Electronic Research Archive

https://www.aimspress.com/journal/era

ERA, 34(4): 2572–2589.
DOI: 10.3934/era.2026119
Received: 28 November 2025
Revised: 14 February 2026
Accepted: 03 March 2026
Published: 24 March 2026

Research article

Exact statistical inference for quantities of loggamma distribution

Bowen Liu[1,*] **and Malwane M. A. Ananda**[2]

[1] School of Science and Engineering, University of Missouri-Kansas City, 5110 Rockhill Road, Kansas City, MO 64110, USA

[2] Department of Mathematical Sciences, University of Nevada, Las Vegas, 4505 S Maryland Pkwy, Las Vegas, NV 89154, USA

* **Correspondence:** Email: bowen.liu@umkc.edu.

Abstract: Transformed gamma distributions are widely used in various fields such as survival analysis, reliability engineering, and environmental studies. Statistical inference for the quantities of the transformed gamma distributions is of great importance in these applications. However, traditional inference methods often rely on large sample approximations, which is not accurate for small sample sizes. To address this issue, we introduced an exact statistical inference framework for quantities of the transformed gamma distributions and applied the framework to loggamma distribution. The performance of the proposed framework was extensively evaluated via comprehensive simulation studies. The results showed that the proposed framework outperforms the parametric bootstrapping method in terms of coverage probability and type 1 error rate. Two real data applications are provided for illustrative purposes.

Keywords: generalized inference; exact inference; parametric inference; gamma distribution; loggamma distribution

1. Introduction

Modeling univariate data with parametric probability distributions is a widely used approach across various scientific fields, including hydrological data modeling, environmental analysis, and insurance loss assessment [1, 2]. Among parametric distributions, the gamma distribution and its transformed family are particularly valuable for analyzing positive and skewed data [3–6]. Their effectiveness in modeling hydrological and financial data is well established. For instance, the log Pearson type 3 distribution, a member of the transformed gamma family, is one of the most frequently used distributions in hydrologic frequency analysis [7]. Similarly, the inverse gamma distribution has found widespread applications in reliability and survival analysis [8, 9].

While the statistical modeling of gamma distributions and their variants has been extensively studied, statistical inference for their parameters and associated quantities remains a crucial area of investigation. Although maximum likelihood estimation (MLE) is widely used for parameter estimation, with inference typically based on asymptotic properties, this approach may not be optimal in all scenarios. Particularly, while asymptotic methods perform well with large samples, many real-world applications involve small sample sizes where such asymptotic inference may be inadequate.

Exact inference methods, which do not rely on large-sample approximations, are essential for providing accurate and reliable results in small-sample scenarios. The broad applications can be found in the literature, such as exact inference for quantities under different assumptions [10–16]. Recently, there is also a growing interest in algebraic exact inference framework within the paradigm of algebraic state space theory (ASST) [17, 18]. Thus, the exact inference methods are becoming increasingly important for practitioners and researchers, especially in contexts where sample sizes are limited or where the assumptions of asymptotic methods may not hold.

Weerahandi and Gamage [10] developed a comprehensive approach for conducting generalized inference on two-parameter continuous distributions. Their method enables the construction of exact tests and confidence intervals for quantities associated with these distributions, based on exact probability statements rather than approximations. They demonstrated its effectiveness for two-parameter gamma distributions through examples of testing gamma scale parameters and means with unknown parameters. Building on this work, Ananda [11] extended the approach to gamma quantiles and demonstrated its satisfactory performance through simulations.

In this paper, we extend this generalized approach to quantities associated with loggamma distribution. Section 2 provides a brief introduction to statistical inference for the gamma distribution. Section 3 provides a summary of generalized inference for two-parameter gamma distributions and introduces methods for the loggamma distribution. Section 4 presents simulation studies in comparison with the parametric bootstrapping method. Section 5 demonstrates the practical application of our proposed procedure using real datasets, with comparisons to conventional methods. Section 6 concludes with final remarks.

2. Preliminaries

2.1. Sufficient statistics

To comprehensively illustrate the generalized inference procedure, we will start by introducing the concept of sufficient statistics. A sufficient statistic is defined as follows:

Definition 1. *A statistic $T(\mathbf{X})$ is sufficient for parameter θ if the conditional distribution of the sample $\mathbf{X}$ given the statistic $T(\mathbf{X})$ does not depend on the parameter θ.*

The sufficiency principle states that if a statistic is sufficient for θ, then any inference on θ should depend on sample $\mathbf{X}$ only through the statistic $T(\mathbf{X})$.

Only through the definition of sufficiency, the derivation of sufficient statistics can be challenging. However, there are some useful theorems that can help us find sufficient statistics more easily. One of the most commonly used theorems is the factorization theorem. The Fisher–Neyman factorization theorem [19, 20] is stated as follows:

Theorem 1 (Factorization criteria)**.** *Let $f(\mathbf{x}|\theta)$ be the joint probability density function (PDF) or probability mass function (PMF) of a sample $\mathbf{X} = (X_1, X_2, ..., X_n)$. A statistic $T(\mathbf{X})$ is sufficient for parameter θ if and only if there exist non-negative functions g and h such that for all sample points $\mathbf{X}$ and all parameter points θ, the joint PDF or PMF can be factorized as*

$$f(\mathbf{x}|\theta) = g(T(\mathbf{x}), \theta)h(\mathbf{x}).$$

Moreover, to completely understand the generalized inference procedure, we also need to introduce the concept of minimal sufficiency. A minimal sufficient statistic is defined as follows:

Definition 2. *A sufficient statistic $T(\mathbf{X})$ is minimal sufficient for parameter θ if, for any other sufficient statistic $T'(\mathbf{X})$, there exists a function f such that $T(\mathbf{X}) = f(T'(\mathbf{X}))$.*

The Lehmann–Scheffé criteria [21] provides a useful method to find minimal sufficient statistics. The criteria is stated as follows:

Theorem 2 (Lehmann–Scheffé criteria)**.** *A statistic $T(\mathbf{X})$ is minimal sufficient for parameter θ if and only if, for any two sample points $\mathbf{x}$ and $\mathbf{y}$, the ratio of their joint PDF or PMF $f(\mathbf{x}|\theta)/f(\mathbf{y}|\theta)$ is constant as a function of θ if and only if $T(\mathbf{x}) = T(\mathbf{y})$.*

This theorem allows us to determine whether a sufficient statistic is minimal sufficient by examining the ratio of joint PDF or PMF for different sample points.

2.2. Gamma distribution and sufficiency

A gamma random variable X is associated with the following PDF:

$$f(x|\alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}, \quad x > 0,$$

where $\alpha > 0$ is the shape parameter and $\beta > 0$ is the scale parameter.

