terms of size/power. In Sect. 5 we present a few applications with real life datasets. Finally, in Sect. 6 we address some important issues related to bootstrap hypothesis testing which are relevant for this paper.

2 A brief literature review on testing H_0

Here we briefly review the main results pertaining to testing H_0 (in (1.2)) under the two cases mentioned above.

When the shape parameters (i.e., δ_i 's) are all assumed to be equal, testing the equality of gamma means boil down to testing the equality of the scale parameters only, i.e., $H_0^* : \lambda_1 = \lambda_2 = \dots = \lambda_k = \lambda$ (*say*). [Shiue and Bain \(1983\)](#) provided an approximate test for $k = 2$ based on the ratio of two sample means. To follow this, first note that if the common shape parameter δ were known, then one needs to look at the statistic (\bar{X}_1/\bar{X}_2) , for $k = 2$, which can be expressed as

$$(\bar{X}_1/\bar{X}_2) = \{(n_2\lambda_2)/(n_1\lambda_1)\}\{(2n_1\delta)/(2n_2\delta)\}W, \tag{2.1}$$

where $W = \{(2\lambda_1n_1\bar{X}_1)/(2n_1\delta)\}/\{2\lambda_2n_2\bar{X}_2/(2n_2\delta)\}$.

Since $(2\lambda_in_i\bar{X}_i)$ follows $\chi^2_{2n_i\delta}$, under $H_0^* : \lambda_1 = \lambda_2$ the ratio (\bar{X}_1/\bar{X}_2) follows $F(2n_1\delta, 2n_2\delta)$ -distribution. As a result, one rejects H_0^* against a suitable alternative if (\bar{X}_1/\bar{X}_2) exceeds the cut-off point(s) of the $F(2n_1\delta, 2n_2\delta)$ -distribution. But when δ is unknown, one can replace δ by $\hat{\delta}_{ML}$ (the MLE of δ based on the combined sample

X_{ij} , $1 \leq j \leq n_i$, $i = 1, 2$), and then use the cut-off point(s) of

$$F(2n_1\hat{\delta}_{ML}, 2n_2\hat{\delta}_{ML})\text{-distribution} \quad (2.2)$$

as the approximate critical region. If $F(k_1, k_2, \gamma)$ denotes the right tail γ -probability cut-off point of the $F(k_1, k_2)$ distribution, then the actual size of testing H_0^* against $H_A^* : \lambda_1 > \lambda_2$ at level α is

$$p(\alpha, \delta) = P_{H_0}(\bar{X}_1/\bar{X}_2 < F(2n_1\hat{\delta}_{ML}, 2n_2\hat{\delta}_{ML}, (1-\alpha))). \quad (2.3)$$

A suitable modification of the above test can be done for the two sided alternative. The above size $p(\alpha, \delta)$ doesn't depend on λ and, as evident from [Shiue and Bain \(1983\)](#) simulation, depends on δ very mildly. The values of $p(\alpha, \delta)$ were tabulated for small (n_1, n_2) which appear to be slightly higher than the prescribed level α . But since $p(\gamma, \delta) \approx p(\gamma, \infty)$ for small n_1 and n_2 , an approximate size α test for H_0^* against $H_A^* : \lambda_1 > \lambda_2$ is to reject H_0 if $(\bar{X}_1/\bar{X}_2) < F(2n_1\hat{\delta}_{ML}, 2n_2\hat{\delta}_{ML}, (1-\alpha_0))$, where α_0 is the value such that $p(\alpha_0, \infty) = \alpha$, and this α_0 depends on (n_1, n_2) through $(n_1 + n_2)$. Table for α_0 was provided for selected values of $(n_1 + n_2)$ and α by [Grice and Bain \(1980\)](#).

Extending the above result for two unequal shape parameters ([Shiue et al. 1988](#)) provided a simple generalization as follows. Similar to (2.1), and using the fact that $\mu_i = (\delta_i/\lambda_i)$,

$$(\bar{X}_1/\mu_1) / (\bar{X}_2/\mu_2) \sim F(2n_1\delta_1, 2n_2\delta_2). \quad (2.4)$$

Since δ_1 and δ_2 are unknown, an approximate level α test for $H_0 : \mu_1 = \mu_2$ versus $H_A : \mu_1 < \mu_2$ based on $\hat{\delta}_{i(ML)} (i = 1, 2)$ is to reject H_0 if $(\bar{X}_1/\bar{X}_2) < F(2n_1\hat{\delta}_{1(ML)}, 2n_2\hat{\delta}_{2(ML)}, (1-\alpha))$. But again the actual size of this approximate test is not α , and to attain level α one needs to use a refined test with critical region

$$(\bar{X}_1/\bar{X}_2) < F(2n_1\hat{\delta}_{1(ML)}, 2n_2\hat{\delta}_{2(ML)}, (1-\alpha_1)), \quad (2.5)$$

where α_1 is computed based on n_1 only. By symmetry, a refined one sided test of $H_0 : \mu_1 = \mu_2$ versus $H_A : \mu_1 > \mu_2$ is obtained by the critical region

$$(\bar{X}_2/\bar{X}_1) < F(2n_2\hat{\delta}_{2(ML)}, 2n_1\hat{\delta}_{1(ML)}, (1-\alpha_2)), \quad (2.6)$$

where α_2 is computed based on n_2 only. The values of α_i ($i = 1, 2$) in (2.5) and (2.6) are readily available from the common table provided by [Grice and Bain \(1980\)](#).

[Tripathi et al. \(1993\)](#) provided a general framework for testing any linear structure involving (δ_i, λ_i) , $1 \leq i \leq k$. For testing the equality of gamma means (see (1.2)), [Tripathi et al. \(1993\)](#) general test statistic Δ_Q (details of which are cumbersome and hence omitted here) asymptotically follows $\chi_{(k-1)}^2$ under H_0 , and hence an approximate size α critical region is: $\Delta_Q > \chi_{(k-1), \alpha}^2$. The above authors made a limited

comparison between the large sample power of their test and that of [Shiue and Bain \(1983\)](#) for $k = 2$, and that too for equal shape parameters. But no comprehensive comparison has been done for small to moderate sample sizes and for unequal shape parameters.

The other asymptotic test which has a wide appeal is the traditional likelihood ratio test (LRT). Define Λ as

$$\Lambda = \sup_{H_0} L / \sup L, \tag{2.7}$$

where $L = L(\lambda_i, \delta_i, 1 \leq i \leq k | x_{ij}, 1 \leq j \leq n_i, 1 \leq i \leq k)$ is the likelihood function, and the supremum in the denominator of (2.7) is taken over the whole parameter space. Asymptotically $(-2 \ln \Lambda)$ follows (under H_0) $\chi^2_{(k-1)}$, and hence the critical region is: $(-2 \ln \Lambda) > \chi^2_{(k-1), \alpha}$. Again, how the LRT performs powerwise for small to moderate sample sizes in comparison to the test by [Tripathi et al. \(1993\)](#) is not known. This will be a part of our goal while we study the power of our proposed CAT as described in the next section.

3 The computational approach test (CAT)

3.1 The general framework

The computational approach test (CAT) discussed here is a slight modification of what was applied in an earlier study to test the equality of several normal means (the classical one-way ANOVA). The CAT is a simple procedure based on a simple set of computational steps which can be implemented easily by applied researchers. Our proposed CAT does not require any knowledge of the complex sampling distribution of a test statistic, and the computational mechanism is such that the CAT finds the critical region automatically. In the case of one-way ANOVA with the usual assumptions, the proposed CAT provides power which is very close to that of the classical F -test (see [Chang and Pal 2008](#)). This motivates us to apply the CAT to test the equality of several gamma means.

For better understanding, we'll consider the two cases separately as mentioned at the end of Sect. 1. One of the features of applying CAT is that both H_0 and H_A need to be represented in terms of a scalar parametric function. Suppose our data X_1, \dots, X_n ($iid \sim f(x|\theta)$), where both X and θ are possibly multidimensional, and the functional form of 'f' is known. Let $\theta = (\theta^{(1)}, \theta^{(2)})$ where $\theta^{(1)}$ is the parameter of interest $\theta^{(2)}$ is the nuisance parameter. To test $H_0 : \theta^{(1)} = \theta_0^{(1)}$ against a suitable alternative H_A , first we express H_0 as $H_0^* : \eta(\theta^{(1)}, \theta_0^{(1)}) = 0$ (against a suitable alternative H_A^*), where η is a scalar valued function. The general methodology of the proposed CAT for testing $H_0^* : \eta(\theta^{(1)}, \theta_0^{(1)}) = 0$ against a suitable alternative at a desired level α is given through the following steps.

Step-1: First obtain $\hat{\theta}_{ML} = (\hat{\theta}_{ML}^{(1)}, \hat{\theta}_{ML}^{(2)})$, the MLE of θ . Hence $\hat{\eta}_{ML} = \eta(\hat{\theta}_{ML}^{(1)}, \theta_0^{(1)})$.

