Communications in Statistics—Simulation and Computation®, 44: 1339–1363, 2015

ISSN: 0361-0918 print / 1532-4141 online
DOI: 10.1080/03610918.2013.818692

Small Sample Tests for Shape Parameters of Gamma Distributions

DULAL K. BHAUMIK,[1,2,6] KUSH KAPUR,[3,6]
NARAYANASWAMY BALAKRISHNAN,[4]
JEROME P. KEATING,[5] AND ROBERT D. GIBBONS[6]

[1]Department of Biostatistics and Psychiatry, University of Illinois at Chicago, Chicago, Illinois, USA
[2]Cooperative Studies Program Coordinating Center (151K), Hines VA Hospital, Hines, Illinois, USA
[3]Clinical Research Center and Department of Neurology, Boston Children's Hospital, Harvard Medical School, Boston, Massatusetts, USA
[4]Department of Mathematics and Statistics, McMaster University, Hamilton, Ontario, Canada
[5]Management Science and Statistics, The University of Texas at San Antonio, One UTSA Circle, San Antonio, Texas, USA
[6]Center for Health Statistics, University of Chicago, Chicago, Illinois, USA

The introduction of shape parameters into statistical distributions provided flexible models that produced better fit to experimental data. The Weibull and gamma families are prime examples wherein shape parameters produce more reliable statistical models than standard exponential models in lifetime studies. In the presence of many independent gamma populations, one may test equality (or homogeneity) of shape parameters. In this article, we develop two tests for testing shape parameters of gamma distributions using chi-square distributions, stochastic majorization, and Schur convexity. The first one tests hypotheses on the shape parameter of a single gamma distribution. We numerically examine the performance of this test and find that it controls Type I error rate for small samples. To compare shape parameters of a set of independent gamma populations, we develop a test that is unbiased in the sense of Schur convexity. These tests are motivated by the need to have simple, easy to use tests and accurate procedures in case of small samples. We illustrate the new tests using three real datasets taken from engineering and environmental science. In addition, we investigate the Bayes' factor in this context and conclude that for small samples, the frequentist approach performs better than the Bayesian approach.

Keywords Bayes' factor; Beta distribution, Dirichlet distribution, Majorization, Schur convex functions.

Mathematics Subject Classification Primary 97K80; Secondary 62F03.

Received March 10, 2013; Accepted June 14, 2013
Address correspondence to Prof. Dulal Bhaumik, University of Illinois at Chicago, Department of Biostatistics and Psychiatry, 1601 W Taylor, M/C 912, Chicago, 60612 United States; E-mail: dbhaumik@uic.edu

1. Introduction

The development of hypothesis tests for a gamma distribution has been problematic whenever the shape parameter is unknown. Test procedures based on maximum likelihood estimators (MLE) are based on large-sample theory and involve test statistics that follow asymptotically normal distributions. Another widely used test procedure is the maximum likelihood ratio test (MLRT), in which the null distribution of a MLRT has an asymptotic chi-square distribution. For small samples, one can have little confidence that the intended nominal Type I error rate will be maintained using traditional large-sample approximations. The focus of this article is to develop new approaches to hypothesis testing that can be applied to data with small sample sizes. In particular, we consider the gamma distribution which can take on a variety of parametric forms, and includes models with increasing, decreasing, and constant hazard rates. For a complete reference of estimation and tests for gamma parameters, we refer to Johnson et al. (1994) and Bowman and Shenton (1988).

In this article, we propose several test procedures for the shape parameter(s) of two-parameter gamma distribution(s). These tests are developed under the constraint that the actual Type I error rate should not exceed the prescribed level of significance(α), even in case of small sample sizes.

1.1. *Applications of the Gamma Family*

The gamma distribution has several practical applications across many different fields. For example, Fang et al. (2007) used gamma distributions to predict life expectancy of patients with newly diagnosed HIV infection, while Manning et al. (2005) and Basu and Rathouz (2005) used it to model health care expenditure, and Whitmore and Neufeldt (2008) predicted the length of stay in hospitals by psychiatric patients. For survival analysis problems, in which endpoints often have long-tailed distributions, gamma distributions are used often to model the survival time, and have been shown to lead to increased power over normal alternatives. Davis (1952) and Barlow and Proschan (1965) explained the importance of a gamma distribution for the failure times of complex systems under continuous repair and maintenance. Das (1995), Stephenson et al. (1988), and Aksoy (2000) used the gamma distributions to model the amount of daily rainfall in a region. In environmental statistics, an important problem is to compare the average of a small number of potentially contaminated measurements to a regulatory standard, usually health-based in nature. Another problem of interest is to compare the average of a small number of potentially impacted measurements with a larger collection of background measurements. The distributions of the analytes of concern are generally right skewed and here again gamma distributions are quite suitable for analyzing these types of data (see Bhaumik and Gibbons, 2006; Gibbons et al., 2009; Krishnamoorthy et al., 2008). Bhaumik et al. (2009) constructed several small-sample tests for the mean of a gamma distribution and studied their properties. For small sample sizes, when the distributional properties cannot be easily verified, routine use of the normal distribution is often misleading. Taken as a whole, gamma distributions are quite useful for applications in many fields, including but not limited to health statistics, environmental monitoring, genetic research, and industrial quality control.

A random variable X that follows the gamma law has its density function as

$$f(x;\kappa,\beta) = \frac{x^{\kappa-1} exp(-x/\beta)}{\Gamma(\kappa)\,\beta^{\kappa}} \times I_{(0,\infty)}(x), \tag{1}$$

where $I(.)$ denotes the usual indicator function. The shape parameter κ is especially interesting to reliability engineers and survival analysts since the gamma hazard function is decreasing, constant, or increasing according to the trichotomy of $\kappa - 1$. β is the scale parameter of the gamma distribution. We use the notation $X \sim \mathcal{G}(\kappa,\ \beta)$. The mean and the variance of $X \sim \mathcal{G}(\kappa,\ \beta)$ are given by

$$E(X) = \kappa\beta \text{ and } V(X) = \kappa\beta^2.$$

Testing the shape parameter: Since the shape parameter (κ) of a gamma distribution is used to characterize the hazard function as mentioned earlier, one problem of interest is to test the null hypothesis $\kappa = \kappa_0$ against the right-sided alternative $\kappa > \kappa_0$ or the left-sided alternative $\kappa < \kappa_0$ for a specified value κ_0. A motivation for consideration of this problem arises from the fact that when setting the null hypothesis to $\kappa = 1$ it relates to testing of exponentiality against gamma alternatives with increasing failure rates and decreasing failure rates, respectively (see Keating et al., 1990). Engineers often model times of occurrence of events in the field of renewal theory using the shape parameter when the data fit the gamma distribution. The coefficient of variation which relates to measuring the efficiency of gears, blades, and deep-groove ball bearings of heavy engines is often modeled as the function of shape parameter of the gamma distribution (Bain et al., 1984). Construction of prediction and tolerance intervals using a transformed gamma random variable is another issue wherein the shape parameter plays an important role. In this context, Aryal et al. (2008) used the normal approximation to a gamma variable when $\kappa > 7$. Krishnamoorthy et al. (2008) used a normal approximation to the cube root of a gamma variable following the well-known Wilson and Hilferty (1931) approximation. This approximation is valid if $\kappa > 1$. Hence, an appropriate testing procedure for the shape parameter is necessary before one uses these approximate results.

1.2. *Small Sample Sizes*

In this section, we introduce the importance of developing methodology for inferences based on small sample sizes taken from populations that follow gamma laws. The assumption of a gamma distribution should be established from studies conducted in the areas of interest based on datasets with larger sample sizes. Examples include volatile organic compounds in ground-water monitoring systems, and the presence of amosite fibers in the environment.

1.2.1. *Rare Counts.* Inhalation of asbestos in any of its many forms is known to be carcinogenic. Chief among these maladies today is the disease, mesothelioma, which is an uncommon cancer that appears in the mesothelial cells lining the chest, abdominal cavities, and lungs. It also causes nonmalignant respiratory diseases such as asbestosis. It is attributed as a risk factor in a multitude of other cancers such as stomach, pharyngeal, laryngeal, esophageal, and colorectal cancers in several epidemiologic cohort and case-control studies. Literature reviews on asbestos raise the issue of analyst-to-analyst variation among counts within as well as between laboratories. Thus, a new specimen may be assigned an observed count by a particular analyst from a particular laboratory but the observed count may deviate significantly from the true number of fibers in the specimen. To connect observed and true counts and characterize uncertainty in these counts, Bhaumik, Kim, and Gibbons (edited by Nelson et al., 2009, pp. 93–108) extended the ideas of Gibbons and Coleman (2001) and Bhaumik and Gibbons (2006) to the case

of a Poisson random variable, which is the appropriate distribution for rare-event count data. As the cdf of the gamma family is directly related to that of the Poisson, we can apply the methodology developed here to problems in rare-event counting processes as well.

Comparison of fiber counts of the same type obtained from different samples is an important problem. In practice, samples with fewer observations (< 5) are discarded as the current methodologies fail to handle this problem efficiently. This motivates us to develop an appropriate test especially for this situation. In cases where the number of fiber counts is high, contamination from asbestos is accepted and the question reduces to quantifying the magnitude of the contamination. However, in cases where the number of fiber counts is modestly small, analysts are uncertain whether the material represents an environmental health hazard. The counts are essential in classifying a specimen as being hazardous. Small samples are subject to size-biased sampling issues (see Zelen, 1974), in that larger asbestos fibers are more readily detected. Zelen's arguments illustrated that early screening processes in breast cancer produced size-biased samples since the length of time that a tumor was of a detectable size formed a sampling weight. Thus, slower growing cancers were over-represented in general since they were inherently more likely to be detected.