Let $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ be an independent identical distributed (IID) sample from the gamma distribution with parameters α and β. The joint PDF of the sample is given by

$$f(\mathbf{x}|\alpha, \beta) = \prod_{i=1}^{n} f(x_i|\alpha, \beta) = \frac{1}{[\Gamma(\alpha)\beta^\alpha]^n} \left(\prod_{i=1}^{n} x_i \right)^{\alpha-1} e^{-\frac{1}{\beta}\sum_{i=1}^{n} x_i} \mathbf{1}(x_{(1)} > 0). \tag{2.1}$$

Define $S = \sum_{i=1}^{n} X_i$ and $P = \prod_{i=1}^{n} X_i$. We shall prove the following propositions.

Proposition 1. *$T(\mathbf{x}) = (S, P)$ is sufficient for $\theta = (\alpha, \beta)$.*

Proof. With the joint PDF in Eq (2.1), define $T(\mathbf{X}) = (\prod_{i=1}^{n} x_i, \sum_{i=1}^{n} x_i)$. Then we can express

$$f(\mathbf{x}|\alpha, \beta) = g(T(\mathbf{x}), \alpha, \beta)\, h(\mathbf{x}),$$

where

$$g(T(\mathbf{x}), \alpha, \beta) = [\Gamma(\alpha)]^{-n} \beta^{-n\alpha} \left(\prod_{i=1}^{n} x_i \right)^{\alpha-1} e^{-\sum_{i=1}^{n} x_i/\beta}, \quad h(\mathbf{x}) = 1(x_{(1)} > 0).$$

Hence, by the factorization theorem, $T(\mathbf{X})$ is sufficient for (α, β).

Proposition 2. *$T(\mathbf{x}) = (S, P)$ is minimal sufficient for $\theta = (\alpha,\beta)$.*

Proof. By the Lehmann–Scheffé criterion, a statistic $T(\mathbf{X})$ is minimal sufficient for the parameter (α,β) if, for any two samples $\mathbf{x} = (x_1, \dots, x_n)$ and $\mathbf{y} = (y_1, \dots, y_n)$, the ratio

$$\frac{f(\mathbf{x}|\alpha,\beta)}{f(\mathbf{y}|\alpha,\beta)}$$

is constant in (α,β) if and only if $T(\mathbf{x}) = T(\mathbf{y})$.

From the joint PDF of the Gamma sample,

$$f(\mathbf{x}|\alpha,\beta) = \frac{1}{[\Gamma(\alpha)\beta^{\alpha}]^n}\left(\prod_{i=1}^{n} x_i^{\alpha-1}\right)\exp\left(-\frac{1}{\beta}\sum_{i=1}^{n} x_i\right).$$

Hence, for two samples $\mathbf{x}$ and $\mathbf{y}$,

$$\frac{f(\mathbf{x}|\alpha,\beta)}{f(\mathbf{y}|\alpha,\beta)} = \frac{\left(\prod_{i=1}^{n} x_i^{\alpha-1}\right)\exp\left(-\frac{1}{\beta}\sum_{i=1}^{n} x_i\right)}{\left(\prod_{i=1}^{n} y_i^{\alpha-1}\right)\exp\left(-\frac{1}{\beta}\sum_{i=1}^{n} y_i\right)}.$$

Simplifying, we get

$$\frac{f(\mathbf{x}|\alpha,\beta)}{f(\mathbf{y}|\alpha,\beta)} = \left(\frac{\prod_{i=1}^{n} x_i}{\prod_{i=1}^{n} y_i}\right)^{\alpha-1}\exp\left\{-\frac{1}{\beta}\left(\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} y_i\right)\right\}.$$

For this ratio to be constant in both α and β, the exponential and power terms must not depend on those parameters. Thus, the ratio is constant in (α,β) if and only if

$$\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \quad \text{and} \quad \prod_{i=1}^{n} x_i = \prod_{i=1}^{n} y_i.$$

Therefore,

$$T(\mathbf{X}) = \left(\sum_{i=1}^{n} X_i, \prod_{i=1}^{n} X_i\right)$$

satisfies the Lehmann–Scheffé condition and is hence a minimal sufficient statistic for (α,β).

Proposition 3. *$V = \frac{P^{1/n}}{S/n}$ is free of parameter β. That is, the distribution of V does not depend on β.*

Proof. Let $Y_i := X_i/\beta$ for $i = 1, \dots, n$. For $X \sim \Gamma(\alpha,\beta)$, the moment generating function is

$$M_X(t) = (1-\beta t)^{-\alpha}, \qquad t < 1/\beta.$$

Then for $Y = X/\beta$,

$$M_Y(t) = \mathbb{E}[e^{tY}] = \mathbb{E}[e^{(t/\beta)X}] = M_X(t/\beta) = (1-\beta(t/\beta))^{-\alpha} = (1-t)^{-\alpha}, \qquad t < 1.$$

Thus, $Y \sim \Gamma(\alpha, 1)$. Since $X_1, \dots, X_n$ are IID $\Gamma(\alpha,\beta)$, we have $Y_1, \dots, Y_n$ as IID $\Gamma(\alpha, 1)$.

Now write

$$S = \sum_{i=1}^{n} X_i = \beta\sum_{i=1}^{n} Y_i, \qquad \left(\prod_{i=1}^{n} X_i\right)^{1/n} = \beta\left(\prod_{i=1}^{n} Y_i\right)^{1/n}.$$

Therefore,

$$V = \frac{\left(\prod_{i=1}^{n} X_i\right)^{1/n}}{S/n} = \frac{\beta\left(\prod_{i=1}^{n} Y_i\right)^{1/n}}{(\beta\sum_{i=1}^{n} Y_i)/n} = \frac{\left(\prod_{i=1}^{n} Y_i\right)^{1/n}}{(\sum_{i=1}^{n} Y_i)/n},$$

which is a function only of $(Y_1, \ldots, Y_n)$ and hence free of β.

Remark 1. *The proof shows V is* scale-free *(independent of β). Writing $U_i = Y_i / \sum_{j=1}^{n} Y_j$ gives $V = n\left(\prod_{i=1}^{n} U_i\right)^{1/n}$ with $(U_1, \ldots, U_n) \sim \text{Dirichlet}(\alpha, \ldots, \alpha)$, so V depends on α only.*

Proposition 4. (S, V) *is sufficient for* (α, β).

Proof. Any one-to-one function of minimal sufficient statistic is also a minimal sufficient statistic. Hence, since there is a one-to-one correspondence defined between (V, S) and (S, P), (V, S) is also minimal suffcient for (α, β).