- Step-2: (i) Assume that H_0 is true, i.e., set $\theta^{(1)} = \theta_0^{(1)}$ (or $\eta(\theta^{(1)}, \theta_0^{(1)}) = 0$). Then find the MLE of $\theta^{(2)}$ from the original data again (i.e., assuming $X_1, \dots, X_n \text{ iid} \sim f(x|\theta_0^{(1)}, \theta^{(2)})$) and call this as the ‘restricted MLE (RMLE)’ of $\theta^{(2)}$ under H_0 , denoted by $\hat{\theta}_{RML}^{(2)}$.
- (ii) Generate artificial sample X_1, \dots, X_n ($\text{iid} \sim f(x|\theta_0^{(1)}, \hat{\theta}_{RML}^{(2)})$) a large number of times (say, M times). For each of these replicated samples, recalculate the MLE of $\theta = (\theta^{(1)}, \theta^{(2)})$ pretending that θ were unknown. From these MLEs retain only the first component (that is associated with $\theta^{(1)}$). Let these recalculated MLE values of $\theta^{(1)}$ from the replicated samples be $\hat{\theta}_{01}^{(1)}, \hat{\theta}_{02}^{(1)}, \dots, \hat{\theta}_{0M}^{(1)}$. Let $\hat{\eta}_{0l} = \eta(\hat{\theta}_{0l}^{(1)}, \theta_0^{(1)})$, $1 \leq l \leq M$.
- (iii) Let $\hat{\eta}_{0(1)} \leq \hat{\eta}_{0(2)} \leq \dots \leq \hat{\eta}_{0(M)}$ be the ordered values of $\hat{\eta}_{0l}$, $1 \leq l \leq M$.
- Step-3: (i) For testing $H_0^* : \eta(\theta^{(1)}, \theta_0^{(1)}) = 0$ versus $H_A^* : \eta(\theta^{(1)}, \theta_0^{(1)}) < 0$ (if such an alternative is meaningful), define $\hat{\eta}_L = \hat{\eta}_{0(\alpha M)}$. Reject H_0^* if $\hat{\eta}_{ML} < \hat{\eta}_L$; and accept H_0^* otherwise. Alternatively, calculate the p -value as: $p\text{-value} = \sum_{l=1}^M I(\hat{\eta}_{0(l)} < \hat{\eta}_{ML})/M$.
- (ii) For testing $H_0^* : \eta(\theta^{(1)}, \theta_0^{(1)}) = 0$ against $H_A^* : \eta(\theta^{(1)}, \theta_0^{(1)}) > 0$ (if such an alternative is meaningful), define $\hat{\eta}_U = \hat{\eta}_{0((1-\alpha)M)}$. Reject H_0^* if $\hat{\eta}_{ML} > \hat{\eta}_U$; and accept H_0^* otherwise. Alternatively, calculate the p -value as: $p\text{-value} = \sum_{l=1}^M I(\hat{\eta}_{0(l)} > \hat{\eta}_{ML})/M$.
- (iii) For testing $H_0^* : \eta(\theta^{(1)}, \theta_0^{(1)}) = 0$ against $H_A^* : \eta(\theta^{(1)}, \theta_0^{(1)}) \neq 0$, define $\hat{\eta}_L = \hat{\eta}_{0((\alpha/2)M)}$ and $\hat{\eta}_U = \hat{\eta}_{0((1-\alpha/2)M)}$. Reject H_0^* if $\hat{\eta}_{ML}$ is either $> \hat{\eta}_U$ or $< \hat{\eta}_L$; accept H_0^* otherwise. Alternatively, the p -value is computed as: $p\text{-value} = 2\min(p_1, p_2)$, where $p_1 = \sum_{l=1}^M I(\hat{\eta}_{0(1)} < \hat{\eta}_{ML})/M$, and $p_2 = \sum_{l=1}^M I(\hat{\eta}_{0(1)} > \hat{\eta}_{ML})/M$.

Remark 3.1 The proposed CAT is heuristic and simple to implement. When there is no nuisance parameter, we do not have any ‘restricted MLE’ for the nuisance parameter to worry about, and the resampling is done from the completely known distribution (under H_0). Thus the CAT provides critical value(s) of the test statistic (which is the MLE of the parameter of interest) through simulation, mimicking the exact distribution of the MLE (under H_0). The presence of nuisance parameter(s) makes the hypothesis testing a bit more challenging. The proposed CAT procedure borrows ideas from the classical likelihood ratio test (LRT) method as well as parametric bootstrap. But parametric bootstrap for testing, as it is described in the literature, is not very clear and uniform. For example, the parametric bootstrap as described in Efron (1982, pp. 29–30) or in Hjorth (1994, pp. 131–133) doesn’t clarify the role played by the nuisance parameter(s) in obtaining the critical values. It is told that by generating data from $f(x|\hat{\theta}_{ML}^{(1)}, \hat{\theta}_{ML}^{(2)})$ one gets a bootstrap estimate, say $\hat{\theta}_{ML}^{(1)*}$, of $\theta^{(1)}$ —the parameter of interest. By repeating $\hat{\theta}_{ML}^{(1)*}$ many times one obtains a bootstrap distribution of $\hat{\theta}_{ML}^{(1)}$ which is then used to draw

inferences. But following the classical LRT method one expects to see maximization of the likelihood function under H_0 . This is indicated as part of parametric bootstrap in Davison and Hinkley (1997, p. 148) briefly which essentially suggests to replicate $T = -2 \ln LR = -2 \ln \left\{ L \left(\theta_0^{(1)}, \hat{\theta}_{RML}^{(2)} \right) / L \left(\hat{\theta}_{ML}^{(1)}, \hat{\theta}_{ML}^{(2)} \right) \right\}$, where $L \left(\theta^{(1)}, \theta^{(2)} \right)$ is the likelihood function, by drawing *iid* from $f \left(x | \theta_0^{(1)}, \hat{\theta}_{RML}^{(2)} \right)$ repeatedly. (Asymptotically, T is supposed to follow a Chi-square distribution under H_0 .) Our proposed CAT uses the idea of replicating data from $f \left(x | \theta_0^{(1)}, \hat{\theta}_{RML}^{(2)} \right)$, but then based on these replicated data repeats estimates of η as $\hat{\eta}_{01}, \hat{\eta}_{02}, \dots, \hat{\eta}_{0M}$. Note that η is the scalar used to represent the null hypothesis as $H_0^* : \eta = 0$. Using the $\hat{\eta}_{0l}$'s to mimic the distribution of $\hat{\eta}_{ML}$ under H_0^* and then getting the critical value(s) for $\hat{\eta}_{ML}$ becomes easy and heuristically appealing. Also, η should be chosen carefully so that it remains invariant under a suitable group of transformations.

3.2 CAT for the gamma model

With the above general idea of CAT as the background, we now see its applications in the gamma mean problem for the two cases discussed earlier.

Case-I: Assume $\delta_1 = \delta_2 = \dots = \delta_k = \delta$ (say).

Under the above assumption our objective is to test $H_0 : \lambda_1 = \lambda_2 = \dots = \lambda_k = \lambda$ (say) versus $H_A : \lambda_i \neq \lambda_j$ for some $i \neq j$; which is same as testing $H_0^* : \eta = \sum_{i=1}^k (\beta_i - \bar{\beta})^2 = 0$ versus $H_A^* : \eta > 0$, where $\beta_i = \ln(\lambda_i)$, $i = 1, 2, \dots, k$, $\bar{\beta} = \sum_{i=1}^k \beta_i / k$.

(a) Step-1: Obtain $(\hat{\lambda}_{1(ML)}, \dots, \hat{\lambda}_{k(ML)}, \hat{\delta}_{ML})$ by solving the following system of equations:

$$\ln \delta - \psi(\delta) = \sum_{i=1}^k w_i \ln R_i; \text{ and } \lambda_i = \delta / \bar{X}_i, w_i = n_i / \sum_{i=1}^k n_i, 1 \leq i \leq k. \quad (3.1)$$

The MLE of η is obtained as $\hat{\eta}_{ML} = \sum_{i=1}^k (\hat{\beta}_{i(ML)} - \hat{\beta}_{ML})^2$. We'll need this $\hat{\eta}_{ML}$ later as our test statistic.

(b) Step-2:

(i) Assume that H_0 is true, i.e., set $\lambda_i = \lambda, 1 \leq i \leq k$. Get the MLEs of δ and λ (called the 'restricted MLEs'), i.e., $\hat{\delta}_{RML}$ and $\hat{\lambda}_{RML}$ respectively, by solving the following system of equations:

$$\ln \delta - \psi(\delta) = \ln R \text{ and } \lambda = \delta / \bar{X}_{..}, \quad (3.2)$$

where $R = \bar{X}_{..} / \bar{X}_{i.} =$ ratio of the grand AM $\bar{X}_{..}$ ($= \sum_{i=1}^k w_i \bar{X}_{i.}$), and the grand GM $\bar{X}_{i.}$ ($= \prod_{i=1}^k \bar{X}_{i.}^{w_i}$) with w_i 's as defined in (3.1).