1.2.2. *Material Fatigue.* Small sample sizes are quite often used in material fatigue tests of large components. For example, the FAA allows for major components such as rotor blades or aircraft wings to be qualified as flight worthy using staircase fatigue tests of samples as small as four. While the staircase fatigue test is only part of the certification, it is the actual life-test demonstration plan. Staircase fatigue tests of rotor blades require separate test facilities, a complex and expensive test apparatus, and long testing periods. Because the test apparatus is unique to the part, multiple test beds for certification of a single part are cost prohibitive. Therefore, the tests are run sequentially. In early tests, cyclic loads are set far in excess of the endurance limit of the part and time to failure is moderate. Following the staircase method, one reduces the magnitude of cyclic stress and thus prolongs the life of the part. In each successive test run, the engineer reduces the magnitude of the stress and the length of the test run is increased. In many of these tests, engineers attempt to reduce the stress load exerted on the last test unit so that it does not fail and produces what engineers refer to as a "suspension" or "run out." The engineer censors the test at some large number of cycles, such as 10^8.

Spiteri et al. (1963) discuss the bending fatigue limit of crankshaft sections based on samples of size 6. This staircase fatigue test represents a problem from automotive industry. This small sample concept is also being imbedded in tests for reliability design in strain testing as seen in the work of Xiong et al. (2002). Electrical engineers resort to Bayesian methods to reduce sample sizes in automotive electronics problems (see Kleyner et al., 1997). Martz and Waller (1982) pioneered the use of Bayesian methods at the Los Alamos National Laboratory in the 1970s as a means to reduce cost and test time in certification of systems in the nuclear industry.

Likewise, redesigned parts can be made in some cases to the exact same specifications as their predecessors except that new parts are made of metal alloys that reduce weight without sacrificing strength. While one might well expect such parts to fail in quite similar ways, we need to compare the distributions of the components, original and redesigned, to ensure that strength has not been compromised.

2. Foundations

Let $X_1, \ldots, X_n$ be n iid random variables having a common density as in (1). Their joint density function is then

$$f(x_1, \ldots, x_n \mid \kappa, \beta) = \frac{1}{[\beta^\kappa \Gamma(\kappa)]^n} \prod_{i=1}^{n} x_i^{\kappa-1} \exp(-x_i/\beta)$$

$$= \frac{1}{[\beta^\kappa \Gamma(\kappa)]^n} \tilde{x}^{n(\kappa-1)} \exp(-n\bar{x}/\beta), \tag{2}$$

where $\bar{x}$ and $\tilde{x}$ denote the arithmetic and geometric means of the random sample. Let $U = \tilde{x}/\bar{x}$ denote the ratio of the geometric mean to the arithmetic mean as a random variable, and $R_n = -ln(U)$. Assume that Y follows a beta distribution with parameters ξ and δ, denoted by $Y \sim \mathcal{B}(\xi,\ \delta)$, and with density function

$$f(y) = \frac{y^{\xi-1}(1-y)^{\delta-1}}{\mathbf{B}(\xi,\ \delta)} \times I_{(0,1)}(y), \tag{3}$$

where $\mathbf{B}(\xi,\ \delta)$ is the complete beta function. We provide a list of results that we will use in following sections. Results $1-5$ are related to sampling distributions of statistics from gamma distributions. For proofs of these results, we refer readers to Glaser (1976b) and Bain and Engelhardt (1975). For Results 6 and 7 related to beta distributions, we refer readers to Rao (1965).

1. $\bar{X}$ and U are jointly sufficient and complete statistics for random samples from the gamma distribution.
2. The distribution of U does not depend on β.
3. The distributions of $\bar{X}$ and U are statistically independent.
4. $2n\kappa R_n$ is approximately distributed as $c\chi^2_\nu$ for appropriate values of c and ν depending on n and κ. For $\kappa > 2$, the distribution of $2n\kappa R_n$ can be approximated by a chi-square distribution with degrees of freedom, (df), $n-1$.
5. Let $T = n\bar{X}$. Then $T \sim \mathcal{G}(n\kappa,\ \beta)$.
6. Let $Y_i = X_i/T$ for $i = 1, 2, \ldots, n-1$. The marginal distribution is $Y_i \sim \mathcal{B}(\kappa, n\kappa)$ for each $i = 1, \ldots, n-1$, while the joint distribution of $Y_1, \ldots, Y_{n-1} \sim \mathcal{D}(\kappa, \ldots, \kappa, n\kappa)$ has a Dirichlet distribution.
7. Let $Y_1 \sim \mathcal{B}(\xi_1,\ \delta_1)$ and $Y_2 \sim \mathcal{B}(\xi_2,\ \delta_2)$, and be independently distributed. If $\xi_1 = \xi_2 + \delta_2$, then $Y_1 Y_2 \sim \mathcal{B}(\xi_2,\ \delta_1 + \delta_2)$.

The maximum likelihood estimators of β and κ, denoted by $\hat{\beta}$ and $\hat{\kappa}$, are solutions to the following equations:

$$R_n = ln(\hat{\kappa}) - \psi(\hat{\kappa}) \ \text{ and } \ \hat{\kappa}\hat{\beta} = \bar{x}, \tag{4}$$

where $\psi(x)$ denotes the digamma or Euler's-psi function. For more results on gamma and beta distributions, we refer the reader to Johnson et al. (1994, 1995).

3. Testing Hypotheses on a Single Shape Parameter

In this section, we first test the shape parameter of a gamma distribution under the assumption that the scale parameter is unknown. $\bar{X}$ alone cannot be used to construct a test statistic for testing the shape parameter as the scale parameter is involved in its distribution.

However, the distribution of U depends only on κ and not on β. Keating et al. (1990) constructed a uniformly most powerful unbiased (UMPU) test for κ based on the ratio of the geometric to arithmetic sample means, U, only by expressing the density function of U in powers of $-ln(U)$. This representation is Glaser's series expansion of the distribution of U (Glaser, 1976b). Keating et al. (1990) noted that Glaser's expression yields a conservative radius of convergence for the series, which is known to converge for all u in the closed interval $[exp(-2\pi/n), 1]$. This condition is problematic whenever the alternative hypothesis is left-sided which occurs frequently when one tests the null hypothesis of exponentiality against a DFR alternative.

Now, we consider the problem of testing the hypothesis

$$H_{01} : \kappa = \kappa_0. \tag{5}$$

Our goal is to develop a test that controls the Type I error rate α even for very small values of n. Glaser (1976b) proved that the distribution of U^n is distributed as a product of $(n-1)$ independent beta distributions, i.e.,

$$U^n \sim \prod_{i=1}^{n-1} Z_i, \tag{6}$$

where $Z_i \sim \mathcal{B}(\kappa, \frac{i}{n}), i = 1, 2, \cdots, n-1$. The density function of the product of independent beta variables is provided by Springer and Thompson (1970) as a Meijer G-function which require evaluation of Mellin integral. It is not easy to determine percentile points from the exact distribution as it is computationally complicated. Hence, we would like to approximate the distribution of $-\sum_{i=1}^{n-1} log(Z_i)$ by a scaled χ^2 distribution (say, by $c\chi^2_\nu$). Using Patnaik's approximation (see Johnson et al., 1995, p. 239) for the distribution of $-log(Z_i)$, we solve for c and ν by equating means and variances of $-\sum_{i=1}^{n-1} log(Z_i)$ with those of $c\chi^2_\nu$. Thus,

$$\nu = \sum_{i=1}^{n-1} c_i^2 \nu_i \Big/ \left(\sum_{i=1}^{n-1} c_i \nu_i\right) \text{ and } c = \left(\sum_{i=1}^{n-1} c_i^2 \nu_i\right)^2 \Big/ \left(\sum_{i=1}^{n-1} c_i \nu_i\right), \tag{7}$$

where $c_i = \frac{1}{2}\frac{\psi'(\kappa)-\psi'(\kappa+i/n)}{\psi(\kappa+i/n)-\psi(\kappa)}$ and $\nu_i = \frac{2[\psi(\kappa+i/n)-\psi(\kappa)]^2}{\psi'(\kappa)-\psi(\kappa+i/n)}$, and ψ and ψ' are digamma and trigamma functions, respectively.