Proposition 5. $V \perp S$. *That is, V and S are statistically independent.*

Proof. Consider the one-to-one transformation:

$$S = \sum_{i=1}^{n} X_i, \qquad U_i = \frac{X_i}{S} \ (i = 1, \ldots, n-1), \qquad U_n = 1 - \sum_{i=1}^{n-1} U_i,$$

whose inverse is $X_i = S\, U_i$ for $i = 1, \ldots, n$. The range is $s > 0$, $u_i > 0$, and $\sum_{i=1}^{n} u_i = 1$.

The Jacobian matrix $J = \frac{\partial(x_1,\ldots,x_n)}{\partial(s,u_1,\ldots,u_{n-1})}$ has entries as follows:

$$\frac{\partial x_i}{\partial s} = u_i, \qquad \frac{\partial x_i}{\partial u_j} = \begin{cases} s & \text{if } i = j \le n-1, \\ -s & \text{if } i = n, \\ 0 & \text{otherwise.} \end{cases}$$

Thus,

$$J = \begin{pmatrix} u_1 & s & 0 & \cdots & 0 \\ u_2 & 0 & s & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ u_{n-1} & 0 & 0 & \cdots & s \\ u_n & -s & -s & \cdots & -s \end{pmatrix}.$$

Hence,

$$\left|\det J\right| = s^{n-1}.$$

The joint density of $\mathbf{X} = (X_1, \ldots, X_n)$ is

$$f_{\mathbf{X}}(\mathbf{x} \mid \alpha, \beta) = \frac{1}{[\Gamma(\alpha)\beta^{\alpha}]^n} \left(\prod_{i=1}^{n} x_i^{\alpha-1}\right) \exp\Big(-\tfrac{1}{\beta}\sum_{i=1}^{n} x_i\Big)\, \mathbf{1}_{\{x_i>0\}}.$$

Substitute $x_i = s\, u_i$ (with $u_n = 1 - \sum_{j=1}^{n-1} u_j$) and multiply by $|\det J|$:

$$\prod_{i=1}^{n} x_i^{\alpha-1} = s^{n(\alpha-1)} \prod_{i=1}^{n} u_i^{\alpha-1}, \qquad \sum_{i=1}^{n} x_i = s.$$

Therefore,

$$\begin{aligned} f_{S,\mathbf{U}}(s,\mathbf{u}) &= f_{\mathbf{X}}(s\mathbf{u})\,|\det J| \\ &= \frac{1}{[\Gamma(\alpha)\beta^{\alpha}]^{n}}\, s^{n(\alpha-1)} \left(\prod_{i=1}^{n} u_i^{\alpha-1}\right) e^{-s/\beta}\, s^{n-1}. \end{aligned}$$

Rearranging the terms, we have

$$f_{S,\mathbf{U}}(s,\mathbf{u}) = \frac{1}{[\Gamma(\alpha)\beta^{\alpha}]^{n}}\, s^{n\alpha-1} e^{-s/\beta} \left(\prod_{i=1}^{n} u_i^{\alpha-1}\right), \quad s > 0,\ \mathbf{u} \in \Delta_{n-1},$$

where $\Delta_{n-1} = \{\mathbf{u} \in \mathbb{R}^{n-1} : u_i > 0, \sum_{i=1}^{n-1} u_i = 1\}$.

The joint PDF factorizes as

$$f_{S,\mathbf{U}}(s,\mathbf{u}) = \underbrace{\frac{1}{\Gamma(n\alpha)\,\beta^{n\alpha}}\, s^{n\alpha-1} e^{-s/\beta}}_{f_S(s)} \times \underbrace{\frac{\Gamma(n\alpha)}{\prod_{i=1}^{n}\Gamma(\alpha)} \Big(\prod_{i=1}^{n} u_i^{\alpha-1}\Big)}_{f_{\mathbf{U}}(\mathbf{u})},$$

so $S \sim \text{Gamma}(n\alpha,\beta)$, $\mathbf{U} \sim \text{Dirichlet}(\alpha,\ldots,\alpha)$, and S and $\mathbf{U}$ are independent.

Finally,

$$V = \frac{(\prod_{i=1}^{n} X_i)^{1/n}}{S/n} = n\left(\prod_{i=1}^{n} U_i\right)^{1/n}$$

depends only on $\mathbf{U}$, hence V is independent of S.

3. Generalized inference

3.1. Generalized inference for two-parameter distributions

The generalized inference procedure for any two-parameter continuous distribution was introduced by Weerahandi and Gamage [10]. Suppose $X_1, X_2, ..., X_n$ forms a random sample from a continuous distribution $f(x;\alpha,\beta)$. Assume that both parameters are not known. Assume that (S, T) are jointly minimal sufficient statistics for parameters (α,β). Consider β as the parameter of interest. Now let $U = U(T;\alpha,\beta) = \Pr(T \le t) = F_T(t;\alpha,\beta) \sim U(0,1)$, where $F_T(t;\alpha,\beta)$ denotes the cumulative distribution function (CDF) of T and $U(0,1)$ stands for a uniform distribution with the support $[0,1]$. At the observed value of $T = t$, denote the corresponding observed $u = U(t;\alpha,\beta) = F_T(t;\alpha,\beta)$. With the observed value of t and fixed β, we can treat $u = u(\alpha)$ as a function of α. Denote u^{-1} as the inverse function of u that satisfies $u^{-1}(u(\alpha)) = \alpha$. Define

$$R_b(T;\alpha,\beta,t) = u^{-1}(U(T)).$$

R_b indeed has the following properties:

- At the observed value of t, $R_b(T;\alpha,\beta,t) = u^{-1}(U(T)) = u^{-1}(u(\alpha)) = \alpha$.
- The distribution of R_b is free of the nuisance parameter α.

R_b is a random quantity that will be used to construct the generalized pivotal quantity (GPQ). Sometimes, R_b is denoted as $\hat{\alpha}(U)$ as an intermediate quantity to construct the subsequent generalized pivotal quantity.

For testing the hypothesis for the quantity of interest Q, assume Q is a function of α and β, where $Q = q(\alpha, \beta)$. Suppose the hypothesis is defined as follows:

$$\begin{cases} \mathbf{H}_0 : Q \leq Q_0, \\ \mathbf{H}_A : Q > Q_0, \end{cases}$$

where Q_0 is a prespecified value. Let $\beta = q^{-1}(Q_0, \alpha)$. That is, through the value of Q_0 and α, we can find the corresponding value of β. Thus, through the intermediate quantity R_b, we can find the corresponding value of β as $\hat{\beta}(U, Q_0) = q^{-1}(Q_0, \hat{\alpha}(U))$. Now we can construct the generalized pivotal quantity for testing the hypothesis. Consider the random variable $W(S; t) = F_{S|T=t}(S|t, \alpha, \beta) \sim U(0, 1)$. The distribution of W given $S = s$ is standard uniform and does not depend on $S = s$. Thus, the unconditional distribution of $W = W(S, T; \alpha, \beta)$ is also uniform and independent of S.