(ii) Generate artificial sample $X_{i1}, \dots, X_{in_i} (= \underline{X}_i, \text{ say})$ *iid* from $G(\hat{\delta}_{RML}, \hat{\lambda}_{RML}), 1 \leq i \leq k$. But once the samples are generated pretend that they

are coming from $G(\delta, \lambda_i), 1 \leq i \leq k$, respectively, and then proceed further to estimate the parameters $(\delta, \lambda_i), 1 \leq i \leq k$, using the system of equations in (3.1) with the above artificial sample. This process needs to be repeated M times.

In the l th replication, $1 \leq l \leq M$, based on $X_i^{(l)}$ (from $G(\hat{\delta}_{RML}, \hat{\lambda}_{RML})$), $1 \leq i \leq k$, get the MLEs of λ_i and δ by using (3.1), and call them as $\hat{\lambda}_{0i}^{(l)}, \hat{\delta}_0^{(l)}$ (but retain only $\hat{\lambda}_{0i}^{(1)}$). Then calculate $\hat{\eta}$ as $\hat{\eta}_{0l} = \sum_{i=1}^k (\hat{\beta}_{0i}^{(l)} - \hat{\beta}_0^{(l)})^2$, where $\hat{\beta}_{0i}^{(l)} = \ln(\hat{\lambda}_{0i}^{(l)})$, $\hat{\beta}_0^{(l)} = \sum_{i=1}^k \hat{\beta}_{0i}^{(l)} / k$.

(iii) Order the $\hat{\eta}_{0l}$ values as $\hat{\eta}_{0(1)} \leq \hat{\eta}_{0(2)} \leq \dots \leq \hat{\eta}_{0(M)}$.

(c) Step-3: The original $\hat{\eta}_{ML}$ (in Step-1) is now compared with the cut-off point $\hat{\eta}_U = \hat{\eta}_{0((1-\alpha)M)}$. Reject H_0 if $\hat{\eta}_{ML} > \hat{\eta}_U$ and accept H_0 otherwise.

Case-II: δ_i 's are completely unknown and possibly unequal.

Under this case our objective is to test $H_0 : \delta_i = \mu\lambda_i, 1 \leq i \leq k$, for some $\mu > 0$; versus $H_A : (\delta_i/\lambda_i) \neq (\delta_j/\lambda_j)$ for some $i \neq j$. Then testing H_0 against H_A is equivalent to testing $H_0^* : \eta = \sum_{i=1}^k (\beta_i - \bar{\beta})^2 = 0$ versus $H_A^* : \eta > 0$, where $\beta_i = \ln(\mu_i), i = 1, 2, \dots, k, \bar{\beta} = \sum_{i=1}^k \beta_i / k$.

(a) Step-1: Obtain $(\hat{\lambda}_{i(ML)}, \hat{\delta}_{i(ML)}), 1 \leq i \leq k$, by solving the system of Eq. (1.4).

The MLEs of μ_i is obtained as $\hat{\mu}_{i(ML)} = (\hat{\delta}_{i(ML)} / \hat{\lambda}_{i(ML)})$. Then, $\hat{\eta}_{ML} = \sum_{i=1}^k (\hat{\beta}_{i(ML)} - \hat{\beta}_{ML})^2$. We'll use this $\hat{\eta}_{ML}$ as our test statistic later.

(b) Step-2:

(i) Assume that H_0 is true, i.e., set $\delta_i = \mu\lambda_i, 1 \leq i \leq k$. The likelihood function (under H_0) is a function of $(\mu, \lambda_1, \dots, \lambda_k)$ only. The 'restricted MLEs' of $\mu, \lambda_1, \dots, \lambda_k$, denoted by $\hat{\mu}_{RML}, \hat{\delta}_{i(RML)}, 1 \leq i \leq k$, are found by solving the following system of equations:

$$\mu\{(1 + \ln\lambda_i) - \psi(\mu\lambda_i) + \ln\tilde{X}_i\} = \tilde{X}_i, 1 \leq i \leq k; \text{ and} \tag{3.3}$$

$$\sum_{i=1}^k n_i \{\lambda_i \ln\lambda_i - \lambda_i \psi(\mu\lambda_i) + \lambda_i \ln(\tilde{X}_i)\} = 0, \tag{3.4}$$

where \tilde{X}_i and \tilde{X}_i are given in (1.3). Define $\hat{\delta}_{i(RML)} = \hat{\mu}_{RML} \hat{\lambda}_{i(RML)}, 1 \leq i \leq k$.

(ii) Generate artificial sample $X_{i1}, \dots, X_{in_i} (= \underline{X}_i, \text{say})$ iid from

$G(\hat{\delta}_{i(RML)}, \hat{\lambda}_{i(RML)}), 1 \leq i \leq k$. But once the samples are generated, pretend that they are coming from $G(\delta_i, \lambda_i), 1 \leq i \leq k$, respectively, and then proceed further to estimate the parameters (δ_i, λ_i) by using the system of Eq. (1.4). Repeat this process M times.

In the l th replication, $1 \leq l \leq M$, based on $X_i^{(l)}$ (from $G(\hat{\delta}_{i(RML)}, \hat{\lambda}_{i(RML)}), 1 \leq i \leq k$) get the MLEs of δ_i 's and λ_i 's by (1.4), and then

- call them as $\hat{\delta}_{0i}^{(l)}$ and $\hat{\lambda}_{0i}^{(l)}$. Recalculate $\hat{\eta}$ as $\hat{\eta}_{0l} = \sum_{i=1}^k (\hat{\beta}_{0i}^{(l)} - \hat{\beta}_0^{(l)})^2$, where $\hat{\beta}_{0i}^{(l)} = \ln(\hat{\delta}_{0i}^{(l)} / \hat{\lambda}_{0i}^{(l)})$, $\hat{\beta}_0^{(l)} = \sum_{i=1}^k \hat{\beta}_{0i}^{(l)} / k$.
- (iii) Order the $\hat{\eta}_{0l}$ values as $\hat{\eta}_{0(1)} \leq \hat{\eta}_{0(2)} \leq \dots \leq \hat{\eta}_{0(M)}$.
- (c) Step-3: The original $\hat{\eta}_{ML}$ (in Step-1) is now compared with the cut-off point $\hat{\eta}_U = \hat{\eta}_{0((1-\alpha)M)}$. Reject H_0 if $\hat{\eta}_{ML} > \hat{\eta}_U$ and accept H_0 otherwise.

Remark 3.2 The choice of our η needs a little clarification, and the success of our proposed CAT depends heavily on the selection of η . Note that η is a scalar expression used to characterize H_0 and H_A , and based on the sampling distribution of $\hat{\eta}_{ML}$, under H_0 , one accepts or rejects H_0 . It has been observed that the CAT works best (in terms of maintaining the desired level and attaining a high power) provided $\hat{\eta}_{ML}$ either behaves roughly like a normal variate (which may happen for $k = 2$ with one-sided H_A) or like a sum of normal variates' squares. Usually, when we have location parameters which can take values over the real line, we can use the standard quadratic (or squared) expression for η , as it is done in classical one-way ANOVA under the normality assumption. But when our parameters are nonnegative, and so are the observations, as we have here under the gamma model, the logarithmic transformation makes the parameters behave like location parameters. Also, the variance stabilizing transformation of asymptotic theory comes handy in attaining asymptotic normality at a faster rate. Think of the above Case-(I) where each $\hat{\lambda}_{i(ML)}$ is asymptotic normal (AN) with asymptotic mean λ_i and asymptotic variance proportional to λ_i^2 . If one thinks of a suitable transformation $g(\hat{\lambda}_{i(ML)})$ which follows $AN(g(\lambda_i), (g'(\lambda_i))^2 \lambda_i^2 c)$, where c is free from λ_i , such that $(g'(\lambda_i)\lambda_i)^2$ is also free from λ_i , then $g(\lambda_i) = \ln \lambda_i$ is the natural choice. As we will see later, in Sect. 4, our proposed CAT, with η as described earlier, does exhibit good size and power behavior even for small samples.

4 Numerical results for size and power comparison

In this section we compare the performance of our proposed CAT with those methods discussed in Sect. 2 in terms of size and power.

Case-I: $\delta_1 = \delta_2 = \dots = \delta_k = \delta$, δ unknown.

The powers of our proposed CAT, the LRT, and Shiue and Bain's test (SBT) are denoted by β_{CAT} , β_{LRT} , and β_{SBT} respectively. The Tables 1 and 2 give the results for $k = 2$ and $\alpha = 0.05$. For convenience λ_1 has been taken as 1.0, and λ_2 is varied from 0.10 to 5.0. Hence, the computed power of a test at $\lambda_2 = 1.0$ represents its size which depends on δ apart from n_i 's. Each simulated power is obtained by $Q = 10^4$ replications, and for the proposed CAT we have used $M = 10^4$. Our simulation study has been restricted to small to moderate sample sizes (i.e., $n_i \leq 25$); because for larger sample sizes (>25) all the tests exhibit very similar performance. Also, the standard error (SE) of our simulated power has been less than 0.002 most of the time. Only in a few instances it has been close to 0.005.