Hence, our test statistic T_1 is $-log(U^n)$ which follows a $c\chi^2_\nu$. In order to evaluate the performance of this test, we consider both right- and left-sided alternatives. As simulation results are similar for both the alternatives, we present here the results for right-sided alternative only. An extensive Monte Carlo simulation study based on 1 million samples indicates that T_1 performs extremely well in controlling simulated Type I error rate for $\kappa = 0.25, 0.50, \ldots, 5$ and $n = 2, 3, 4, 5, 10, 20$. We have also compared the simulated Type I error rates of this test with those of Bhaumik et al. (2009) (BKG) and the likelihood ratio test (LRT). The results of T_1, BKG and LRT are reported in Table 1. Inspection of Table 1 shows that the performance of T_1 is exact for $n = 2$. This is not surprising as the distribution of U^n is exactly beta and we have used this exact distribution to carry out the simulation. The performance of the LRT is unsatisfactory for small values of n, but it improves significantly for $n = 20$. Comparing T_1 and BKG we see that for smaller values of κ, T_1 has a better performance. In addition, we have simulated power curves for all these three tests. As the LRT has a much higher Type I error rate, its power curve is elevated

Table 1
Simulated Type I error rates of T_1, BKG, and LRT for the right tail alternatives

	n = 2			*n = 3*			*n = 4*			*n = 5*			*n = 10*			*n = 20*		
κ	T_1	BKG	LRT	T_1	BKG	LRT	T_1	BKG	LRT	T_1	BKG	LRT	T_1	BKG	LRT	T_1	BKG	LRT
0.250	0.051	0.085	0.178	0.059	0.064	0.153	0.058	0.062	0.127	0.053	0.057	0.102	0.051	0.053	0.068	0.051	0.052	0.061
0.500	0.051	0.069	0.176	0.059	0.060	0.152	0.055	0.059	0.128	0.051	0.056	0.104	0.051	0.052	0.067	0.050	0.051	0.061
0.750	0.050	0.058	0.171	0.058	0.056	0.150	0.054	0.054	0.129	0.052	0.054	0.106	0.051	0.052	0.067	0.050	0.051	0.061
1.000	0.050	0.057	0.166	0.050	0.053	0.148	0.054	0.053	0.129	0.051	0.053	0.107	0.050	0.051	0.067	0.050	0.051	0.061
1.250	0.049	0.056	0.160	0.051	0.053	0.144	0.049	0.053	0.128	0.049	0.053	0.108	0.050	0.051	0.067	0.050	0.051	0.061
1.500	0.050	0.052	0.154	0.051	0.052	0.141	0.051	0.051	0.126	0.052	0.052	0.108	0.050	0.051	0.067	0.050	0.050	0.061
1.750	0.050	0.052	0.149	0.050	0.052	0.138	0.051	0.051	0.125	0.049	0.051	0.108	0.050	0.050	0.068	0.050	0.050	0.061
2.000	0.051	0.051	0.144	0.051	0.052	0.135	0.049	0.050	0.123	0.050	0.051	0.108	0.049	0.050	0.067	0.050	0.050	0.060
2.250	0.051	0.051	0.139	0.048	0.051	0.132	0.052	0.051	0.122	0.051	0.051	0.108	0.050	0.050	0.067	0.050	0.050	0.060
2.500	0.049	0.050	0.134	0.049	0.054	0.129	0.052	0.049	0.121	0.050	0.050	0.108	0.050	0.050	0.067	0.050	0.050	0.060
2.750	0.049	0.052	0.130	0.048	0.050	0.126	0.050	0.050	0.119	0.052	0.050	0.108	0.051	0.049	0.068	0.050	0.050	0.060
3.000	0.051	0.051	0.125	0.049	0.050	0.123	0.049	0.051	0.118	0.052	0.050	0.107	0.050	0.050	0.067	0.050	0.050	0.060
3.250	0.050	0.050	0.121	0.049	0.049	0.120	0.050	0.051	0.116	0.051	0.050	0.112	0.050	0.050	0.067	0.050	0.050	0.060
3.500	0.050	0.049	0.117	0.048	0.050	0.117	0.050	0.050	0.115	0.051	0.050	0.111	0.050	0.051	0.066	0.050	0.050	0.060
3.750	0.050	0.051	0.113	0.051	0.051	0.114	0.052	0.051	0.113	0.049	0.050	0.111	0.050	0.050	0.067	0.050	0.049	0.060
4.000	0.050	0.050	0.109	0.049	0.049	0.111	0.051	0.049	0.112	0.051	0.051	0.110	0.049	0.050	0.067	0.050	0.050	0.060
4.250	0.049	0.051	0.105	0.052	0.049	0.109	0.048	0.050	0.110	0.049	0.049	0.110	0.050	0.050	0.067	0.050	0.050	0.060
4.500	0.049	0.049	0.101	0.049	0.051	0.106	0.050	0.050	0.109	0.048	0.050	0.109	0.050	0.050	0.067	0.050	0.050	0.059
4.750	0.049	0.052	0.098	0.052	0.049	0.104	0.049	0.050	0.108	0.051	0.051	0.109	0.051	0.049	0.067	0.050	0.050	0.059
5.000	0.049	0.049	0.096	0.049	0.050	0.102	0.049	0.050	0.107	0.049	0.050	0.108	0.050	0.050	0.066	0.050	0.050	0.059

significantly compared to T_1 and BKG. There is no trivial mathematical method to adjust the Type I error rate and hence calibration of the power curve of LRT is nearly impossible.

Example 3.1. *We now apply our results to the amosite asbestos data collected from NYSDOH 25 (see Bhaumik, Kim, and Gibbons, edited by Nelson et al., 2009, pp. 93–108). The NYSDOS 25 data were measured by transmission electron microscopy (TEM) for testing and laboratory assessment. The unit of TEM measurement of asbestos data is structures per square millimeter. We first check the distributional assumption based on the gamma quantile-quantile plots and Anderson-Darling goodness-of-fit tests: gamma distributions fit well to 13 out of 14 sites (sample sizes of these sites vary from 23 to 29) (Kim, 2011). These results indicate that gamma distributions fit well to amosite asbestos data collected from NYSDOS 25. Thus, this gives the basis for analyzing NYSDOS 25 data using gamma distributions for small samples. Initially, sites with smaller samples* (≤ 5) *were discarded from the study as appropriate statistical methodologies were not available. The mean permissible exposure limit (PEL) of amosite fiber suggested by the Environmental Protection Agency (EPA) is 70 structures per square millimeter. The United States Department of Labor routinely applies the NIOSH strategy and recommends the coefficient of variation (cv) for fibers to be* 0.13. *Following these guidelines and using formulae for the mean and variance of the gamma distribution, we compute* $\kappa = 59.31$ *and* $\beta = 1.18$ *for the permissible exposure level. Now, consider the site 5099 which has only 4 measurements. Scale adjusted values of these measurements are 96.84, 92.97, 73.84, and 81.71 and the estimated* $\kappa = 86.34$. *Based on the test* T_1, *we do not have enough evidence to believe that the shape parameter of this dataset is significantly greater than the permissible value of* 59.31 *(*p*-value* ≈ 0.56*).*

4. Multiple Gamma Populations

In introducing comparison of multiple gamma populations, we consider the problem of competing risks. Consider a homogeneous population of n individuals with lives that are at risk to p diseases or competing causes of death, such as cardiovascular disease, cancer, diabetes, etc. (see Crowder, 2001; Pintille, 2006).

Let X_ℓ, $\ell = 1, \ldots, n$, be the observed lifetime of an individual and let Δ_ℓ be the indicator which specifies the cause of death for the ℓth individual. The random variable Δ_ℓ has support on the set $\{1, \ldots, p\}$. Specifically, if the person expires due to the ith disease, we have

$$\Delta_\ell = i \quad \text{and} \quad \Pr(\Delta_\ell = i) = \pi_i.$$

The natural restriction is that $\sum_{i=1}^{p} \pi_i = 1$. It follows that the conditional distribution of an individual's lifetime due to the ith cause is denoted by $X \mid \Delta = i$ and has the same pdf as the ith family as

$$f(x, \Delta = i) = f(x \mid \Delta_\ell = i) \Pr(\Delta_\ell = i) = f_i(x) \pi_i.$$

Using the previous expression, we can write a general expression for the likelihood of N independent deaths due to p competing causes as

$$L = \prod_{i=1}^{p} \left[\prod_{j=1}^{n_i} f_i\left(x_{ij}\right) \pi_i \right],$$

where x_{ij} is the jth lifetime due to the ith cause of death, n_i is the number of deaths due to the ith cause, and $N = \sum_{i=1}^{p} n_i$. Within the context of competing risks, we introduce multiple independent Gamma distributions each representing a particular cause of death. We now address the estimation and testing issues related to this problem.

4.1. *Testing Shape Parameters of Multiple Gamma PopuLations*

In this section, we assume that we have p independent gamma populations, and would like to compare their shape parameters. Assume that we have n_i independent observations from the ith population, i.e., $x_{ij} \sim \mathcal{G}(\kappa_i,\ \beta_i)\ j = 1, 2, \ldots, n_i$ and $i = 1, 2, \cdots p$. Our hypothesis is $H_{02} : \kappa_1 = \kappa_2 = \cdots = \kappa_p$ against the alternative hypothesis H_{a2} that H_{02} is not true. There are many problems in a wide variety of fields where this issue is of interest. In engineering, when components of a product are made in different plants for the purpose of using them in the same system, it is essential to check whether they are equivalent. In environmental monitoring problems, we often have a small number of background samples from a series of hydraulically upgradient (of a potential industrial hazard such as a landfill) ground-water monitoring wells, which we want to show are equivalent prior to pooling. In studies of this type, generally the shape parameter governs the rule as data are often rescaled and it is assumed that $\beta = 1$. In the first part of this section, we develop a test for H_{02} under the assumption that $x_{ij} \sim \mathcal{G}(\kappa_i,\ \beta)$. Let $Z = \sum_{i=1}^{p} \sum_{j=1}^{n_i} x_{ij}$ and $z_{ij} = \frac{x_{ij}}{Z}$. The joint distribution of z_{ij}'s is Dirichlet with the following density function:

$$f(z_{11}, \cdots, z_{pn_p}) = \frac{\Gamma(n_1\kappa_1 + \cdots + n_p\kappa_p)}{(\Gamma(\kappa_1))^{n_1} \cdots (\Gamma(\kappa_p))^{n_p}} \times z_{11}^{(\kappa_1-1)} \cdots z_{1n_1}^{(\kappa_1-1)} \cdots z_{p1}^{(\kappa_p-1)} \cdots z_{pn_p}^{(\kappa_p-1)}. \tag{8}$$

Marshall and Olkin (1979) called this distribution the Dirichlet Type III distribution. We now construct an unbiased test (in the sense of Schur convexity) for H_{02} using z_{ij}'s and their joint distribution presented above. Let us denote this test by T_2 and state the result in the form of a theorem.