The generalized testing variable is defined as

$$R_0 = \frac{W}{w(s, t; \hat{\alpha}(U), \hat{\beta}(U, Q_0))}, \tag{3.1}$$

where $w(s, t; \hat{\alpha}(U), \hat{\beta}(U, Q_0))$ is the observed value of W at the observed data (s, t) and the estimated parameters $(\hat{\alpha}(U), \hat{\beta}(U, Q_0))$.

The generalized p-value for testing $\mathbf{H}_0$ is

$$\begin{aligned} p &= P(R_0 \leq 1) \\ &= P(W \leq w(s, t; \hat{\alpha}(U), \hat{\beta}(U, Q_0))), \end{aligned} \tag{3.2}$$

or $1 - p$ depending on the stochastic monotonicity of R. In real practice, the value of p is computed based on Monte Carlo methods by drawing the random samples from U.

3.2. Generalized inference for gamma distribution

Previous research has shown that the inference for the functions of gamma distribution parameters can be successfully performed with the above procedure [10]. In order to perform the procedure on gamma distribution, we need to find the sufficient statistics for the distribution. Let $X_1, X_2, ..., X_n$ be a random sample from a gamma distribution $X \sim G(\alpha, \beta)$. The PDF of X is as follows:

$$f(x|\alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-x/\beta}, \quad x > 0. \tag{3.3}$$

In Section 2, we have shown that $S = \sum X_i$ and $P = \prod X_i$ are minimal sufficient statistics for the parameters α and β with factorization criterion and the Lehmann–Scheffé criteria. However, S and P are not independent of each other. Bhaumik et al. [22, 23] utilized the following statistic:

$$T = \frac{P^{1/n}}{S/n}. \tag{3.4}$$

We have also shown in detail that the distribution of T is free of θ and the distribution of T is independent of S in Section 2.

Therefore, the procedure for two-parameter distributions can be utilized without finding the conditional distribution of S given T. However, the CDF of T is difficult to solve for, so an empirical CDF method will be used.

Consider testing the following hypothesis for a gamma sample:

$$\begin{cases} \mathbf{H}_0 : Q = q(\alpha,\beta) \le q_0, \\ \mathbf{H}_A : Q = q(\alpha,\beta) > q_0. \end{cases} \tag{3.5}$$

Let $\beta = q^{-1}(Q_0, \alpha)$. The pseudocode for computing the generalized p-value given data X is as follows (n is the sample size of X; N_1 specifies the number of replicates for simulation of the CDF of T; N_2 denotes the number of replicates for the computation of generalized p-value.):

Algorithm 1 Generalized inference procedure for gamma distribution

Input: $(X, n, N_1, N_2, q^{-1}(Q_0, \alpha), Q_0)$

1: Compute observed $T = t$ and $S = s$ with sample X.
2: Prepare the empirical CDF of T that evaluates at $T = t$: $F_T(t)$ with observed t, prespecified N_1, and sample size n.
3: Generate the uniform sample $u_1, u_2, ..., u_{N_2}$ with sample size being N_2.
4: For each $i \in 1, 2, ..., N_2$, numerically solve $F_T(t, \alpha_i) = u_i$ for α_i. Denote the solution as $\hat{k}_i$.
5: Find $\hat{\beta}_i = q^{-1}(Q_0, \hat{\alpha}_i)$.
6: Compute $w_i = \int_0^s \frac{1}{\Gamma(n\hat{\alpha}_i)\hat{\beta}_i^{n\hat{\alpha}_i}} x^{n\hat{\alpha}_i - 1} e^{-x/\hat{\beta}_i}\, dx$ for $i = 1, 2, ..., N_2$.
7: Find the mean of all w_i over N_2 replicates. Denote it as $p(q_0)$.
8: If R in (2) is stochastically decreasing in Q, then p-value equals $p(q_0)$; otherwise, p-value equals $1 - p(q_0)$. This is the p-value for testing the hypothesis in (3.5).
9: The $(1-\gamma)100\%$ equal-tail confidence interval can be found by finding $[q_L, q_U]$ such that $p(q_L) = \frac{\gamma}{2}$ and $p(q_U) = 1 - \frac{\gamma}{2}$.

The construction of the confidence interval of the quantity is straightforward based on the property of the generalized pivotal quantity R. The percentiles of R will be used to construct the confidence intervals of the parameter of interest.

It should be noted that this generalized inference framework provides an exact probability statement in the sense that the constructed GPQ has a distribution free of nuisance parameters under the assumed model. Consequently, the resulting confidence intervals and tests possess exact finite-sample validity in principle, without relying on asymptotic approximations.

However, in practice, the distribution of the GPQ is evaluated via Monte Carlo simulation. Thus, the reported confidence limits and p-values are Monte Carlo approximations to the exact generalized inference procedure. The theoretical guarantee pertains to the pivotal construction itself, while the numerical accuracy depends on the Monte Carlo sample size.

3.3. Application to transformed gamma distributions

Suppose we define a family of distributions $\mathcal{F}$ such that for all $Y \in \mathcal{F}$, $Y = g(X)$, where X follows a gamma distribution and g is a continuous monotone function that does not introduce any new parameters. This family $\mathcal{F}$ contains a wide range of distributions such as the inverse gamma distribution ($g(X) = 1/X$), loggamma distribution ($g(X) = \log(X)$), transformed gamma distribution ($g(X) = X^{1/\tau}, \tau > 0$) [1,2], and the inverse transformed gamma distribution ($g(X) = X^{1/\tau}, \tau > 0$) [1,2]. As we provided the procedures for generalized inference of quantities of the gamma distribution, we now extend these procedures to the family $\mathcal{F}$, starting with the following theorem.

Theorem 3. *Suppose $X = \{X_i : i = 1, 2, ..., n\}$ is an IID sample from a gamma distribution with shape parameter α and scale parameter β. Let $Y = \{Y_i : Y_i = g(X_i), i = 1, 2, ..., n\}$, where g is monotone and g^{-1} has a continuous derivative on the domain of Y_i for all $i = 1, 2, ..., n$. Let $S = \sum X_i$, $P = \prod X_i$, and $T = \frac{P^{1/n}}{S/n}$. Suppose S and T are jointly sufficient statistics derived for the gamma sample X. Then $S^* = \sum g^{-1}(Y_i)$ and $T^* = \frac{(\prod g^{-1}(Y_i))^{1/n}}{S^*/n}$ are jointly sufficient for the parameters α and β associated with Y.*

Proof. Since S and T are jointly sufficient for α and β for the gamma sample X, we have

$$\begin{aligned} f_X(x|\alpha,\beta) &= \prod_{i=1}^{n} f_X(x_i|\alpha,\beta) \\ &= \frac{1}{[\Gamma(\alpha)\beta^\alpha]^n}\left(\prod_{i=1}^{n} x_i^{\alpha-1}\right)\exp\left(-\frac{1}{\beta}\sum_{i=1}^{n} x_i\right). \end{aligned}$$