Remark 4.1 The trends that emerge from our simulation study with $k = 2$ are worth noticing.

Table 1 Power of three tests for $k = 2$ and $n_1 = n_2$ with $\alpha = 0.05$ (Case-I)

(n_1, n_2)	δ	Power	λ_2					
			0.5	0.75	1.0	1.25	2.0	5.0
(10, 10)	2.5	β_{CAT}	0.653	0.166	0.055*	0.121	0.651	1.000
		β_{LRT}	0.691	0.196	0.068*	0.143	0.692	1.000
		β_{SBT}	0.716	0.216	0.078*	0.158	0.715	1.000
	5.0	β_{CAT}	0.913	0.284	0.055*	0.191	0.910	1.000
		β_{LRT}	0.927	0.322	0.069*	0.221	0.929	1.000
		β_{SBT}	0.937	0.349	0.079*	0.241	0.939	1.000
	7.5	β_{CAT}	0.980	0.396	0.052*	0.267	0.982	1.000
		β_{LRT}	0.986	0.440	0.067*	0.303	0.986	1.000
		β_{SBT}	0.989	0.470	0.078*	0.331	0.989	1.000
	10.0	β_{CAT}	0.996	0.497	0.053*	0.330	0.997	1.000
		β_{LRT}	0.997	0.539	0.066*	0.373	0.998	1.000
		β_{SBT}	0.998	0.572	0.079*	0.405	0.998	1.000
(25, 25)	2.5	β_{CAT}	0.966	0.350	0.048*	0.243	0.969	1.000
		β_{LRT}	0.969	0.364	0.052*	0.256	0.972	1.000
		β_{SBT}	0.971	0.374	0.055*	0.265	0.974	1.000
	5.0	β_{CAT}	1.000	0.613	0.049*	0.417	0.999	1.000
		β_{LRT}	1.000	0.627	0.053*	0.433	0.999	1.000
		β_{SBT}	1.000	0.638	0.057*	0.446	0.999	1.000
	7.5	β_{CAT}	1.000	0.786	0.054*	0.566	1.000	1.000
		β_{LRT}	1.000	0.796	0.058*	0.580	1.000	1.000
		β_{SBT}	1.000	0.804	0.064*	0.595	1.000	1.000
	10.0	β_{CAT}	1.000	0.887	0.050*	0.688	1.000	1.000
		β_{LRT}	1.000	0.895	0.054*	0.701	1.000	1.000
		β_{SBT}	1.000	0.900	0.058*	0.712	1.000	1.000

* indicates size of a test

- (i) First note that the sizes of the tests are observed at $\lambda_2 = 1.0$ (marked with a ‘*’) for various δ values. Both the LRT and the SBT are consistently liberal with size greater than α , the nominal level. On the other hand, the CAT has been found to have size always very close to α . Whatever little difference we see between α and the CAT’s size can be attributed to plus-minus two SE’s of variation. Also, the size for all the three tests are mildly dependent on δ . The variation of size over δ is not significant if we take the standard error (SE) of the simulation into consideration.
- (ii) Size-wise the CAT is the best as it almost attains α . What came as a surprise is the size of our usual LRT which, though more liberal than the CAT, is better than the SBT. Interestingly, [Shiue and Bain \(1983\)](#) didn’t consider comparing their test (i.e., SBT) with the LRT. All our simulation results have shown that the LRT’s size is between those of the CAT and SBT. As a result, the power of the SBT is higher than that of the LRT, which in turn is marginally better

Table 2 Power of three tests for $k = 2$ and $n_1 \neq n_2$ with $\alpha = 0.05$ (Case-I)

(n_1, n_2)	δ	Power	λ_2					
			0.5	0.75	1.0	1.25	2.0	5.0
(10, 15)	2.5	β_{CAT}	0.739	0.204	0.055*	0.128	0.720	1.000
		β_{LRT}	0.757	0.218	0.066*	0.155	0.762	1.000
		β_{SBT}	0.769	0.231	0.075*	0.172	0.782	1.000
	5.0	β_{CAT}	0.957	0.348	0.052*	0.222	0.952	1.000
		β_{LRT}	0.962	0.371	0.063*	0.256	0.961	1.000
		β_{SBT}	0.966	0.393	0.071*	0.278	0.967	1.000
	7.5	β_{CAT}	0.995	0.484	0.053*	0.308	0.992	1.000
		β_{LRT}	0.996	0.509	0.061*	0.344	0.993	1.000
		β_{SBT}	0.997	0.531	0.072*	0.370	0.994	1.000
	10.0	β_{CAT}	1.000	0.577	0.048*	0.390	0.999	1.000
		β_{LRT}	1.000	0.606	0.059*	0.431	0.999	1.000
		β_{SBT}	1.000	0.632	0.068*	0.463	0.999	1.000
(5, 25)	2.5	β_{CAT}	0.605	0.170	0.051*	0.088	0.528	0.996
		β_{LRT}	0.587	0.162	0.058*	0.131	0.625	0.999
		β_{SBT}	0.580	0.157	0.063*	0.152	0.661	0.999
	5.0	β_{CAT}	0.881	0.272	0.053*	0.145	0.835	1.000
		β_{LRT}	0.877	0.266	0.063*	0.191	0.881	1.000
		β_{SBT}	0.878	0.269	0.068*	0.215	0.897	1.000
	7.5	β_{CAT}	0.972	0.368	0.048*	0.211	0.954	1.000
		β_{LRT}	0.971	0.365	0.059*	0.260	0.968	1.000
		β_{SBT}	0.972	0.373	0.066*	0.287	0.973	1.000
	10.0	β_{CAT}	0.995	0.457	0.056*	0.278	0.987	1.000
		β_{LRT}	0.995	0.459	0.063*	0.328	0.991	1.000
		β_{SBT}	0.995	0.468	0.072*	0.358	0.994	1.000

* indicates size of a test

than that of CAT for obvious reasons. Thus, our proposed CAT has the clear advantage of attaining the required size α . Though we have presented the results for $\alpha = 0.05$ here, similar trends have been observed for $\alpha = 0.10$ also.

Remark 4.2 Grice and Bain (1980) provided a table where one can start with a different nominal level α_0 (as the target level) so that the SBT ends up achieving the size approximately equal to α . For example, when $n_1 = n_2 = 10$ and our nominal level $\alpha = 0.05$, note that SBT’s actual size is consistently higher and is about 0.08. Therefore, to apply the SBT properly, one should start with a different nominal level, say $\alpha_{SBT} (<\alpha)$, as target so that the actual size comes around 0.05. We have found that Grice and Bain (1980) table of α_{SBT} ($= \alpha_0$, according to their notation) is not very accurate (may be due to relatively less sophisticated computational resources available at that time). For $n_1 = n_2 = 10$ and $\alpha = 0.05$, α_{SBT} should be 0.028, not 0.038 as suggested by Grice and Bain (1980). The Table 3 gives the size and power of SBT

Table 3 Power and size of SBT for $n_1 = n_2 = 10$ with $\alpha = \alpha_{\text{SBT}} = 0.028$ (Case-I)

δ	λ_2					
	0.5	0.75	1.0	1.25	2.0	5.0
2.5	0.634	0.159	0.049*	0.113	0.632	1.000
5.0	0.903	0.273	0.051*	0.185	0.911	1.000
7.5	0.979	0.388	0.050*	0.252	0.979	1.000
10.0	0.997	0.486	0.049*	0.313	0.997	1.000

* indicates size of a test

for target level $\alpha_{\text{SBT}} = 0.028$. After the size correction, the power of SBT is almost identical, if not marginally inferior, to the power of CAT. It would have been better to provide the power of size corrected SBT together with those of the other tests in Tables 1 and 2 (instead of the separate Table 3). But obtaining each α_{SBT} , which is mildly dependent on δ , but heavily dependent upon sample sizes, is a tedious task. Each α_{SBT} must be obtained by trial and error method, and hence time consuming. Therefore, Table 3 is presented as a demonstration which provides the required size and corresponding power of SBT for $n_1 = n_2 = 10$ and $\alpha = 0.05$. While the existing tests may appear to have higher power than CAT, it is due to the fact that the other tests are liberal to begin with, and when they are adjusted for size, then the CAT is almost as good as, if not marginally better than, the other tests. This can be seen by comparing the results of Tables 1 and 3.