Theorem 4.1. *Let $x_{i1}, x_{i2}, \cdots, x_{in_i}$ be a random sample from the* i*th gamma population with shape parameter κ_i and scale parameter β, where $i = 1, \cdots, p$. For testing the hypothesis H_{02} against the alternative H_{a2}, there exists an unbiased test in the sense of Schur convexity given by: reject H_{02} if $\Phi(z_{11}, \cdots, z_{pn_p}) < D_\alpha$, where $\Phi(z_{11}, \cdots, z_{pn_p}) = \prod_{i=1}^{p} \prod_{j=1}^{n_i} z_{ij}$ and D_α is the $100\alpha\%$ point of the distribution of the test-statistic $\Phi(z_{11}, \cdots, z_{pn_p})$ when all the κ's are equal.*

Proof. $\Phi(z_{11}, \cdots, z_{pn_p})$ is a Schur concave function of random variables $z_{11}, \cdots, z_{pn_p}$. Note that $\sum_{i=1}^{p} \sum_{j=1}^{n_i} z_{ij} = 1$ and the indicator function $I_{\{\Phi(z_{11}, \cdots, z_{pn_p}) < D_\alpha\}}$ is a Schur convex function. Dirichlet Type III distribution parameterized by the vector $\kappa_1, \cdots, \kappa_p$ has the property that expectation of a Schur convex function leads to a Schur convex function of $\kappa_1, \cdots, \kappa_p$ (Marshall and Olkin, 1979, chap. 11). Hence, $\Psi(\kappa_1, \cdots, \kappa_p) = E(I_{\{\Phi(z_{11}, \cdots, z_{pn_p}) < D_\alpha\}})$ is a Schur convex function of $\kappa_1, \cdots, \kappa_p$. Thus under H_{02}, $\Psi(\kappa_1, \cdots, \kappa_p)$ takes its minimum value at $(\kappa, \cdots, \kappa)$ and is majorized by $(\kappa_1, \cdots, \kappa_p)$. But, the value of $(\kappa_1, \cdots, \kappa_p)$ is $(\kappa, \cdots, \kappa)$ under H_{02}. Therefore, the proposed test is unbiased in the sense of Schur convexity.

Table 2
Simulated Type I error rates of LRT and T_2

κ,κ	(2.00, 2.00)	(2.50, 2.50)	(3.00, 3.00)	(3.50, 3.50)	(4.00, 4.00)
			$n_1 = n_2 = 3$		
LRT	0.127	0.128	0.129	0.128	0.127
T_2	0.050	0.049	0.050	0.049	0.052
			$n_1 = n_2 = 5$		
LRT	0.093	0.093	0.091	0.090	0.089
T_2	0.051	0.050	0.050	0.051	0.050

In order to implement the result stated in Theorem 1, we need the common value of κ, say κ_0, under H_{02}. We call κ_0 the target value. The critical value D_α of Theorem 1 depends on the target value κ_0. The power of the test at $(\kappa_1, \cdots, \kappa_p)$ is always greater than α whenever $\sum_{i=1}^{p} n_i\kappa_i = \kappa_0 \sum_{i=1}^{p} n_i$.

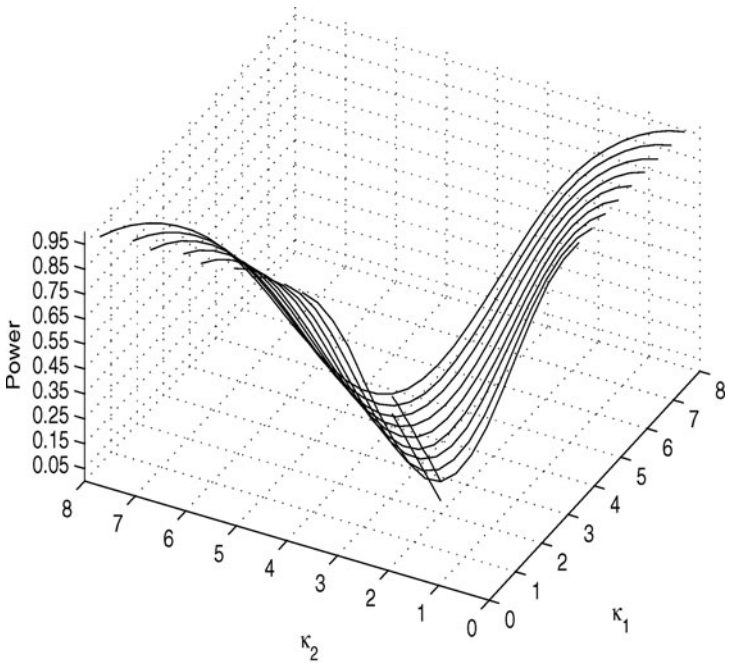
We have performed an extensive simulation study in order to evaluate T_2 and compared its Type I error rates with those of likelihood ratio test (LRT) (see Appendix) for two gamma populations with $n_1 = n_2 = 3$, $n_1 = n_2 = 5$ and several combinations of $\kappa_1 = \kappa_2$. The results are presented in Table 2. From this table, we conclude that T_2 controls the Type I error rates at its nominal value (5%) for all combinations of κ_1 and κ_2. However, the LRT has inflated Type I error rates for both $n_1 = n_2 = 3$ and $n_1 = n_2 = 5$.

We have also compared power curves of T_2 and LRT by means of Monte Carlo simulation for some combinations of κ_1 and κ_2, *i.e.*, $\kappa_1 = 2, 2.25, \ldots, 4$ and $\kappa_2 = 2, 2.25, \ldots, 4$ with the constraint that $\sum_{i=1}^{2} \kappa_i = 2\kappa_0$. In Figure 1, we see that T_2 controls Type I error rates for all the combinations of κ_1 and κ_2 when $n_1 = n_2 = 3$. The power of T_2 at (κ_1, κ_2) is always greater than the nominal rate (5%) as we move along the direction of the constraint. However, the LRT has higher power than T_2 along the constrained direction at the expense of inflated Type I error rates. This figure also shows the reference line of nominal rates (5%) for visualization. Similar results are shown in Figure 2 for $n_1 = n_2 = 5$. Comparing Figures 1 and 2, we see that the power curve of T_2 becomes steeper for larger values of n. We see an improvement in Type I error rates of the LRT as n increases, but it still remains inflated.

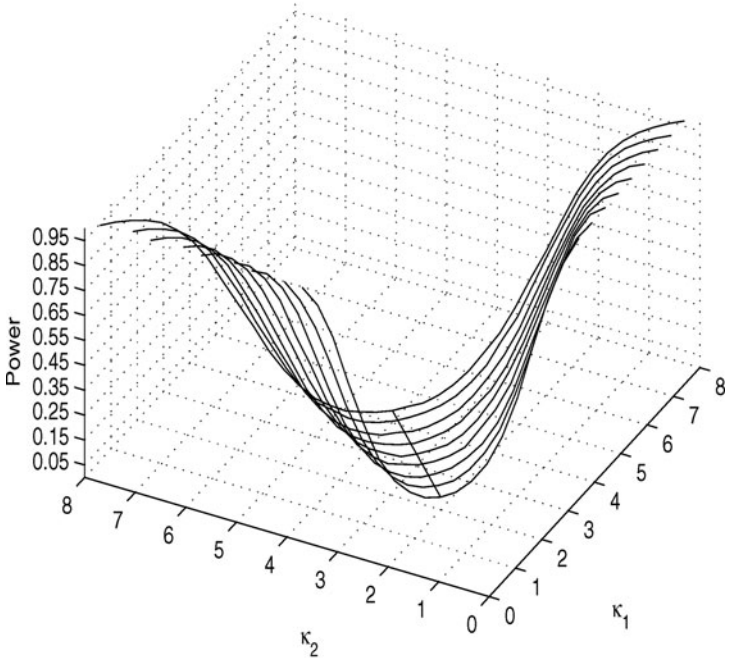
In addition, we present simulated power of T_2 under one-sided alternative in Table 3. The concept of one-sided alternative is under the theory of majorization, i.e., the sum of two components are fixed and the first component is larger than the second component. In Table 3, we show that for more dispersed components, we achieve better power.

