Estimation of $P[Y \mid X]$ for generalized Pareto distribution

manijeh mahmoodi
sadegh rezaei
rasool tahmasbi
Estimation of $P[Y < X]$ for generalized Pareto distribution

Sadegh Rezaei$^a$, Rasool Tahmasbi$^b$,*, Manijeh Mahmoodi$^b$

$^a$Department of Statistics, Shahid Chamran University of Ahvaz, Ahvaz, Iran
$^b$Department of Statistics, Amirkabir University of Technology (Polytechnic), Tehran, Iran

ARTICLE INFO

Article history:
Received 28 December 2008
Received in revised form 23 July 2009
Accepted 24 July 2009
Available online 4 August 2009

Keywords:
Bayes estimator
Generalized Pareto distribution
Maximum likelihood estimator
Bootstrap confidence intervals
Asymptotic distributions

ABSTRACT

This paper deals with the estimation of $P[Y < X]$ when $X$ and $Y$ are two independent generalized Pareto distributions with different parameters. The maximum likelihood estimator and its asymptotic distribution are obtained. An asymptotic confidence interval of $P[Y < X]$ is constructed using the asymptotic distribution. Assuming that the common scale parameter is known, MLE, UMVUE, Bayes estimation of $R$ and confidence interval are obtained. The ML estimator of $R$, asymptotic distribution and Bayes estimation of $R$ in general case is also studied. Monte Carlo simulations are performed to compare the different proposed methods.

© 2009 Elsevier B.V. All rights reserved.

1. Introduction

In reliability contexts, inferences about $R = P[Y < X]$, where $X$ and $Y$ independent distributions, are a subject of interest. For example, in mechanical reliability of a system, if $X$ is the strength of a component which is subject to stress $Y$, then $R$ is a measure of system performance. The system fails, if at any time the applied stress is greater than its strength.

The problem of estimating of $R$ have been widely used in the statistical literature. The maximum likelihood estimator (MLE) of $P[Y < X]$, when $X$ and $Y$ are normally distributed, has been considered by Downtown (1973), Govindarajulu (1967), Woodward and Kelley (1977) and Owen et al. (1977). Tong (1974, 1977) considered the estimation of $P[Y < X]$ when $X$ and $Y$ are independent exponential random variables. Awad et al. (1981) considered the MLE of $R$, when $X$ and $Y$ have bivariate exponential distribution. Ahmad et al. (1997) and Surles and Padgett (1998, 2001) considered the estimation of $P[Y < X]$, where $X$ and $Y$ are Burr type X random variable. The gamma case is studied in Constantine and Karson (1986). The theoretical and practical results on the theory and applications of the stress–strength relationships in industrial and economic systems during the last decades are collected and digested in Kotz et al. (2003). The class of life-time distributions (in particular, exponential and gamma) is considered by Nadarajah (2003). Estimation of $P[Y < X]$ from logistic (Nadarajah, 2004a), Laplace (Nadarajah, 2004b), exponential case with common location parameter (Baklizi and El-Masri, 2004), Burr type III (Mokhlis, 2005), beta (Nadarajah, 2005a), gamma (Nadarajah, 2005b), bivariate exponential (Nadarajah and Kotz, 2006) and Weibull (Kundu and Gupta, 2006) distributions are also studied. Kundu and Gupta (2005) considered the estimation of $P[Y < X]$, when $X$ and $Y$ are independent generalized exponential distribution. Recently, inferences on reliability in two-parameter exponential stress–strength model (Krishnamoorthy et al., 2007) and ML estimation of system reliability for Gompertz distribution (Saraçoğlu and Kaya, 2007) are considered. Kakade et al. (2008) studied exponentiated Gumbel case and Abd Elfattah and Marwa (to appear) studied exponential case based on censored samples.

* Corresponding author.
E-mail addresses: srezaei@aut.ac.ir (S. Rezaei), rtahmasbi@rtssp.org, rtahmasbi@ieee.org (R. Tahmasbi), mmahmoodi@rtssp.org (M. Mahmoodi).

0378-3758/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.jspi.2009.07.024
In this paper, we focus on estimation of $R = P[Y < X]$, where $X$ and $Y$ follow the generalized Pareto (GP) distribution with different parameters. We obtain the MLE of $R$ and its asymptotic distribution. The asymptotic distribution is used to construct an asymptotic confidence interval. Two bootstrap confidence intervals of $R$ are also proposed. Assuming that the common scale parameter is known, the maximum likelihood estimator of $P[Y < X]$, confidence intervals, UMVUE and Bayes estimation of $R$ are obtained.

This paper is organized as follows: In the next Section, the GP distribution is introduced. Estimation of $R$ with common scale parameter is given in Section 3. In this section, the ML estimator of $R$, asymptotic distribution and bootstrap confidence intervals are presented. Estimation of $R$ if the common scale parameters are known is discussed in Section 4. In this section MLE, UMVUE and Bayes estimation of $R$ are discussed. In Section 5, the estimation of $R$ in general case is studied. The ML estimator of $R$, asymptotic distribution and Bayes estimation of $R$ are presented in Section 5. The different proposed methods have been compared using Monte Carlo simulations and their results have been reported in Section 6. In Section 7, a numerical example is illustrated and the results of different methods are compared.