The joint PDF of Y is then

$$\begin{aligned} f_Y(y|\alpha,\beta) &= \prod_{i=1}^{n} f_X(g^{-1}(y_i))\left|\frac{dg^{-1}(y_i)}{dy_i}\right| \\ &= \frac{1}{[\Gamma(\alpha)\beta^\alpha]^n}\left(\prod_{i=1}^{n} g^{-1}(y_i)^{\alpha-1}\right)\exp\left(-\frac{1}{\beta}\sum_{i=1}^{n} g^{-1}(y_i)\right)\prod_{i=1}^{n}\left|\frac{dg^{-1}(y_i)}{dy_i}\right|. \end{aligned}$$

Define

$$S^* = \sum_{i=1}^{n} g^{-1}(y_i), \quad P^* = \prod_{i=1}^{n} g^{-1}(y_i), \quad T^* = \frac{(P^*)^{1/n}}{S^*/n}.$$

Substituting these into the joint PDF, we can rewrite

$$f_Y(y|\alpha,\beta) = h(y)\ \underbrace{\frac{1}{[\Gamma(\alpha)\beta^\alpha]^n}(T^*)^{n(\alpha-1)}(S^*)^{n(\alpha-1)}\exp\left(-\frac{S^*}{\beta}\right)}_{\text{depends only on } S^*,\ T^*,\ \alpha,\ \beta},$$

where

$$h(y) = \prod_{i=1}^{n}\left|\frac{dg^{-1}(y_i)}{dy_i}\right| \Big/ [(T^*)^{n(\alpha-1)}(S^*)^{n(\alpha-1)}]$$

is free of α and β.

Thus, the joint PDF factorizes into a product of two components: one depending on $(S^*, T^*, \alpha, \beta)$ and another depending only on the sample y. By the factorization theorem, (S^*, T^*) are jointly sufficient for (α, β).

Since $X_i = g^{-1}(Y_i)$, it follows that $S^* = S$ and $T^* = T$. Therefore, S and T are jointly sufficient for α and β associated with the transformed gamma sample Y. Moreover, by the property of minimal sufficient statistics, (S^*, T^*) are also minimal sufficient for (α, β) associated with the transformed gamma sample Y.

Therefore, for any inference on a transformed gamma sample, the generalized inference procedure developed for the gamma family can be directly utilized, provided an appropriate transformation function g is defined. In this paper, we focus on the loggamma distribution (denoted by LG). If $Y \sim LG(\alpha, \beta)$, the inference on functions of the parameters of a loggamma distribution can be performed using the generalized inference procedure described in Algorithm 1.

4. Simulations

We conducted comprehensive simulations to evaluate the performance of the generalized inference and parametric bootstrapping methods when testing the 0.5^{th} and 0.9^{th} quantile of loggamma distributions.

Let X be a gamma random variable with the PDF defined in 3.3. Consider $Y = \exp(X)$. The PDF of Y can be found by utilizing the transformation method for continuous random variables. Specifically, the PDF of $LG(\alpha, \beta)$ can be derived as follows:

$$f_Y(y) = f_X(\ln y)\left|\frac{dx}{dy}\right| = \frac{1}{\Gamma(\alpha)\beta^\alpha}(\ln y)^{\alpha-1}e^{-(\ln y)/\beta}\frac{1}{y} = \frac{1}{\Gamma(\alpha)\beta^\alpha}(\ln y)^{\alpha-1}y^{-(1+1/\beta)}, \qquad y > 1.$$

Thus, the quantile function of $LG(\alpha, \beta)$ can be expressed as follows:

$$Q(p|\alpha, \beta) = \exp(G^{-1}(p|\alpha, \beta)),$$

where $G^{-1}(p|\alpha, \beta)$ is the quantile function of gamma distribution with shape parameter α and scale parameter β. The likelihood function of a random sample $Y_1, Y_2, ..., Y_n$ from $LG(\alpha, \beta)$ is given by

$$L(\alpha, \beta \mid Y_1, \dots, Y_n) = \left[\frac{1}{\Gamma(\alpha)\beta^\alpha}\right]^n \prod_{i=1}^{n}(\ln y_i)^{\alpha-1}\, y_i^{-\left(1+\frac{1}{\beta}\right)}, \quad \alpha > 0,\ \beta > 0,\ y_i > 1.$$

Let $\hat{\alpha}$ and $\hat{\beta}$ be the MLE of α and β, respectively, based on the observed data. The parametric bootstrapping procedure is as follows:

- Step 1: Generate a random sample of size n from $LG(\hat{\alpha}, \hat{\beta})$ and compute the MLE of α and β, denoted by $\hat{\alpha}^*$ and $\hat{\beta}^*$, respectively.
- Step 2: Repeat Step 1 for B times to obtain B replicates of $(\hat{\alpha}^*, \hat{\beta}^*)$.
- Step 3: For each bootstrap replicate, compute the target quantile $q_p^* = Q(p \mid \hat{\alpha}^*, \hat{\beta}^*)$. Using the B values $\{q_p^*\}_{b=1}^{B}$, form the $(1-\gamma)$ confidence interval for the quantile via the percentile method, i.e., take the $\gamma/2$ and $1-\gamma/2$ empirical quantiles of $\{q_p^*\}$.

We considered various combinations of shape parameters ($\alpha = 1, 2$), scale parameters ($\beta = 0.5, 1$), and sample sizes ($n = 10, 15, 20$). For each combination, we generated 1000 random samples from the specified loggamma distribution. We then applied both the generalized inference and parametric bootstrapping methods to construct 90% confidence intervals for the p-th quantile and evaluated their coverage probabilities. Additionally, we assessed the type 1 error rates for testing the null hypothesis $H_0 : q \leq q_p$ at a significance level of $\gamma = 0.05$, where q_p stands for the p^{th} theoretical quantile for the corresponding loggamma distribution. The results of the simulation study are summarized in Tables 1 and 2.

For the implementation of generalized inference, the required distribution functions are evaluated using Monte Carlo estimators of the form

$$\widehat{F}(x) = \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}\{R_i \leq x\},$$

where R_i denotes the i-th simulated generalized pivotal quantity generated according to Algorithm 1. Specifically, each R_i is constructed by generating auxiliary random variables (e.g., uniform variates), obtaining the corresponding generalized parameter values, and computing the resulting generalized test quantity. Thus, $\widehat{F}(x)$ represents the empirical cumulative distribution function of the simulated generalized variables.