Remark 4.3 We now extend our computations for $k = 3$. Note that [Shiue and Bain \(1983\)](#) method is not available for three populations. Therefore, in Table 4 we provide the size and power values for CAT and LRT only. This table is made for $\lambda_1 = 1$, and various combinations of (λ_2, λ_3) . Again, the CAT is found to be attaining the nominal level very closely whereas the LRT is a liberal one in terms of size (attained at $(\lambda_2, \lambda_3) = (1, 1)$). Since the trend remains same for other sample sizes, we have presented only limited results in Table 4.

Case-II: δ_i 's are all unknown and possibly unequal

First we consider the results for only two populations (i.e., $k = 2$). Here we have now four test methods available, namely—CAT, LRT, SBT and [Tripathi et al. \(1993\)](#) test (henceforth denoted by ‘TGPT’). The following tables provide the power and size of these tests for various combinations of (δ_1, δ_2) and λ_2 , where λ_1 is taken as 1.0 for convenience. Note that, in general, the size is attained when $\delta_2 = \delta_1 (\lambda_2/\lambda_1)$ (and marked by a ‘*’). This is reported in Tables 5 and 6 for $\alpha = 0.05$. Similar results hold for $\alpha = 0.10$, but not reported here.

The results for three populations (i.e., $k = 3$) are provided in Table 7. Again for more than two populations we don’t have [Shiue and Bain \(1983\)](#) method, and hence the comparison boils down to the remaining three tests.

Remark 4.4 The results of Tables 5–7 again reveal that the CAT appears to have the best size, in terms of attaining the nominal level. Surprisingly, the worst performers are SBT and TGPT. With $\alpha = 0.05$, the size of TGPT can be as high as 0.09

Table 4 Power of two tests for $k = 3$ and $n_1 = n_2 = n_3$ with $\alpha = 0.05$ (Case-I)

(n_1, n_2, n_3)	δ	Power	(λ_2, λ_3)					
			(0.1, 0.1)	(0.1, 1)	(0.5, 1)	(1, 1)	(1, 2)	(1, 5)
(10, 10, 10)	2.5	β_{CAT}	1.000	1.000	0.650	0.050*	0.659	1.000
		β_{LRT}	1.000	1.000	0.725	0.068*	0.700	1.000
	5.0	β_{CAT}	1.000	1.000	0.930	0.050*	0.934	1.000
		β_{LRT}	1.000	1.000	0.951	0.069*	0.947	1.000
	7.5	β_{CAT}	1.000	1.000	0.990	0.048*	0.993	1.000
		β_{LRT}	1.000	1.000	0.994	0.065*	0.994	1.000
10.0	β_{CAT}	1.000	1.000	0.998	0.050*	1.000	1.000	
	β_{LRT}	1.000	1.000	0.999	0.069*	1.000	1.000	
(25, 25, 25)	2.5	β_{CAT}	1.000	1.000	0.979	0.049*	0.984	1.000
		β_{LRT}	1.000	1.000	0.984	0.055*	0.984	1.000
	5.0	β_{CAT}	1.000	1.000	1.000	0.050*	1.000	1.000
		β_{LRT}	1.000	1.000	1.000	0.056*	1.000	1.000
	7.5	β_{CAT}	1.000	1.000	1.000	0.051*	1.000	1.000
		β_{LRT}	1.000	1.000	1.000	0.057*	1.000	1.000
	10.0	β_{CAT}	1.000	1.000	1.000	0.048*	1.000	1.000
		β_{LRT}	1.000	1.000	1.000	0.055*	1.000	1.000

* indicates size of a test

(see Table 6, $(n_1, n_2) = (10, 25)$). This was overlooked by [Tripathi et al. \(1993\)](#), and only large sample results were provided by these authors to justify their method.

Remark 4.5 All the numerical results provided in this paper have been obtained by the programming language R which is slow but appears to work well. We also tried Fortran which works faster, but runs into difficulties when $k > 2$. It seems that for $k > 2$, solving the system of Eqs. (1.4) and/or (3.3)–(3.4) using Fortran is problematic. Therefore, to maintain uniformity throughout the paper, we have relied on R.

5 Applications with real life datasets

It is customary to assume normality to compare several population means, especially in the form of traditional one-way ANOVA. But the normality assumption may not be valid, and/or questionable when the observable variable is nonnegative and/or standard normality tests may not be reliable due to very small sample sizes. This happens in hydrological studies where datasets tend to have positively skewed distribution. In such a situation it would be a more pragmatic approach to assume gamma distributions for the underlying datasets. Also, a gamma distribution can incorporate shapes ranging from heavily positively skewed to almost symmetric through its shape parameter which can give an applied researcher more leverage as shown in the following three examples. These examples demonstrate how the results from a gamma

Table 5 Power of four tests for $k = 2$ and $\alpha = 0.05$ (Case-II)

(n_1, n_2)	(δ_1, δ_2)	Power	λ_2					
			0.5	0.75	1.0	1.25	2.0	5.0
(10, 10)	(1, 2)	β_{CAT}	0.920	0.679	0.410	0.225	0.046*	0.582
		β_{LRT}	0.928	0.699	0.431	0.241	0.059*	0.673
		β_{SBT}	0.934	0.715	0.449	0.258	0.073*	0.706
		β_{TGPT}	0.954	0.774	0.511	0.305	0.063*	0.472
	(1, 5)	β_{CAT}	0.999	0.999	0.994	0.965	0.694	0.049*
		β_{LRT}	1.000	0.998	0.993	0.968	0.703	0.065*
		β_{SBT}	1.000	0.999	0.993	0.970	0.716	0.077*
		β_{TGPT}	1.000	1.000	0.999	0.996	0.868	0.095*
	(2.5, 5)	β_{CAT}	1.000	0.967	0.768	0.450	0.050*	0.931
		β_{LRT}	1.000	0.972	0.789	0.480	0.065*	0.958
		β_{SBT}	1.000	0.976	0.810	0.509	0.077*	0.968
		β_{TGPT}	1.000	0.986	0.858	0.566	0.075*	0.924
	(5, 5)	β_{CAT}	0.903	0.278	0.053*	0.170	0.900	1.000
		β_{LRT}	0.925	0.318	0.069*	0.206	0.923	1.000
		β_{SBT}	0.935	0.353	0.082*	0.237	0.936	1.000
		β_{TGPT}	0.923	0.334	0.077*	0.222	0.924	1.000
	(5, 10)	β_{CAT}	1.000	0.998	0.967	0.574	0.051*	0.997
		β_{LRT}	1.000	0.999	0.973	0.627	0.064*	0.999
		β_{SBT}	1.000	0.999	0.978	0.673	0.080*	0.999
		β_{TGPT}	1.000	0.999	0.986	0.721	0.079*	0.998
(10, 20)	(1, 2)	β_{CAT}	0.962	0.741	0.467	0.247	0.049*	0.598
		β_{LRT}	0.965	0.746	0.476	0.255	0.067*	0.740
		β_{SBT}	0.967	0.757	0.498	0.269	0.077*	0.775
		β_{TGPT}	0.994	0.878	0.656	0.396	0.081*	0.467
	(1, 5)	β_{CAT}	1.000	1.000	0.995	0.975	0.731	0.043*
		β_{LRT}	1.000	1.000	0.996	0.977	0.740	0.064*
		β_{SBT}	1.000	1.000	0.996	0.979	0.752	0.074*
		β_{TGPT}	1.000	1.000	1.000	0.999	0.917	0.100*
	(2.5, 5)	β_{CAT}	1.000	0.987	0.825	0.504	0.053*	0.944
		β_{LRT}	1.000	0.990	0.839	0.533	0.068*	0.977
		β_{SBT}	1.000	0.992	0.853	0.567	0.082*	0.983
		β_{TGPT}	1.000	0.998	0.920	0.666	0.083*	0.938
	(5, 5)	β_{CAT}	0.967	0.359	0.053*	0.228	0.958	1.000
		β_{LRT}	0.973	0.388	0.065*	0.276	0.975	1.000
		β_{SBT}	0.979	0.414	0.078*	0.307	0.979	1.000
		β_{TGPT}	0.987	0.454	0.074*	0.256	0.963	1.000
	(5, 10)	β_{CAT}	1.000	0.998	0.982	0.792	0.054*	1.000
		β_{LRT}	1.000	0.998	0.986	0.819	0.071*	1.000
		β_{SBT}	1.000	0.999	0.991	0.845	0.086*	1.000
		β_{TGPT}	1.000	1.000	0.997	0.895	0.089*	1.000

* indicates size of a test

Table 6 Power of four tests for $k = 2$ and $\alpha = 0.05$ (Case-II)