4.2. *Bayesian Analysis of Multiple Gamma Populations*

In this section, we discuss the Bayesian approach to hypothesis testing. To quantify the evidence in favor of a hypothesis, Jeffreys (1935) developed the Bayes' factor (BF). The BF is the ratio of the posterior odds to the prior odds of the alternative hypothesis. Hence, when the prior odds ratio is set to be 1, then the BF is the posterior odds. The priors should be chosen properly as the BF is sensitive to that choice. Let the probability densities of X under H_0 and H_1 be denoted by $p(X|H_0)$ and $p(X|H_1)$, respectively. The BF denoted by BF_{10} is then $p(X|H_1)/p(X|H_0)$. We are assuming that $p(H_0) = p(H_1) = 0.5$. Denote the prior density of parameter θ_i under H_i by $\pi(\theta_i|H_i)$. $p(X|\theta_i, H_i)$ is the likelihood function

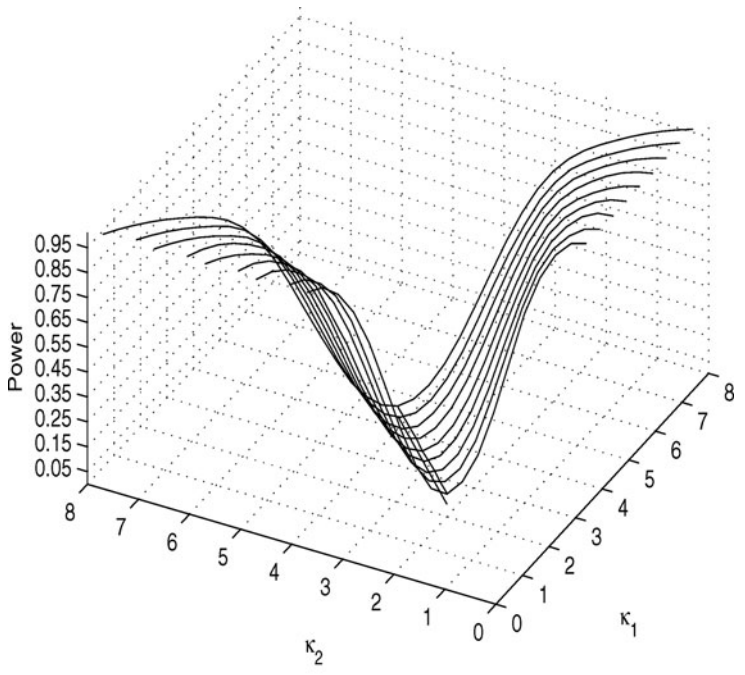


(a) LRT

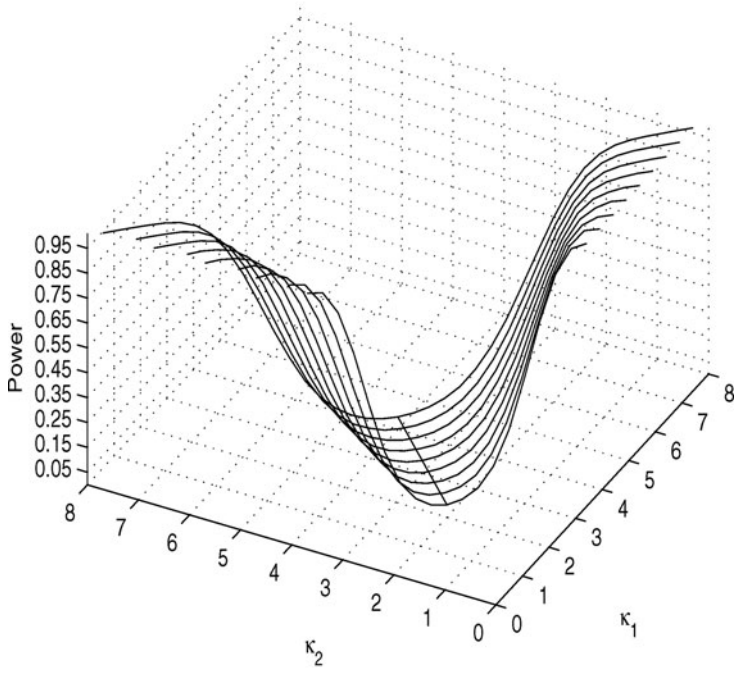


(b) T_2

Figure 1. Power curves for testing equality of shape parameters $H_{02} : \kappa_1 = \kappa_2$ for $n_1 = n_2 = 3$.



(a) LRT



(b) T_2

Figure 2. Power curves for testing equality of shape parameters $H_{02} : \kappa_1 = \kappa_2$ for $n_1 = n_2 = 5$.

Table 3
Simulated power of T_2

κ_1	κ_2	Power(T_2)	κ_1	κ_2	Power(T_2)
7.20	0.80	0.999	3.80	0.20	0.999
6.20	1.80	0.841	3.00	1.00	0.502
5.20	2.80	0.264	2.60	1.40	0.172
4.00	4.00	0.050	2.00	2.00	0.049

of the θ_i under H_i, where $i = 0, 1$ which is obtained by integrating the likelihood function over the parameter space as

$$p(X|H_i) = \int p(X|\theta_i, H_i)\pi(\theta_i|H_i)d\theta_i. \tag{9}$$

Here, $p(X|H_i)$ is the marginal probability under H_i. B_{10} is the ratio of these marginal probabilities under two different hypotheses and it is closely related to the likelihood ratio statistic where the parameter θ_i is eliminated by maximization rather than integration. The BF provides a summary of the evidence in favor of one hypothesis as opposed to the other. Kass and Raftery (1995) provide the following table to interpret the value of B_{10}.

Numerical computation of B_{10} for high-dimensional parametric space can be performed using well established methods like Laplace's approximation and Markov Chain Monte Carlo methods (MCMC), particularly the Metropolis-Hastings and Gibbs sampling technique. In the context of testing multiple gamma populations, under H_0 the gamma distributions are identical and independently distributed and under H_1 the gamma distributions are independent but with different shape parameters.

We use the following conjugate prior for the shape parameter (κ) of the gamma distribution provided by Miller (1980) and Fink (1995) when the scale parameter is assumed to be known. Let $S = n\bar{X}$ and $P = \tilde{X}^n$. The conjugate prior with hyperparameters $a, b, c > 0$ is defined as

$$\begin{aligned} \pi(\kappa, \theta|a, b, c) &= \frac{a^{\kappa-1}}{K\Gamma(\kappa)^b\theta^{c\kappa}} \quad \kappa > 0 \\ &= 0 \ \ otherwise, \\ where \ \ K &= \int_0^\infty \frac{a^{\kappa-1}}{\Gamma(\kappa)^b\theta^{c\kappa}}d\kappa. \end{aligned} \tag{10}$$

This prior density implies past data or a hypothetical experiment with a sample size b and a product of observations a. The posterior distribution of κ is specified by updated hyperparameters

$$a' = aP, \quad b' = b + n, \quad c' = c + n. \tag{11}$$

In order to see the performance of BF for small sample sizes ($n = 3, 5, 10, 20$), we considered 2, 3, and 4 independent gamma populations. We implemented a cautious adaptive Romberg method of integration to compute the marginal densities for the BFs. This integration technique is easy to implement and is more efficient in terms of reducing numerical

Table 4
Bayes' factor table

$log_{10}(B_{10})$	B_{10}	Evidence against H_0
0 to 0.5	1 to 3.2	Not worth more that a bare mention
0.5 to 1	3.2 to 10	Substantial
1 to 2	10 to 100	Strong
> 2	> 100	Evidence

errors than approximate methods such as Gaussian quadrature or MCMC techniques. We have assumed $\theta = 1$, and two distinct prior densities with hyperparameters $(a = P,\ b = n)$ and $(a = P,\ b = n/2)$. Under H_0, we considered three distinct values of $\kappa = 2, 3, 4$, and under H_1 we have considered various combinations of κ. Based on 10, 000 simulations, we have computed proportions of BF meeting the criteria defined in Table 4 for each combination of sample sizes, parameters, hyperparameters, and number of populations. The results of our simulation study are presented in Tables 5–10. Inspecting Tables 5–10, we see that BF is very sensitive to the choice of sample size as well as the prior distributions. For example, in Table 8, when $\kappa_1 = 1$ and $\kappa_2 = 3$, the proportion of $BFs > 3.2$ (substantial evidence for H_1 as opposed to H_0) for $n = 3, 5, 10, 20$ are 0.627, 0.788, 0.908, 0.996, respectively. This means that when sample size is increased from 3 to 20, the corresponding proportions of $BFs > 3.2$ is also increased from 0.627 to 0.996. For this combination of parameters and sample sizes, when the hyperparameters of the priors are changed from $a = P,\ b = n$ to $a = P,\ b = n/2$, we see from Table 9 that corresponding proportions of $BFs > 3.2$ are 0.251, 0.267, 0.398, 0.642. In other words, simply due to change in the priors, the proportions of $BFs > 3.2$ change from 0.627 to 0.251. In addition, we observe that proportions of $BFs > 3.2$ are not that sensitive to changes in the hyperparameter a, when b is fixed. Similar findings are seen in Tables 7–10.

The proportions of BFs under H_0 when all the κ's are equal should be ideally 0 (analogous to Type I error rate). Inspecting Tables 7–10, we find that for equal κ's, the proportion of BFs vary drastically depending on the number of populations, sample size, and prior distribution. For example, in the case of two populations, when H_0 is $(2, 2)$, $n = 3$ and prior is $(a = P,\ b = n)$, the proportion of $BFs > 3.2$ is 0.074 (see Table 5), but that proportion becomes 0.014 (see Table 6) for the prior $(a = P,\ b = n/2)$, being five times less. Inspecting Tables 7 and 9, the corresponding proportion of $BFs > 3.2$ for three and four populations with $n = 3$ and the prior $(a = P,\ b = n)$ are 0.167 and 0.210. In general, when number of populations increases, the proportions of $BFs > 3.2$ also increase irrespective of the prior distribution, but the rate of increment decreases as the sample size increases.