Confidence limits are obtained by numerically solving the corresponding estimating equation using Brent's bracketing root-finding algorithm (implemented via `uniroot` in R), which combines bisection, secant, and inverse quadratic interpolation steps to ensure stable and guaranteed convergence within the specified search interval.

For computational efficiency, we set $N_1 = 500$ and $N_2 = 1000$ in Algorithm 1, where N_1 controls the Monte Carlo approximation of the pivotal distribution and N_2 determines the number of generalized samples used to evaluate probabilities. For the parametric bootstrapping procedure, we set $B = 10{,}000$.

Table 1. Performance of the generalized inference and parametric bootstrapping when testing the 0.9^{th} quantile of loggamma distributions. (GI: generalized inference; PB: parametric bootstrapping)

α	β	Sample size n	Coverage probability (90% CI)		Type 1 error: $H_0 : q \leq q_{0.9}$ ($\gamma = 0.05$)	
			GI	PB	GI	PB
1	0.5	10	0.913	0.782	0.050	0.216
1	0.5	15	0.896	0.822	0.061	0.173
1	0.5	20	0.903	0.864	0.056	0.135
1	1	10	0.939	0.842	0.061	0.153
1	1	15	0.932	0.868	0.062	0.129
1	1	20	0.915	0.883	0.069	0.107
2	0.5	10	0.949	0.779	0.029	0.218
2	0.5	15	0.927	0.837	0.034	0.161
2	0.5	20	0.902	0.867	0.059	0.132
2	1	10	0.906	0.820	0.094	0.178
2	1	15	0.907	0.832	0.092	0.166
2	1	20	0.915	0.868	0.084	0.129

Table 2. Performance of the generalized inference and parametric bootstrapping when testing the 0.5^{th} quantile of loggamma distributions. (GI: generalized inference; PB: parametric bootstrapping)

α	β	Sample size n	Coverage probability (90% CI)		Type 1 error: $H_0 : q \leq q_{0.5}$ ($\gamma = 0.05$)	
			GI	PB	GI	PB
1	0.5	10	0.913	0.828	0.040	0.096
1	0.5	15	0.905	0.935	0.047	0.039
1	0.5	20	0.906	0.945	0.042	0.028
1	1	10	0.918	0.918	0.026	0.046
1	1	15	0.902	0.930	0.048	0.037
1	1	20	0.902	0.929	0.045	0.037
2	0.5	10	0.907	0.894	0.026	0.062
2	0.5	15	0.906	0.935	0.038	0.038
2	0.5	20	0.905	0.929	0.037	0.035
2	1	10	0.912	0.893	0.018	0.054
2	1	15	0.909	0.924	0.036	0.035
2	1	20	0.887	0.924	0.043	0.035

From the simulation results presented in Tables 1 and 2, it is evident that the generalized inference method consistently outperforms the parametric bootstrapping method in terms of coverage probabilities for the 90% confidence intervals across all combinations of shape parameters, scale parameters, and sample sizes. The generalized inference method achieves coverage probabilities close to the nominal level of 90%, while the parametric bootstrapping method exhibits significantly lower coverage probabilities, particularly for smaller sample sizes. Furthermore, the type 1 error rates for testing the null hypothesis $H_0 : q \leq q_p$ indicate that the generalized inference method maintains error rates close to the nominal level of 0.05, whereas the parametric bootstrapping method shows inflated type 1 error rates, especially for smaller sample sizes. These findings suggest that the generalized inference method is more reliable and accurate for inference on the 0.9^{th} quantile of loggamma distributions compared to the parametric bootstrapping method.

5. Real data applications

In this section, we applied the generalized inference procedure for the loggamma distribution. Specifically, we compared the performance of the generalized inference procedure with that of the parametric bootstrapping procedure. The real data analysis is mainly illustrative for loggamma inference with small sample sizes. Goodness of fit tests were conducted to evaluate the applicability of the loggamma distribution for the data sets.

5.1. Example 1: The fatigue life of deep-groove ball bearings

The data set contains the number of a million revolutions before failure for each of the 23 ball bearings in the reliability test. Previously, the data set was utilized by researchers to determine the service life of ball bearings [24]. The data are as follows:

17.88	28.92	33.00	41.52	42.12	45.60	48.80	51.84	51.96
54.12	55.56	67.80	68.64	68.64	68.88	84.12	93.12	98.64
105.12	105.84	127.92	128.04	173.40				

Previous analyses used Weibull and Gamma models for the bearing fatigue life. The loggamma distribution is also suitable because of its heavy-tail feature. Goodness of fit was evaluated with the Kolmogorov–Smirnov (KS) and Anderson–Darling (AD) tests. The KS test statistic is defined as

$$D = \sup_x |F_n(x) - F(x|\hat{\alpha}, \hat{\beta})|,$$

where $F_n(x)$ is the empirical CDF and $F(x|\hat{\alpha}, \hat{\beta})$ is the CDF of loggamma distribution with MLEs $\hat{\alpha}$ and $\hat{\beta}$. The range of D is between 0 and 1, and a smaller value of D indicates a better fit.

The AD test statistic is defined as follows:

$$A_n = -n - \frac{1}{n} \sum_{i=1}^{n} (2i - 1)[\ln F(Y_{(i)}|\hat{\alpha}, \hat{\beta}) + \ln(1 - F(Y_{(n-i+1)}|\hat{\alpha}, \hat{\beta}))],$$

where $Y_{(i)}$ is the i-th order statistic of the sample. Similar to KS test, a smaller value of A_n indicates a better fit. The p-values for both tests are obtained through Monte Carlo simulations.

Table 3 illustrates the test results when comparing the fitted model to the real data. KS test suggests that the loggamma distribution is a good fit for the ball bearing data ($D = 0.11$, $p = 0.99$). AD test also supports the same conclusion ($A_n = 0.34$, $p = 0.90$).

Table 3. Goodness of fit test for loggamma distribution with ball endurance data.

Model	MLE	KS test statistic (p-value)	AD test statistic (p-value)
Loggamma	$\hat{\alpha} = 60.57$ $\hat{\beta} = 0.069$	$D = 0.13$ $(p = 0.99)$	$A_n = 0.34$ $(p = 0.90)$

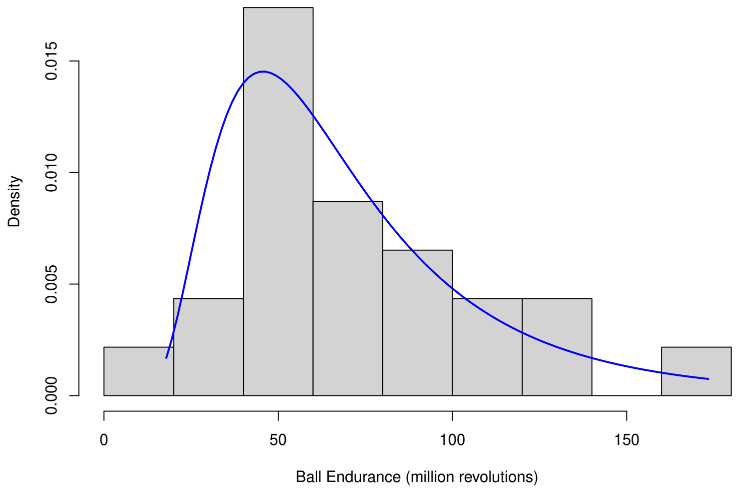

Figure 1. Histogram of ball endurance data with fitted loggamma PDF.