(n_1, n_2)	(δ_1, δ_2)	Power	λ_2					
			0.5	0.75	1.0	1.25	2.0	5.0
(20, 10)	(1, 2)	β_{CAT}	0.989	0.846	0.560	0.301	0.053*	0.787
		β_{LRT}	0.992	0.864	0.592	0.323	0.061*	0.811
		β_{SBT}	0.993	0.874	0.609	0.340	0.069*	0.829
		β_{TGPT}	0.984	0.836	0.556	0.310	0.059*	0.787
	(1, 5)	β_{CAT}	0.997	1.000	1.000	1.000	0.928	0.053*
		β_{LRT}	1.000	1.000	1.000	1.000	0.925	0.060*
		β_{SBT}	1.000	1.000	1.000	1.000	0.927	0.065*
		β_{TGPT}	1.000	1.000	1.000	1.000	0.958	0.064*
	(2.5, 5)	β_{CAT}	1.000	0.997	0.915	0.614	0.054*	0.992
		β_{LRT}	1.000	0.998	0.930	0.648	0.065*	0.993
		β_{SBT}	1.000	0.999	0.939	0.669	0.074*	0.994
		β_{TGPT}	1.000	0.998	0.926	0.653	0.068*	0.994
	(5, 5)	β_{CAT}	0.960	0.344	0.052*	0.234	0.971	1.000
		β_{LRT}	0.978	0.409	0.064*	0.256	0.976	1.000
		β_{SBT}	0.982	0.444	0.076*	0.282	0.982	1.000
		β_{TGPT}	0.965	0.377	0.071*	0.308	0.989	1.000
	(5, 10)	β_{CAT}	1.000	1.000	0.997	0.887	0.046*	1.000
		β_{LRT}	1.000	1.000	0.998	0.905	0.056*	1.000
		β_{SBT}	1.000	1.000	0.998	0.917	0.067*	1.000
		β_{TGPT}	1.000	1.000	0.998	0.910	0.063*	1.000
(10, 25)	(1, 2)	β_{CAT}	0.967	0.763	0.484	0.263	0.047*	0.603
		β_{LRT}	0.968	0.769	0.494	0.269	0.064*	0.769
		β_{SBT}	0.970	0.776	0.510	0.282	0.077*	0.804
		β_{TGPT}	0.995	0.909	0.683	0.441	0.086*	0.483
	(1, 5)	β_{CAT}	1.000	1.000	0.996	0.977	0.740	0.045*
		β_{LRT}	1.000	1.000	0.997	0.979	0.746	0.067*
		β_{SBT}	1.000	1.000	0.997	0.978	0.758	0.082*
		β_{TGPT}	1.000	1.000	1.000	0.998	0.923	0.110*
	(2.5, 5)	β_{CAT}	1.000	0.990	0.847	0.531	0.049*	0.948
		β_{LRT}	1.000	0.992	0.861	0.558	0.064*	0.981
		β_{SBT}	1.000	0.993	0.880	0.586	0.081*	0.986
		β_{TGPT}	1.000	0.998	0.942	0.706	0.088*	0.943
	(5, 5)	β_{CAT}	0.971	0.370	0.052*	0.233	0.964	1.000
		β_{LRT}	0.977	0.401	0.067*	0.284	0.978	1.000
		β_{SBT}	0.982	0.431	0.080*	0.317	0.982	1.000
		β_{TGPT}	0.991	0.485	0.079*	0.258	0.968	1.000
	(5, 10)	β_{CAT}	1.000	1.000	0.987	0.794	0.049*	0.999
		β_{LRT}	1.000	1.000	0.990	0.821	0.073*	1.000
		β_{SBT}	1.000	1.000	0.993	0.847	0.090*	1.000
		β_{TGPT}	1.000	1.000	0.998	0.902	0.089*	0.999

* indicates size of a test

Table 7 Power of three tests for $k = 3$ with $\alpha = 0.05$ (Case-II)

(n_1, n_2, n_3)	(δ_2, δ_3)	Power	(λ_2, λ_3)						
			(0.1, 0.1)	(0.1, 1)	(0.5, 1)	(1, 1)	(1, 2)	(1, 5)	(1, 10)
(10, 15, 20)	(2.5, 5)	β_{CAT}	1.000	1.000	0.044*	0.847	0.915	1.000	0.935
		β_{LRT}	1.000	1.000	0.061*	0.797	0.953	1.000	1.000
		β_{TGPT}	1.000	1.000	0.086*	0.918	0.931	1.000	1.000
	(5, 10)	β_{CAT}	0.990	1.000	0.909	0.986	0.038*	1.000	1.000
		β_{LRT}	1.000	1.000	0.856	1.000	0.065*	1.000	1.000
		β_{TGPT}	1.000	1.000	0.974	1.000	0.089*	1.000	1.000

* indicates size of a test

model compare with those from a normal model, being very similar to not so similar.

Even though Sects. 3 and 4 consider the two cases (Case-I: equal shape parameters; Case-II: unequal shape parameters), one needs to decide which case to follow for a given dataset. Our proposed CAT can be used again to decide which case does the given dataset come under.

One can think of testing first $H'_0 : \delta_i = \delta (\forall i)$ versus $H'_A : \delta_i \neq \delta_j$, for $i \neq j$; i.e., test $H'_0 : \eta = \sum_{i=1}^k (\gamma_i - \bar{\gamma})^2 = 0$ versus $H'_A : \eta > 0$, where $\gamma_i = \ln(\delta_i)$, $i = 1, 2, \dots, k$, $\bar{\gamma} = \sum_{i=1}^k \gamma_i / k$.

(a) Step-1: Obtain $\hat{\lambda}_{i(ML)}$, $\hat{\delta}_{i(ML)}$ by solving the system of equations in (1.4). Get $\hat{\eta}_{ML}$.

(b) Step-2:

(i) Assume that H'_0 is true; i.e., $\delta_i = \delta \forall i$. Find the ‘restricted MLEs’ $(\hat{\lambda}_{i(RML)}, i = 1, 2, \dots, k, \hat{\delta}_{RML})$ of $(\lambda_1, \dots, \lambda_k, \delta)$ by solving the following system of equations (with $w_i = n_i / \sum_{i=1}^k n_i$):

$$\sum_{i=1}^k w_i \ln \lambda_i - \psi(\delta) + \sum_{i=1}^k w_i \ln \tilde{X}_i = 0; \quad \text{and} \quad \lambda_i = \delta / \tilde{X}_i, \quad 1 \leq i \leq k. \tag{5.1}$$

(ii) Generate artificial sample X_{i1}, \dots, X_{in_i} iid from $G(\hat{\delta}_{RML}, \hat{\lambda}_{i(RML)})$, $1 \leq i \leq k$; then pretend that it is from $G(\delta_i, \lambda_i)$, $1 \leq i \leq k$, respectively, and proceed to estimate the parameters (δ_i, λ_i) , $1 \leq i \leq k$ using (1.4). This process needs to be repeated M times.

In the l th replication, $1 \leq l \leq M$, the estimated parameters (δ_i, λ_i) (based on the artificial sample) are denoted by $\hat{\delta}_{0i}^{(l)}, \hat{\lambda}_{0i}^{(l)}$. Then calculate $\hat{\eta}$ as $\hat{\eta}_{0l} = \sum_{i=1}^k (\hat{\gamma}_{0i}^{(l)} - \hat{\gamma}_0^{(l)})^2$ where $\hat{\gamma}_{0i}^{(l)} = \ln(\hat{\delta}_{0i}^{(l)})$ and $\hat{\gamma}_0^{(l)} = \sum_{i=1}^k \hat{\gamma}_{0i}^{(l)} / k$.

(iii) Order the $\hat{\eta}_{0l}$ values as $\hat{\eta}_{0(1)} \leq \hat{\eta}_{0(2)} \leq \dots \leq \hat{\eta}_{0(M)}$.

(c) Step-3: The original $\hat{\eta}_{ML}$ (in Step-1) is now compared with the cut-off point $\hat{\eta}_U = \hat{\eta}_{0((1-\alpha)M)}$. Alternatively, the p -value is $\approx \sum_{l=1}^M I(\hat{\eta}_{0(l)} > \hat{\eta}_{ML}) / M$.

Table 8 Virginia ground water well yield data with and without rock fractures

Without fractures	With fractures
0.001, 0.003, 0.007, 0.020,	0.020, 0.031, 0.086, 0.130,
0.030, 0.040, 0.041, 0.077,	0.160, 0.160, 0.180, 0.300,
0.100, 0.454, 0.490, 1.020	0.400, 0.440, 0.510, 0.720, 0.950

Example 5.1 [Wright \(1985\)](#) reported the following ground water yield (in gal/min/ft) data from two types of wells in southwestern Virginia. These are small datasets: 12 wells were observed from valleys underlain by unfractured rocks, and 13 wells from valleys underlain by fractured rocks (see Table 8).

It is tempting to use the standard two normal populations technique (with or without common standard deviation) to see if there is any difference in mean yields. But a simple histogram of each sample shows a heavy positive skewness. Therefore, the normality assumption is not tenable. In fact, a $Q-Q$ plot for each sample shows a heavy departure from normality. Hence, it would be appropriate to assume Gamma distribution for each sample.