For unequal κ's, the proportion of $BFs > 3.2$ are expected to be close to 1 depending on the difference of the components. For example, when the scenarios are $(\kappa_1 = 1,\ \kappa_2 = 7)$ and $(\kappa_1 = 1,\ \kappa_2 = 3,\ \kappa_3 = 5,\ \kappa_4 = 7)$, the proportions of $BFs > 3.2$ are greater than 0.95 irrespective of the sample sizes and prior distributions. However, the evidence becomes weaker when the difference in the components and sample sizes are smaller.

Prior distributions play a vital role in the computation of the Bayes' factor. Generally, a combination of relevant data and information from the literature are used while choosing the prior distribution. For example, Johnson et al. (2005) used binomial-beta model for the outcomes of eleven launches of new vehicles conducted by companies with limited

Table 5
Proportion of Bayes' factor for two populations with prior parameters $a = P$, $b = n$

κ, κ		n				κ, κ		n				κ, κ		n			
2, 2	*BF*	3	5	10	20	3, 3	*BF*	3	5	10	20	4, 4	*BF*	3	5	10	20
	> 1	0.397	0.370	0.390	0.395		> 1	0.392	0.387	0.406	0.421		> 1	0.426	0.410	0.411	0.381
	> 3.2	0.074	0.075	0.081	0.090		> 3.2	0.076	0.079	0.081	0.082		> 3.2	0.083	0.082	0.094	0.073
	> 10	0.023	0.018	0.027	0.020		> 10	0.019	0.020	0.028	0.022		> 10	0.026	0.026	0.016	0.022
	> 100	0.003	0.000	0.001	0.002		> 100	0.002	0.002	0.001	0.000		> 100	0.002	0.002	0.001	0.000
1, 3	*BF*	3	5	10	20	1, 5	*BF*	3	5	10	20	2, 4	*BF*	3	5	10	20
	> 1	0.888	0.911	0.955	0.999		> 1	0.994	0.999	1.000	1.000		> 1	0.784	0.835	0.910	0.961
	> 3.2	0.627	0.788	0.908	0.996		> 3.2	0.942	0.995	1.000	1.000		> 3.2	0.426	0.647	0.831	0.932
	> 10	0.402	0.648	0.851	0.986		> 10	0.849	0.983	1.000	1.000		> 10	0.227	0.466	0.743	0.897
	> 100	0.129	0.330	0.671	0.933		> 100	0.574	0.879	0.999	1.000		> 100	0.055	0.140	0.507	0.779
1, 7	*BF*	3	5	10	20	2, 6	*BF*	3	5	10	20	3, 5	*BF*	3	5	10	20
	> 1	0.9998	0.9999	1.0000	1.0000		> 1	0.971	0.987	0.999	1.000		> 1	0.704	0.767	0.789	0.995
	> 3.2	0.9952	0.9999	1.0000	1.0000		> 3.2	0.833	0.950	0.996	1.000		> 3.2	0.337	0.532	0.591	0.950
	> 10	0.9809	0.9960	1.0000	1.0000		> 10	0.654	0.887	0.983	1.000		> 10	0.162	0.326	0.451	0.868
	> 100	0.8930	0.9937	1.0000	1.0000		> 100	0.320	0.641	0.957	1.000		> 100	0.036	0.072	0.201	0.596

Table 6
Proportion of Bayes' factor for two populations with prior parameters $a = P, b = n/2$

κ, κ		n				κ, κ		n				κ, κ		n			
2, 2	*BF*	3	5	10	20	3, 3	*BF*	3	5	10	20	4, 4	*BF*	3	5	10	20
BF	> 1	0.073	0.025	0.020	0.009		> 1	0.345	0.236	0.121	0.035		> 1	0.168	0.096	0.017	0.000
	> 3.2	0.014	0.001	0.001	0.000		> 3.2	0.145	0.090	0.045	0.009		> 3.2	0.069	0.039	0.004	0.000
	> 10	0.002	0.000	0.000	0.000		> 10	0.068	0.044	0.022	0.004		> 10	0.029	0.020	0.003	0.000
	> 100	0.000	0.000	0.000	0.000		> 100	0.016	0.018	0.006	0.001		> 100	0.009	0.007	0.001	0.000
1, 3	*BF*	3	5	10	20	1, 5	*BF*	3	5	10	20	2, 4	*BF*	3	5	10	20
	> 1	0.526	0.536	0.645	0.775		> 1	0.880	0.881	0.891	0.917		> 1	0.517	0.551	0.701	0.780
	> 3.2	0.251	0.267	0.398	0.642		> 3.2	0.738	0.781	0.797	0.883		> 3.2	0.301	0.325	0.416	0.480
	> 10	0.124	0.127	0.222	0.480		> 10	0.602	0.677	0.736	0.835		> 10	0.204	0.212	0.308	0.350
	> 100	0.024	0.021	0.055	0.227		> 100	0.385	0.489	0.587	0.743		> 100	0.106	0.110	0.262	0.301
1, 7	*BF*	3	5	10	20	2, 6	*BF*	3	5	10	20	3, 5	*BF*	3	5	10	20
	> 1	0.996	0.997	1.000	1.000		> 1	0.955	0.955	0.988	0.998		> 1	0.865	0.915	0.961	0.988
	> 3.2	0.975	0.993	1.000	1.000		> 3.2	0.877	0.914	0.977	0.995		> 3.2	0.635	0.761	0.897	0.978
	> 10	0.954	0.987	1.000	1.000		> 10	0.785	0.866	0.965	0.995		> 10	0.462	0.628	0.837	0.961
	> 100	0.887	0.959	0.990	1.000		> 100	0.594	0.771	0.936	0.989		> 100	0.245	0.398	0.688	0.917

Table 7
Proportion of Bayes' factor for three populations with prior parameters $a = P, b = n$

κ, κ, κ			n			κ, κ, κ			n		
2, 2, 2	*BF*	3	5	10	20	3, 3, 3	*BF*	3	5	10	20
	> 1	0.528	0.516	0.523	0.525		> 1	0.492	0.472	0.501	0.480
	> 3.2	0.167	0.161	0.151	0.154		> 3.2	0.135	0.155	0.158	0.153
	> 10	0.051	0.055	0.053	0.046		> 10	0.038	0.052	0.046	0.051
	> 100	0.006	0.005	0.007	0.006		> 100	0.003	0.006	0.004	0.005
1, 2, 3	*BF*	3	5	10	20	1, 3, 5	*BF*	3	5	10	20
	> 1	0.882	0.958	0.998	1.000		> 1	0.989	0.999	1.000	1.000
	> 3.2	0.666	0.843	0.979	1.000		> 3.2	0.944	0.990	1.000	1.000
	> 10	0.459	0.674	0.931	1.000		> 10	0.859	0.978	1.000	1.000
	> 100	0.162	0.348	0.765	0.998		> 100	0.609	0.905	0.999	1.000

Table 8
Proportion of Bayes' factor for three populations with prior parameters $a = P$, $b = n/2$

κ, κ, κ		n				κ, κ, κ		n			
2, 2, 2	*BF*	3	5	10	20	3, 3, 3	*BF*	3	5	10	20
	> 1	0.046	0.007	0.013	0.001		> 1	0.357	0.244	0.086	0.019
	> 3.2	0.011	0.002	0.000	0.000		> 3.2	0.210	0.140	0.050	0.008
	> 10	0.006	0.001	0.000	0.000		> 10	0.116	0.078	0.029	0.004
	> 100	0.002	0.000	0.000	0.000		> 100	0.042	0.032	0.011	0.001
1, 2, 3	*BF*	3	5	10	20	1, 3, 5	*BF*	3	5	10	20
	> 1	0.370	0.391	0.508	0.660		> 1	0.815	0.819	0.820	0.820
	> 3.2	0.176	0.202	0.336	0.523		> 3.2	0.686	0.713	0.717	0.748
	> 10	0.089	0.114	0.227	0.392		> 10	0.577	0.611	0.650	0.708
	> 100	0.021	0.028	0.065	0.202		> 100	0.360	0.455	0.505	0.611

Table 9
Proportion of Bayes' factor for four populations with prior parameters $a = P, b = n$

$\kappa, \kappa, \kappa, \kappa$			n			$\kappa, \kappa, \kappa, \kappa$			n		
2, 2, 2, 2	*BF*	3	5	10	20	3, 3, 3, 3	*BF*	3	5	10	20
	> 1	0.550	0.550	0.540	0.550		> 1	0.553	0.549	0.568	0.570
	> 3.2	0.210	0.230	0.209	0.230		> 3.2	0.197	0.222	0.235	0.234
	> 10	0.076	0.072	0.085	0.090		> 10	0.074	0.086	0.103	0.105
	> 100	0.012	0.009	0.009	0.013		> 100	0.011	0.009	0.012	0.013
1, 2, 3, 4	*BF*	3	5	10	20	1, 3, 5, 7	*BF*	3	5	10	20
	> 1	0.973	0.996	1.000	1.000		> 1	1.000	1.000	1.000	1.000
	> 3.2	0.893	0.978	1.000	1.000		> 3.2	0.999	1.000	1.000	1.000
	> 10	0.774	0.930	0.999	1.000		> 10	0.993	1.000	1.000	1.000
	> 100	0.480	0.770	0.990	1.000		> 100	0.939	0.998	1.000	1.000