The histogram of the data with the fitted loggamma PDF is shown in Figure 1.

Figure 2 presents the comparison of lengths of 90% confidence intervals for the shape parameter α and scale parameter β using the generalized inference method and the parametric bootstrapping method. It can be observed that the lengths of the confidence intervals obtained from the generalized inference method are shorter than those from the parametric bootstrapping method for both parameters, indicating that the generalized inference method provides more precise estimates in this case.

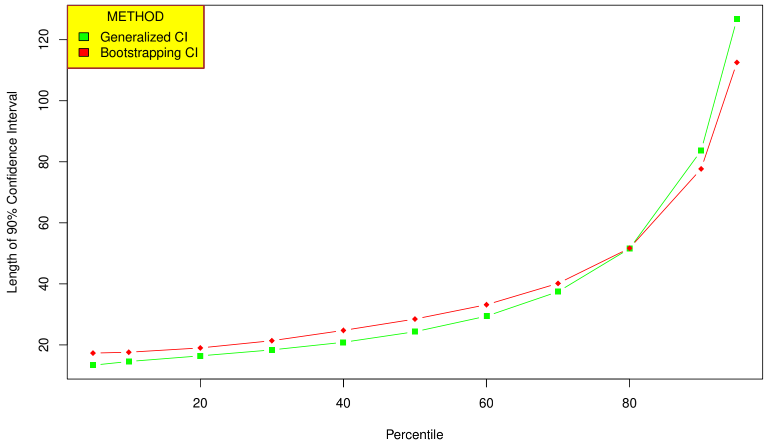

Figure 2. The comparison of lengths of 90% confidence intervals when using two different methods.

5.2. *Example 2: Stream flows for the Harricana River*

The second data set contains the maximum river discharges (in m^3/s) recorded during the month of September over the period of 1940–1968 for the Harricana River in Canada.

119.0	126.0	39.9	101.0	43.3	62.6	51.0	49.6	32.8
61.4	18.7	39.1	88.6	49.6	92.6	22.9	70.8	65.4
86.1	62.0	84.7	106.0	26.8	81.6	38.8	117.0	65.7

This hydrological data was previously used by researchers [11, 25] to investigate the confidence intervals for upper quantiles. Similar to the first example, we fitted a loggamma distribution to the data and evaluated the goodness of fit using the KS and AD tests. The results are summarized in Table 4. Both the KS test ($D = 0.11, p = 0.99$) and AD test ($A_n = 0.42, p = 0.82$) indicate that the loggamma distribution is an appropriate model for the Harricana River data.

Table 4. Goodness-of-fit test for loggamma distribution with Harricana River data.

Model	MLE	KS test statistic (p-value)	AD test statistic (p-value)
Loggamma	$\hat{\alpha} = 62.20$ $\hat{\beta} = 0.066$	$D = 0.11$ $(p = 0.99)$	$A_n = 0.42$ $(p = 0.82)$

The histogram of the data with the fitted loggamma PDF is shown in Figure 3.

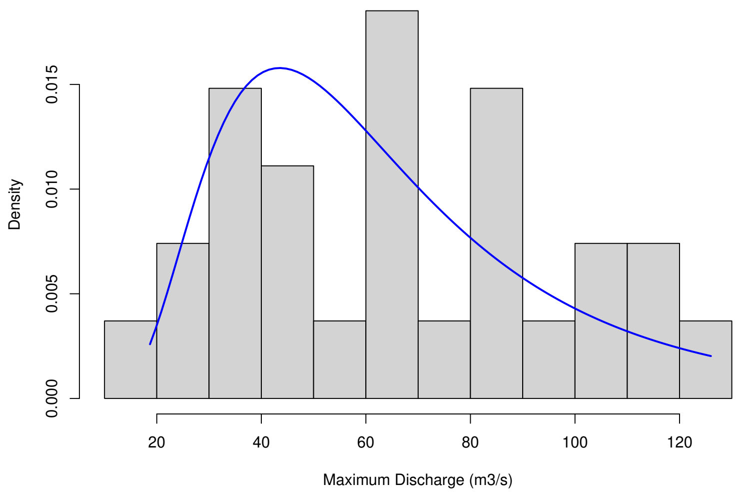

Figure 3. Histogram of Harricana River data with fitted loggamma PDF.

Figure 4 presents the comparison of lengths of 90% confidence intervals for the shape parameter α and scale parameter β using the generalized inference method and the parametric bootstrapping method. Similar to the first example, the lengths of confidence intervals obtained from the generalized inference method are shorter than those from the parametric bootstrapping method for both parameters, indicating that the generalized inference method provides more precise estimates in this case.

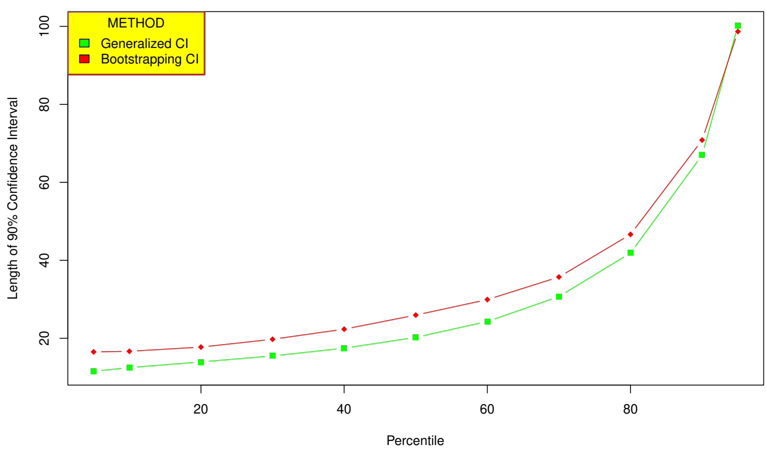

Figure 4. The comparison of lengths of 90% confidence intervals when using two different methods.