First, we need to decide whether the shape parameters are equal or not. For the above dataset, the four estimated parameters are $\hat{\delta}_1 = 0.4342$, $\hat{\lambda}_1 = 2.2824$, $\hat{\delta}_2 = 1.1854$, $\hat{\lambda}_2 = 3.7707$. By applying the CAT above with $M = 10^5$ to test the equality of shape parameters, we get the p -value ≈ 0.0596 which is not too large or not too small either.

Next, we choose the Case-I, and apply the corresponding CAT to test the equality of means. The resultant p -value (to compare the means) is ≈ 0.3341 . Note that, if one carries out a similar test assuming normality erroneously, then the p -value (with pooled standard deviation) comes out to be 0.3065, which is large enough to conclude that the mean yields are same. If one assumes Case-II gamma model, then the p -value for testing the equality of means comes out to be 0.3441. Hence we conclude that the means water yields from the two types of wells are same at 5 or 10% level.

Example 5.2 In a comprehensive monitoring study on availability of ground water to neighboring domestic wells near Bel Air, Harford county, Maryland, [Duigon and Cooper \(1999\)](#) obtained data from various types of wells. The following dataset, extracted from the original one, provides the well yields (or ‘specific capacity’ in gal/min/ft) based on four topographic settings: G = Flood Plain, H = Hilltop, S = Hillside and W = Upland draw.

The dataset (in Table 9) is heavily unbalanced with very few observations under the last two categories (W and G). Individual sample histograms for the other two categories (S and H) show positive skewness. Therefore, assuming gamma model for each category seems plausible.

By applying the CAT above with $M = 10^5$ to test the equality of shape parameters, we get the p -value ≈ 0.9061 . Next, we choose the Case-I, and apply the corresponding CAT again to test the equality of means. The resultant p -value (to compare the means) is ≈ 0.1034 . On the other hand, a one-way homoscedastic ANOVA, to test the

Table 9 Maryland ground water well data based on topography

S	H	W	G
0.22, 1.33, 0.75, 0.18, 0.01, 0.16, 0.28, 0.87, 0.02, 0.10, 0.03, 0.05, 0.86, 5.00, 0.04, 4.00, 0.37, 0.38, 0.11, 0.10, 0.02, 0.01, 0.05, 0.17, 0.46, 0.16, 1.33, 0.14, 2.86, 0.13, 7.50, 4.50, 0.03, 0.003, 0.05, 0.02, 0.04, 0.75, 0.52, 5.00, 0.35 (+)	0.12, 0.15, 0.03, 0.08, 0.07, 0.12, 0.01, 0.03, 0.02, 0.01, 1.94, 0.21, 0.06, 0.83, 0.15, 0.10, 0.14	0.43, 0.22, 0.22, 8.00	0.44, 0.01, 0.12

(+) One observation (=0.00) was dropped since no pumping was done. The entry 0.003 is approximate

Table 10 Sierra Nevada dataset on chloride concentration in water

Rock type	Concentration										
Granodiorite	6.0	0.5	0.4	0.7	0.8	6.0	5.0	0.6	1.2		
	0.3	0.2	0.5	0.5	10.0	0.2	0.2	1.7	3.0		
Quartz Monzonite	1.0	0.2	1.2	1.0	0.3	0.1	0.1	0.4	3.2		
	0.3	0.4	1.8	0.9	0.1	0.2	0.3	0.5			

equality of four means, yields a p -value of 0.1475 which is not too different from the gamma model result. Therefore, it is concluded that mean well yields from the four topographic settings are same at 5 or 10% level.

Example 5.3 Ferth et al. (1964) measured chemical compositions of water in springs draining differing rock types. The dataset in Table 10 provides chloride concentration (in mg/litre) in water samples from two rock types in Sierra Nevada, California and Nevada.

First, we need to decide whether the shape parameters are equal or not. For the above dataset, the four estimated parameters are $\hat{\delta}_1 = 0.7594, \hat{\lambda}_1 = 0.3616, \hat{\delta}_2 = 1.1359, \hat{\lambda}_2 = 1.6092$. By applying the CAT above with $M = 10^5$ to test the equality of shape parameters, we get the p -value ≈ 0.3593 .

Next, we choose the Case-I, and apply the corresponding CAT again to test the equality of two means. The resultant p -value (to compare the means) is ≈ 0.0066 , small enough to conclude that means are different. (Note that, if one carries out a similar test assuming normality erroneously, then the p -value (with pooled standard deviation) comes out to be 0.0567, which is large enough at 5% level to conclude that the mean concentrations are same).

6 Further comments on CAT

The success of our proposed CAT in terms of nearly attaining the nominal level and maintaining a high level of power compared to the other tests depends on the selection of η (which has been discussed thoroughly in Remark 3.2), and the resampling done from the restricted model, under H_0 , where each nuisance parameter is replaced by its RMLE (see Sect. 3.1, Step-2(i)). One may wonder whether the usual (unrestricted) MLE of the nuisance parameters could have done a better job as this has been the topic of debate in the literature.

Hall and Wilson (1991) highlighted two important guidelines for bootstrap hypothesis testing. Their first guideline states that when one wants to estimate the critical value, resampling must be done in a way that reflects the null hypothesis. This must be done regardless of the distribution that generated the data. This guideline is crucial to the success of bootstrap hypothesis testing and has been mentioned by other researchers also [see, for example, Martin (2007) and Allison (2008)]. In their second guideline, Hall and Wilson (1991) recommended that bootstrap hypothesis tests should be based on test statistics that are (asymptotically) pivotal.

In our proposed CAT, deriving the RMLE under H_0 , and using this RMLE for resampling is meant to follow the above first guideline put forward by Hall and Wilson (1991). Though our test statistic $\hat{\eta}_{ML}$ is not scaled, to make it a proper pivot, the scaling factor remains hidden in the critical value(s) through the generated $\hat{\eta}_{0l}$ values (see Sect. 3.1, Step-2(ii)). One may find a suitable scaling factor to scale $\hat{\eta}_{ML}$, but that is not needed for our CAT, and hence no extra effort is needed to find such a suitable scaling factor.

One may wonder whether the resampling ought to have been done using the MLE, rather than the RMLE, of the nuisance parameters in the restricted model under H_0 . This is based on the practice suggested by Hall and Wilson (1991) first guideline. In our opinion, this is counter-intuitive for several reasons.

First of all, using (unrestricted) MLE, rather than the RMLE, goes very much against Hall and Wilson (1991) first guideline. One can see the follow-up comments by Tibshirani (1991) where he agrees with the first guideline, but criticizes Hall and Wilson (1991) practice of using the aforementioned (unrestricted) MLE saying—“They are right, but do not follow their own advice in the article.” About Hall and Wilson (1991) first guideline Martin (2007) stresses that—“... when constructing tests of a certain size, it is fundamental to the procedure that the distribution of the test statistic *assuming that the null hypothesis holds* be considered. ... The first recommendation—resampling under null—is certainly appropriate, but its application usually needs careful thought as to how such sampling might actually be conducted.”

Secondly, it has to be kept in mind that, Hall and Wilson (1991) (and others’) suggestions for bootstrap hypothesis testing are for a general nonparametric setup. The suggested guidelines are supposed to work asymptotically, and no way do they ensure good size and/or power behavior for small to moderate sample sizes. In our parametric gamma model, where exact sampling distributions are impossible to obtain for small to moderate sample sizes, the proposed CAT works remarkably well as demonstrated by our comprehensive simulation results. For large sample sizes all

available tests are almost identical in terms of size and power. Hence, for large sample sizes, one does not need to do any parametric bootstrap since the LRT can do the hypothesis testing reasonably well. Our proposed CAT has been motivated by the real-life applications, is shown in Sect. 5, where small samples are quite common.