Table 10
Proportion of Bayes' factor for four populations with prior parameters $a = P, b = n/2$

$\kappa, \kappa, \kappa, \kappa$		n				$\kappa, \kappa, \kappa, \kappa$		n			
2, 2, 2, 2	*BF*	3	5	10	20	3, 3, 3, 3	*BF*	3	5	10	20
	> 1	0.023	0.050	0.040	0.000		> 1	0.330	0.195	0.070	0.000
	> 3.2	0.010	0.002	0.000	0.000		> 3.2	0.226	0.118	0.045	0.000
	> 10	0.006	0.000	0.000	0.000		> 10	0.150	0.073	0.030	0.000
	> 100	0.001	0.000	0.000	0.000		> 100	0.055	0.026	0.009	0.000
1, 2, 3, 4	*BF*	3	5	10	20	1, 3, 5, 7	*BF*	3	5	10	20
	> 1	0.505	0.437	0.332	0.194		> 1	0.993	0.998	1.000	1.000
	> 3.2	0.363	0.319	0.224	0.142		> 3.2	0.986	0.992	1.000	1.000
	> 10	0.249	0.224	0.144	0.117		> 10	0.975	0.987	0.999	1.000
	> 100	0.119	0.103	0.081	0.061		> 100	0.934	0.973	0.998	1.000

Table 11
Scale adjusted samples from three sites

Site	5099	5209	879Q
1	96.84	36.78	16.91
2	92.97	51.74	29.39
3	73.84	48.46	21.21
4	81.71		
$\hat{\kappa}$	86.34	45.66	22.17

launch-design experience and Miller (1980) used gamma-gamma model in the context of survival analysis. Contrary to the Bayesian point estimates like posterior mode and mean, the BF is sensitive to the choice of prior. In order to avoid the problem of prior choice, one can use the Schwarz criterion, but it requires a large sample in order to draw an appropriate conclusion. In our simulation study, we find that there are some cases where results do not stabilize even for $n = 20$.

To illustrate the above result, we now provide two examples from environmental science and engineering.

Example 4.1. *Comparing asbestos fiber counts of the same type obtained from different samples is an important problem. In practice, samples with fewer observations (<5) are discarded as the current methodologies fail to handle this problem efficiently. We consider fiber counts obtained from three sites:* 5099*,* 5209*, and* 879*Q. The scale-adjusted samples, along with the estimate of* κ *from each site, are presented in Table 11.*

In this example, we illustrate first our methodology for two populations: 5099 and 5209. We set the null hypothesis as $H_{02} : \kappa_1 = \kappa_2$. The null value is set to the permissible exposure level value of $\kappa = 59.31$. We then find $T_2 = 8.248342e-07$ and since the critical value is $1.093542e-06$, we reject the null hypothesis. We also compute the Bayes' factor B_{10} for these two sites. The values of BFs are $B_{10} = 0.5061621$ and $B_{10} = 0.6244722$ for hyperparameters $a = P$, $b = n$ and $a = P$, $b = n/2$, respectively. Hence, we do not find any evidence to reject the null hypothesis using BFs.

Similarly, we consider the hypothesis for all the three populations, i.e., $H_{02} : \kappa_1 = \kappa_2 = \kappa_3$. In this scenario, we consider three null values: (i) permissible exposure level value of $\kappa = 59.31$, (ii) arithmetic mean, $\kappa = 51.39$, and (iii) geometric mean, $\kappa = 44.37$. The critical values for these three null hypotheses are $8.662195e-11$, $8.48668e-11$ and $8.258074e-11$, respectively, whereas the value of $T_2 = 2.056885e-11$. Hence, we reject the null hypothesis in each scenario. Again, the values of BFs are $B_{10} = 0.2598837$ and $B_{10} = 0.3124744$ for hyperparameters $a = P$, $b = n$, and $a = P$, $b = n/2$, respectively. Here again, we do not find any evidence to reject the null hypothesis using BFs.

Example 4.2. *Consider the failure data in Feiveson and Kulkarni (2000) of stress-ruptures of Kevlar-wrapped pressure vessels. During an experimental period of 5 years, 86 of 97 vessels experienced stress rupture and their failure times were recorded. All the remaining vessels had been pressurized to the lowest stress fraction. An important characteristic of the vessels is that their Kevlar strands were manufactured in lots known as "Spools." These data can also be found in Glaser (1983). Consider the data for two samples obtained from*

Spool 1 at two different stress fractions. For a stress fraction of 0.791, the failure times in hours are: 453.4, 664.5, 930.4, *and* 1755.5. *The geometric and arithmetic means are* 837.66 *and* 951.01, *respectively, with their ratio being 0.8808. Our estimates of* κ *and* θ *based on this sample are* $\hat{\kappa} = 4.099$ *and* $\hat{\theta} = 232.05$. *For a stress fraction of* 0.853, *the failure times in hours are:* 444.4, 755.2, 952.2, *and* 1108.2. *The geometric and arithmetic means are* 771.43 *and* 815.00, *respectively, with their ratio being 0.9465. Our estimates of* κ *and* θ *based on this sample are* $\hat{\kappa} = 9.26$ *and* $\hat{\theta} = 87.95$. *To set the null value, we aggregate the two samples into one sample of size 8. The MLE of* κ *for this combined sample is* $\kappa = 5.485$. *Following the same procedure described in the above example, we have the value of* $T_2 = 1.506527e-08$ *and the critical value as* $7.467173e-06$. *The values of BFs are* $B_{10} = 72.05573$ *and* $B_{10} = 671.6531$ *for hyperparameters* $a = P$, $b = n$ *and* $a = P$, $b = n/2$, *respectively. Hence, we reject the null hypothesis by using* T_2 *as well as by the values of BFs.*

5. Discussion

We have developed two tests for the shape parameter(s) of gamma distributions. In the one-sample case, we have developed test (T_1) which allows us to determine if the data are exponentially distributed ($\kappa = 1$) versus the gamma alternatives of increasing failure rate ($\kappa > 1$) or decreasing failure rate ($\kappa < 1$). The most important feature of this test is that it maintains the nominal Type I error rate even for extremely small sample sizes. Next, we considered multiple gamma populations with the interest lying in comparing the shape parameters of the gamma distributions between two or more populations. Tests of equality of gamma distributions are important in numerous applied areas including but not limited to risk assessment, environmental monitoring, reliability, and inter-laboratory calibration. The proposed new tests provide excellent results for small samples. In addition, we discuss the BF in this context and compare its performance with the frequentist approach. For large samples, frequentist tests have a tendency to reject the null hypothesis for large samples whereas the BF does not. For example, when $n = 20$ and the prior is $a = P$, $b = n/2$, BFs provide a strong evidence for the null hypotheses for all two, three, and four populations whereas in case of the prior $a = P$, $b = n$, the evidence for the null is a little weak. The Bayesian framework helps us to investigate the sensitivity of our study to variations of prior distributions. If we feel that our prior information will have a very little impact on the information in the data, then the BF can be used as an alternative to the classical approach. The assumption of the conjugate family of distributions, reduces the burden of numerical computation significantly. The classical approach developed in this article for comparing gamma populations is seen to have an edge over the Bayesian approach in the case of small sample sizes.

Acknowledgments

The authors thank Dr. James Webber of the Wadsworth Center of the New York State Department of Health for providing the Asbestos data and also for providing helpful comments.

Funding

This work was supported in part by grants from the National Institute of Health (R01 MH69353).

Appendix: Maximum Likelihood Estimates of Parameters

In this appendix, we discuss the maximum likelihood estimation procedure for estimating the model parameters under the assumption that each gamma population has its own scale and shape parameters. Thus, we have $\kappa_1, \kappa_2, \ldots, \kappa_p$ as shape parameters and $\beta_1, \beta_2, \ldots, \beta_p$ as scale parameters. Denote the vector of shape parameters by $\boldsymbol{\kappa}$ and the vector of scale parameters by $\boldsymbol{\beta}$.

The log-likelihood function for the complete data is given by

$$\mathcal{L}_0\left(\boldsymbol{\kappa}, \boldsymbol{\beta} | \mathbf{x}_1 \ldots \mathbf{x}_p\right) = \sum_{i=1}^{p} \left(n_i(\kappa_i - 1)\right) \ln\left(\tilde{X}_i\right) - \sum_{i=1}^{p} \left(n_i\right) \frac{\bar{X}_i}{\beta_i} - \sum_{i=1}^{p} n_i \left\{\ln\left[\Gamma\left(\kappa_i\right)\right]\right\} - \sum_{i=1}^{p} \left(n_i \kappa_i\right) \ln\left(\beta_i\right).$$

We take partial derivatives with respect to κ_i and β_i and solve the system of equations to obtain the corresponding estimates. Consider hypothesis H_{02} when the unknown shape parameters are equal. Under these assumptions, the log-likelihood function simplifies to

$$\mathcal{L}_0\left(\kappa, \beta | \mathbf{x}_1 \ldots \mathbf{x}_p\right) = (\kappa - 1) \sum_{i=1}^{p} n_i \ln\left(\tilde{X}_i\right) - \beta^{-1} \sum_{i=1}^{p} n_i \bar{X}_i - n \left\{\ln\left[\Gamma\left(\kappa\right)\right]\right\} - n\kappa \ln\left(\beta\right).$$

Under H_{02}, the maximum likelihood estimator of the common shape parameter, κ, satisfies

$$\hat{\kappa}\widehat{\beta} = \sum_{i=1}^{p} \nu_i \bar{X}_i, \quad \text{where } \nu_i = \frac{n_i}{n}.$$

It follows that

$$\ln\left(\frac{\prod_{i=1}^{p} \tilde{X}^{\nu_i}}{\sum_{i=1}^{p} \nu_i \bar{X}_i}\right) = \ln\left(W\right) = \psi\left(\kappa\right) - \ln\left(\kappa\right).$$

Therefore, the maximum likelihood estimator satisfies the same equation as in the one-sample problem with the exception that the ratio of the geometric mean of the observations within one sample is replaced with the weighted geometric mean of the p geometric means to the weighted arithmetic means of the individual samples. The statistic, W, treats the values as though they all came from the sample and has the same distribution as that of Bartlett's test statistic for equal variances among p independent and normally distributed populations which have respective sample sizes, $n_1, \ldots, n_p$. Hence, the test statistic follows the distribution laid out by Glaser (1973, 1976a, 1976b, 1980), Nandi (1980), and Dyer and Keating (1980).