6. Conclusions

Statistical inference for quantities of gamma-related distributions has been an important topic in statistics due to their wide applications in various fields. In this paper, we have demonstrated that the previously proposed generalized inference method can be effectively applied to the transformed gamma distributions to construct confidence intervals for quantities of interest that can be expressed as functions of the shape and rate parameters of the underlying gamma distribution. We have conducted simulation studies to evaluate the performance of the proposed method in terms of coverage probability and average length of the confidence intervals. The results indicate that our method provides satisfactory coverage probabilities close to the nominal level while maintaining shorter average lengths compared to the parametric boostrapping method under the small sample size assumption. Furthermore, we have applied the generalized inference procedure for constructing confidence intervals for the quantiles of the log-gamma distribution. Through two real data examples, we have demonstrated the performance of our proposed method in providing more precise estimates compared to the parametric bootstrapping method. Since asymptotic inference depends on large-sample approximations and parametric bootstrapping typically relies on point estimates (such as MLEs) whose accuracy itself is justified asymptotically, the proposed method provides a practical and reliable alternative for small-sample settings where traditional approaches may perform poorly. However, it is important to note that the generalized inference method may involve higher computational costs due to the simulation-based nature of the approach. Future research could focus on optimizing the computational efficiency of the method and exploring its applicability to other distributions and more complex models.

Use of AI tools declaration

Grammarly was used to correct the grammar structures and spellings in the creation of this article.

Conflict of interest

The authors declare there are no conflicts of interest.

References

1. C. Kleiber, S. Kotz, *Statistical Size Distributions in Economics and Actuarial Sciences*, John Wiley & Sons, 2003. https://doi.org/10.1002/0471457175
2. S. A. Klugman, H. H. Panjer, G. E. Willmot, *Loss Models: From Data to Decisions*, John Wiley & Sons, 2012.
3. R. L. Prentice, A log gamma model and its maximum likelihood estimation, *Biometrika*, **61** (1974), 539–544. https://doi.org/10.2307/2334737
4. P. Veazie, O. Intrator, B. Kinosian, C. S. Phibbs, Better performance for right-skewed data using an alternative gamma model, *BMC Med. Res. Methodol.*, **23** (2023), 298. https://doi.org/10.1186/s12874-023-02113-1

5. M. Amini, S. M. T. K. MirMostafaee, J. Ahmadi, Log-gamma-generated families of distributions, *Statistics*, **48** (2014), 913–932. https://doi.org/10.1080/02331888.2012.748775

6. B. Liu, M. M. A. Ananda, Analyzing insurance data with an exponentiated composite inverse gamma-Pareto model, *Commun. Stat. Theory Methods*, **52** (2023), 7618–7631. https://doi.org/10.1007/978-94-017-1431-0_15

7. V. P. Singh, Log-Pearson type III distribution, in *Entropy-based Parameter Estimation in Hydrology*, Springer, (1998), 252–274. https://doi.org/10.1007/978-94-017-1431-0

8. A. G. Glen, On the inverse gamma as a survival distribution, in *Computational Probability Applications*, Springer, (2017), 15–30. https://doi.org/10.1007/978-3-319-43317-2_2

9. C. T. Lin, B. S. Duran, T. O. Lewis, Inverted gamma as a life distribution, *Microelectron. Reliab.*, **29** (1989), 619–626. https://doi.org/10.1016/0026-2714(89)90352-1

10. S. Weerahandi, J. Gamage, A general method of inference for two-parameter continuous distributions, *Commun. Stat. Theory Methods*, **45** (2016), 2612–2625. https://doi.org/10.1080/03610926.2014.887109

11. M. Ananda, Generalized statistical inference for quantiles of two-parameter gamma distribution, *J. Math. Stat.*, **20** (2024), 1–12. https://doi.org/10.3844/jmssp.2024.1.12

12. S. M. Jordan, K. Krishnamoorthy, Exact confidence intervals for the common mean of several normal populations, *Biometrics*, **52** (1996), 77–86. https://doi.org/10.2307/2533146

13. P. L. H. Yu, Y. Sun, B. K. Sinha, On exact confidence intervals for the common mean of several normal populations, *J. Stat. Plann. Inference*, **81** (1999), 263–277. https://doi.org/10.1016/S0378-3758(99)00052-X

14. S. Weerahandi, C. R. Yu, Exact distributions of statistics for making inferences on mixed models under the default covariance structure, *J. Stat. Distrib. Appl.*, **7** (2020), 4. https://doi.org/10.1186/s40488-020-00105-w

15. S. Weerahandi, Generalized confidence intervals, *J. Am. Stat. Assoc.*, **88** (1993), 899–905. https://doi.org/10.2307/2290779

16. K. W. Tsui, S. Weerahandi, Generalized p-values in significance testing of hypotheses in the presence of nuisance parameters, *J. Am. Stat. Assoc.*, **84** (1989), 602–607. https://doi.org/10.2307/2289949

17. Y. Yan, D. Cheng, J. E. Feng, H. Li, J. Yue, Survey on applications of algebraic state space theory of logical systems to finite state machines, *Sci. China Inf. Sci.*, **66** (2023), 111201. https://doi.org/10.1007/s11432-022-3538-4

18. G. K. K. King, O. Papaspiliopoulos, M. Ruggiero, Exact inference for a class of hidden Markov models on general state spaces, *Electron. J. Stat.*, **15** (2021), 2832–2875. https://doi.org/10.1214/21-EJS1841

19. R. A. Fisher, On the mathematical foundations of theoretical statistics, *Philos. Trans. A Math. Phys. Eng. Sci.*, **222** (1922), 309–368. https://doi.org/10.1098/rsta.1922.0009

20. J. Neyman, Su un teorema concernente le cosiddette statistiche sufficienti, *G. Ist. Ital. Attuari*, **6** (1935), 320–334.

21. E. L. Lehmann, H. Scheffé, Completeness, similar regions, and unbiased estimation-Part I, in *Selected Works of E. L. Lehmann*, Springer, (2011), 233–268. https://doi.org/10.1007/978-1-4614-1412-4_23

22. D. K. Bhaumik, K. Kapur, R. D. Gibbons, Testing parameters of a gamma distribution for small samples, *Technometrics*, **51** (2009), 326–334. https://doi.org/10.1198/tech.2009.07038

23. R. E. Glaser, The ratio of the geometric mean to the arithmetic mean for a random sample from a gamma distribution, *J. Am. Stat. Assoc.*, **71** (1976), 480–487. https://doi.org/10.2307/2285338

24. J. Lieblein, M. Zelen, Statistical investigation of the fatigue life of deep-groove ball bearings, *J. Res. Natl. Bur. Stand.*, **57** (1956), 273–316. http://doi.org/10.6028/jres.057.033

25. F. Ashkar, T. B. M. J. Ouarda, Approximate confidence intervals for quantiles of gamma and generalized gamma distributions, *J. Hydrol. Eng.*, **3** (1998), 43–51. https://doi.org/10.1061/(ASCE)1084-0699(1998)3:1(43)

AIMS Press