Thirdly, we would like to use one of the examples used by Martin (2007) to show why (unrestricted) MLE, instead of RMLE, can lead to absurd results. We refer to Martin (2007) Case-3 with added parametric model restrictions. Assume $X_{i1}, \dots, X_{in_i} \text{ iid } N(\mu_i, \sigma^2)$, $i = 1, 2$, where our goal is to test $H_0 : \mu_1 = \mu_2$ versus $H_A : \mu_1 \neq \mu_2$. Note that $\hat{\sigma}_{ML}^2 = S/(n_1 + n_2)$ and $\hat{\mu}_{i(ML)} = \bar{X}_i = \sum_{j=1}^{n_i} X_{ij}/n_i$, where $S = S_1 + S_2$ and $S_i = \sum_1^{n_i} (X_{ij} - \bar{X}_i)^2$. Let us first see what happens if the (unrestricted) MLE of all parameters are used for our resampling scheme. In Step-1, $\hat{\eta}_{ML} = (\hat{\mu}_{1(ML)} - \hat{\mu}_{2(ML)})$. In Step-2, l th replication would yield $X_{i1}^{(l)}, \dots, X_{in_i}^{(l)} \text{ iid } N(\hat{\mu}_{i(ML)}, \hat{\sigma}_{ML}^2)$, which in turn gives $\hat{\eta}_{0l} = (\bar{X}_1^{(l)} - \bar{X}_2^{(l)}) \sim N(\hat{\mu}_{1(ML)} - \hat{\mu}_{2(ML)}, \hat{\sigma}_{ML}^2(1/n_1 + 1/n_2))$. Thus, after generating $\hat{\eta}_{01}, \hat{\eta}_{02}, \dots, \hat{\eta}_{0M}$, for sufficiently large M , the critical values $\hat{\eta}_L$ and $\hat{\eta}_U$ are nothing but $\{(\hat{\mu}_{1(ML)} - \hat{\mu}_{2(ML)}) - z_{(\alpha/2)}\hat{\sigma}_{ML}\sqrt{(1/n_1 + 1/n_2)}\}$ and $\{(\hat{\mu}_{1(ML)} - \hat{\mu}_{2(ML)}) + z_{(\alpha/2)}\hat{\sigma}_{ML}\sqrt{(1/n_1 + 1/n_2)}\}$ respectively. Therefore, $P(\text{Accept } H_0) = P(\hat{\eta}_L < \hat{\eta}_{ML} < \hat{\eta}_U) = 1$ always, which is absurd. This absurdity takes place because not only did we not enforce $H_0 : \mu_1 = \mu_2$ for the location parameters, but also we did not take H_0 into consideration for estimating σ^2 before resampling. Now, let us enforce H_0 only for the location parameters, but not for σ^2 . That means, in Step-2, l th replication would yield $X_{i1}^{(l)}, \dots, X_{in_i}^{(l)} \text{ iid } N(\bar{X}_{..}, \hat{\sigma}_{ML}^2)$, where $\bar{X}_{..}$ is the grand sample mean of μ where $H_0 : \mu_1 = \mu_2 = \mu$ holds. In this situation, $\hat{\eta}_{0l} = (\bar{X}_1^{(l)} - \bar{X}_2^{(l)}) \sim N(0, \hat{\sigma}_{ML}^2(1/n_1 + 1/n_2))$, $1 \leq l \leq M$. For sufficiently large M , $\hat{\eta}_L \approx -z_{(\alpha/2)}\hat{\sigma}_{ML}\sqrt{(1/n_1 + 1/n_2)}$ and $\hat{\eta}_U \approx z_{(\alpha/2)}\hat{\sigma}_{ML}\sqrt{(1/n_1 + 1/n_2)}$. The power of the test is then $P(\text{Reject } H_0) = P(|\bar{X}_1 - \bar{X}_2| > z_{(\alpha/2)}\hat{\sigma}_{ML}\sqrt{(1/n_1 + 1/n_2)})$, which can be simplified after several steps as: $P(F > c)$ where the random variable $F \sim F(1, n_1 + n_2 - 2 | \lambda) =$ non-central F -distribution with dfs 1 and $(n_1 + n_2 - 2)$, and noncentrality parameter $\lambda = (\mu_1 - \mu_2)^2 / \{\sigma^2(1/n_1 + 1/n_2)\}$; the constant c is given as $c = (z_{(\alpha/2)})^2(n_1 + n_2 - 2)/(n_1 + n_2)$. Though this power function is monotonically increasing in λ , i.e., as one deviates from H_0 , the size of the test is not very satisfactory for small to moderate sample sizes. For example, with $\alpha = 0.05$, (i) $n_1 = n_2 = 5$ yields $P(F > c) \cong 0.12$; and (ii) $n_1 = n_2 = 10$ yields $P(F > c) \cong 0.08$; that means, using unrestricted MLE of σ^2 gives a very liberal test. Here we have used the normal example because the size of the CAT can be obtained explicitly (as $P(F > c)$ for sufficiently large M) which is found to be quite liberal for small samples. Similar pattern also holds for the gamma case, but needs to be seen only by simulation.

We do not claim that the proposed CAT is the ‘‘best’’ test procedure. However, for the given gamma problem, it works fine compared to the other existing tests. If, for a given problem, one has an ‘optimal’ test, then there won’t be any need for the CAT. On the other hand, if, for a given problem, no ‘optimal’ test is known, then the proposed CAT is worth a try as also shown in Chang and Pal (2008).

7 Concluding remark

The proposed CAT, though computationally intensive, is a very useful procedure to handle any dataset as long as we have any parametric model. It is quite versatile in the sense that CAT can be used for any parameter of interest. For example, under the gamma model we can test the equality of medians (instead of means) by defining η properly. One just needs to make sure that the underlying simulation and computations can be done effectively and accurately. Currently we are applying our CAT for testing the equality of several Weibull and lognormal means, and it will be reported in due course. Also, a copy of our simulation written in R can be made available to any interested reader.

Acknowledgments We sincerely thank the three anonymous referees for going over our earlier manuscript meticulously. Their comments and suggestions have helped us significantly in improving the presentation of this paper. The second author's research has been supported partially by a research grant from the National Science Council (98-2118-M-130-002).

References

- Aksoy H (2000) Use of gamma distribution in hydrological analysis. *Turk J Eng Environ Sci* 24(6):419–428
- Allison JS (2008) Bootstrap-based hypothesis testing. Ph.D. Thesis, North-West University, Potchefstroom, South Africa
- Bougeault P (1982) Cloud-ensemble relations based on the ability distribution for the higher order models of the planet layer. *J Atmos Sci* 39:2691–2700
- Chang CH, Pal N (2008) A computational approach to one-way ANOVA and its model robustness. *Far East J Theor Stat* 25(1):111–134
- Coe R, Stern RD (1982) Fitting models to daily rainfall data. *J Appl Meteorol* 21:1024–1031
- Coles B, Vrbik P, Giacometti RD, Rothstein SM (2008) Gamma distribution model to provide a direct assessment of the overall quality of quantum monte carlo-generated electron distributions. *J Phys Chem* 112(10):2012–2017
- Davison AC, Hinkley DV (1997) *Bootstrap methods and their application*. Cambridge University Press, Cambridge
- Duigon MT, Cooper BF (1999) Impact of a public water supply well on availability of ground water to neighboring domestic wells near bel air, Maryland. Interim Technical Report, Department of Natural Resources, Maryland Geological Survey (www.mgs.md.gov/hydro/pub/pdf/impact.pdf)
- Efron B (1982) The Jackknife, Bootstrap, and other resampling plans. In: CBMS-NSF regional conference series in applied mathematics, vol 38. Society for Industrial and Applied Mathematics, Philadelphia
- Ferth JH, Roberson CE, Polzer WL (1964) Sources of mineral constituents in water from granite rocks, Sierra Nevada, California and Nevada. US Geological Survey Water Supply Paper 1535-I
- Greenwood JA, Durand D (1960) Aids for fitting the gamma distribution by maximum likelihood. *Technometrics* 2:55–65
- Grice JV, Bain LJ (1980) Inferences concerning the mean of the gamma distribution. *J Am Stat Assoc* 75:929–933
- Hall P, Wilson SR (1991) Two guidelines for Bootstrap hypothesis testing. *Biometrics* 47:757–762
- Hjorth JSU (1994) *Computer intensive statistical methods: validation, model selection and Bootstrap*. Chapman and Hall, London
- Li LW, Kooi PS, Leong MS, Yeo TS (2007) A gamma distribution of raindrop sizes and its application to Singapore's tropical environment. *Microw Opt Technol Lett* 7(5):253–257
- Martin MA (2007) Bootstrap hypothesis testing for some common statistical problems: a critical evaluation of size and power properties. *Comput Stat Data Anal* 51:6321–6342
- McDonald JB, Jensen BC (1979) An analysis of estimators of alternative measures of income inequality associated with the gamma distribution function. *J Am Stat Assoc* 74:856–860

- Ming Tan C, Raghavan N, Roy A (2007) Application of gamma distribution in electromigration for submicron interconnects. *J Appl Phys* 102(10):3703–3711
- Shiue WK, Bain LJ (1983) A two-sample test of equal gamma distribution scale parameters with unknown common shape parameter. *Technometrics* 25(4):377–381
- Shiue WK, Bain LJ, Engelhardt M (1988) Test of equal gamma-distribution means with unknown and unequal shape parameters. *Technometrics* 30:169–174
- Simpson J (1972) Use of the gamma distribution in single-cloud rainfall analysis. *Mon Weather Rev* 100(4):309–312
- Singh VP, Singh K (1985) Derivation of the gamma distribution by using the principle of maximum entropy (POME). *Water Resour Bull* 21:941–952
- Singh SK (2004) Simplified use of gamma-distribution/nash model for runoff modeling. *J Hydrol Eng* 9(3):240–243
- Tibshirani R (1991) Bootstrap hypothesis testing (correspondence). *Biometrics* 47:969–970
- Tripathi RC, Gupta RC, Pair RK (1993) Statistical tests involving several independent gamma distributions. *Ann Inst Stat Math* 45(4):773–786
- Wright WG (1985) Effects of fracturing on well yields in the coal field areas of wise and dickenson counties, southwestern Virginia; US geological survey water resources investigations report, pp 85–4061