References

Aksoy, H. (2000). Use of gamma distribution in hydrological analysis. *Turkish Journal of Engineering & Environmental Sciences* 2:419–428.

Aryal, S., Bhaumik, D. K., Mathew, T., Gibbons, R. D. (2008). Approximate tolerance limits and prediction limits for the gamma distribution. *Journal of Applied Statistical Science* 16:103–111.

Bain, L. J., Engelhardt, M. (1975). A two-moment chi-square approximation for the statistic $Log(\frac{\bar{x}}{\tilde{x}})$. *Journal of the American Statistical Association* 70:948–950.

Bain, L. J., Engelhardt, M., Shiue, W.-K. (1984). Approximate tolerance limits and confidence limits on reliability for the gamma distribution. *IEEE Transactions on Reliability* 33:184–187.

Barlow, R. E., Proschan, F. (1965), *Mathematical Theory of Reliability*. New York: John Wiley and Sons.

Basu, A., Rathouz, P. J. (2005). Estimating marginal and incremental effects on health outcomes using flexible link and variance function models. *Biostatistics* 6:93–109.

Bhaumik, D. K., Gibbons, R. D. (2006). One-sided approximate prediction intervals for at least p of m observations from a gamma population at each of r locations. *Technometrics* 48:112–119.

Bhaumik, D. K., Kapur, K., Gibbons, R. D. (2009). Testing parameters of a gamma distribution for small samples. *Technometrics* 51:326–334.

Bowman, K. O., Shenton, L. R. (1988), *Properties of Estimators for the Gamma Distribution*. New York: Marcel Dekker.

Crowder, M. (2001), *Classical Competing Risks*. New York: Chapman & Hall.

Das, S. C. (1995). Fitting truncated Type III curves to rainfall data. *Australian Journal of Physics* 8:298–304.

Davis, D. J. (1952). An analysis of some failure data. *Journal of the American Statistical Association* 47:113–150.

Dyer, D. D., Keating, J. P. (1980). On the determination of critical values for Bartlett's test. *Journal of the American Statistical Association* 75:313–319.

Fang, C. T., Chang, Y. Y., Hsu, H. M., Twu, S. J., Chen, K. T., Lin, C. C., Huang, L. Y., Chen, M. Y., Hwang, J. S., Wang, J. D., Chuang, C. Y. (2007). Life expectancy of patients with newly-diagnosed HIV infection in the era of highly active antiretroviral therapy. *QJM: An International Journal of Medicine* 100:97–105.

Feiveson, A. H., Kulkarni, P. M. (2000). Reliability of space-shuttle pressure vessels with random batch effects. *Technometrics* 42:332–344.

Fink, D. (1995). A compendium of conjugate priors. In progress report: Extension and enhancement of methods for setting data quality objectives. *DOE contract 95-831*.

Gibbons, R. D., Bhaumik, D. K., Aryal, S. (2009). *Statistical Methods for Groundwater Monitoring.* 2nd ed. Hoboken, New Jersey: John Wiley and Sons.

Gibbons, R. D., Coleman, D. E. (2001), *Statistical Methods For Detection And Quantification of Environmental Contamination*. New York: John Wiley and Sons.

Glaser, R. E. (1973). Inferences for a Gamma Distributed Random Variable with Both Parameters Unknown with Applications to Reliability. Tech. Rep. 154, Department of Statistics, Stanford University.

Glaser, R. E. (1976a). Exact critical values for Bartlett's test for homogeneity of variances. *Journal of the American Statistical Association* 71:488–490.

Glaser, R. E. (1976b). The ratio of the geometric mean to the arithmetic mean for a random sample from a gamma distribution. *Journal of the American Statistical Association* 71:480–487.

Glaser, R. E. (1980). A characterization of Bartlett's statistic involving incomplete beta functions. *Biometrika* 67:53–58.

Glaser, R. E. (1983). *Statistical analysis of Kevlar 49/epoxy composite stress-rupture data*. Tech. Rep. UCID-19849, Lawrence Livermore National Laboratory, Livermore, CA.

Jeffreys, H. (1935). Some tests of significance, treated by the theory of probability. *Proceedings of the Cambridge Philosophy Society* 31:203–222.

Johnson, N. L., Kotz, S., Balakrishnan, N. (1994). *Continuous Univariate Distributions*, Vol. 1, 2nd ed.New York: John Wiley and Sons.

Johnson, N. L., Kotz, S., Balakrishnan, N. (1995), *Continuous Univariate Distributions*, Vol. 2, 2nd ed.New York: John Wiley and Sons.

Johnson, V. E., Moosman, M., Cotter, P. (2005). A hierarchical model for estimating the reliability of complex systems. *IEEE Transactions on Reliability*, 54, 224–231.

Kass, R. E., Raftery, A. E. (1995). Bayes factor. *Journal of the American Statistical Association* 90:773–795.

Keating, J. P., Glaser, R. E., Ketchum, N. S. (1990). Testing hypotheses about the shape parameter of a gamma distribution. *Technometrics* 32:67–82.

Kim, Y. (2011). *Generalized Linear Mixed Model and Calibration for a Gamma Random Variable: Application to Asbestos Fibers*. Ph.D. thesis, University of Illinois at Chicago.

Kleyner, A., Bhagath, S., Gasparini, M., Robinson, J., Bender, M. (1997). Bayesian techniques to reduce the sample size in automotive electronics attribute testing. *Microelectronics and Reliability* 37:879–883.

Krishnamoorthy, K., Mathew, T., Mukherjee, S. (2008). Normal based methods for a Gamma distribution: Prediction and tolerance intervals and stress-strength reliability. *Technometrics* 50:69–78.

Manning, W. G., Basu, A., Mullahy, J. (2005). Generalized modeling approaches to risk adjustment of skewed outcomes. *Journal of Health Economics*, 24, 465–488.

Marshall, A. W. and Olkin, I. (1979), *Inequalities: Theory of Majorization and Its Applications*. New York: Academic Press.

Martz, H. F., Waller, R. A. (1982), *Bayesian Reliability Analysis*. New York: John Wiley and Sons.

Miller, R. B. (1980). Bayesian analysis of the two-parameter gamma distribution. *Technometrics* 22:65–69.

Nandi, S. B. (1980). On the exact distribution of a normalized ratio of weighted geometric mean to the unweighted arithmetic mean in samples from gamma distributions. *Journal of the American Statistical Association* 75:217–220.

Nelson, A. R., Liverman, C. T., Eide, E. A., Abt, E. (2009), *A Review of the NIOSH Roadmap for Research on Asbestos Fibers and Other Elongate Mineral Particles*. Washington, D.C.: The National Academies Press.

Pintille, M. (2006), *Competing Risks: A Practical Perspective*. West Sussex, England: John Wiley and Sons.

Rao, C. R. (1965), *Linear Statistical Inference and Its Applications*. New York: John Wiley and Sons.

Spiteri, P., Ho, S., Lee, Y. L. (1963). Assessment of bending fatigue limit for crankshaft sections with inclusion of residual stresses. *International Journal of Fatigue* 29:318–329.

Springer, M. D., Thompson, W. E. (1970). The distribution of products of beta, gamma and Gaussian random variables. *SIAM Journal on Applied Mathematics* 18:721–737.

Stephenson, D. B., Kumar, K., Doblas-Reyes, F. -J., Royer, J. F., Chauvin, F., Pezzulli, S. (1988). Extreme daily rainfall events and their impact on ensemble forecasts of the Indian monsoon. *Monthly Weather Review* 127:1954–1966.

Whitmore, G. A. and Neufeldt, A. H. (2008). An application of statistical models in mental health research. *Bulletin of Mathematical Biophysics* 32:563–579.

Wilson, E. B. and Hilferty, M. M. (1931). The distribution of chi-squares. *Proceedings of the National Academy of Sciences* 17:684–688.

Xiong, J., Shenoi, R. A., Gao, Z. (2002). Small sample theory for reliability design. *Journal of Strain Analysis* 37:87–92.

Zelen, M. (1974). Problems in cell kinetics and early detection of disease, In: Proschan, F., Serfling, R. J., eds. *Reliability and Biometry*. Philadelphia: SIAM.