2. Generalized Pareto distribution

A random variable $X$ is said to have generalized Pareto distribution, if its probability density function (pdf) is given by

$$f_{\xi, \mu, \sigma}(x) = \frac{1}{\sigma} \left(1 + \frac{x - \mu}{\sigma}\right)^{-(1/\xi + 1)}$$

where $\mu, \xi \in \mathbb{R}$ and $\sigma \in (0, +\infty)$. For convenience, we reparametrize this distribution by defining $\xi/\sigma = \lambda$, $1/\xi = \alpha$ and $\mu = 0$.

Therefore,

$$f(x) = \lambda(1 + \lambda x)^{-(\alpha + 1)}; \quad x > 0.$$  

The cumulative distribution function is defined by

$$F(x) = 1 - (1 + \lambda x)^{-\alpha}$$

for $\lambda > 0$ and $\alpha > 0$. Here $\alpha$ and $\lambda$ are the shape and scale parameters, respectively. It is also well known that this distribution has decreasing failure rate (DFR) property. This distribution is also known as Pareto distribution of the second type or Lomax distribution. Shi et al. (1999) used generalized Pareto distribution to estimate the size of the maximum inclusion in clean steels and application of this distribution to reinsurance is discussed by Kremer (1997).

3. Estimation of $R$ with common scale parameter

In this section, we investigate the properties of $R$, when the common scale parameter $\lambda$, is the same, and the general case is studied in Section 5.

3.1. Maximum likelihood estimator of $R$

To investigate the properties of $R$, denote by $GP(\alpha, \lambda)$ the distribution of reparametrized GP. Let $X \sim GP(\alpha, \lambda)$ and $Y \sim GP(\beta, \lambda)$, where $X$ and $Y$ are independent random variables. Therefore,

$$R = P[Y < X] = \int_0^\infty P(Y < X | X = x)P(X = x) \, dx$$

$$= \int_0^\infty \alpha \lambda(1 + \lambda x)^{-(\alpha + 1)}(1 - (1 + \lambda x)^{-\beta}) \, dx$$

$$= \frac{\beta}{\alpha + \beta}. \quad (1)$$

To compute the MLE of $R$, suppose $X_1, X_2, \ldots, X_n$ is a random sample from $GP(\alpha, \lambda)$ and $Y_1, Y_2, \ldots, Y_m$ is a random sample from $GP(\beta, \lambda)$. Therefore, the log-likelihood function of the observed sample is

$$L(\alpha, \beta, \lambda) = n \ln \alpha + m \ln \beta + (n + m) \ln \lambda - (x + 1) \sum_{i=1}^n \ln(1 + \lambda x_i) - (\beta + 1) \sum_{j=1}^m \ln(1 + \lambda y_j). \quad (2)$$
The MLEs of \( \alpha, \beta \) and \( \lambda \) say \( \hat{\alpha}, \hat{\beta}, \) and \( \hat{\lambda} \), respectively, can be obtained as the solutions of

\[
\frac{\partial L}{\partial \alpha} = \frac{n}{\alpha} - \sum_{i=1}^{n} \ln(1 + \lambda x_i), \tag{3}
\]

\[
\frac{\partial L}{\partial \beta} = \frac{m}{\beta} - \sum_{j=1}^{m} \ln(1 + \lambda y_j), \tag{4}
\]

\[
\frac{\partial L}{\partial \lambda} = \frac{n + m}{\alpha} - (\alpha + 1) \sum_{i=1}^{n} \frac{x_i}{1 + \lambda x_i} - (\beta + 1) \sum_{j=1}^{m} \frac{y_j}{1 + \lambda y_j}. \tag{5}
\]

From (3)-(5), we obtain

\[
\hat{\alpha} = \frac{n}{\sum_{i=1}^{n} \ln(1 + \lambda x_i)}, \tag{6}
\]

\[
\hat{\beta} = \frac{m}{\sum_{j=1}^{m} \ln(1 + \lambda y_j)}, \tag{7}
\]

and \( \hat{\lambda} \) can be obtained as the solution of the non-linear equation

\[
f(\hat{\lambda}) = \frac{n + m}{\hat{\alpha}} - \sum_{i=1}^{n} \frac{n}{1 + \lambda x_i} \sum_{i=1}^{n} x_i \ln(1 + \lambda x_i) - \sum_{j=1}^{m} \frac{m}{1 + \lambda y_j} \sum_{j=1}^{m} y_j \ln(1 + \lambda y_j) - \sum_{i=1}^{n} \frac{n}{1 + \lambda x_i} + \sum_{j=1}^{m} \frac{m}{1 + \lambda y_j}.
\]

Therefore, \( \hat{\lambda} \) can be obtained as a solution of the non-linear equation of the form

\[
g(\lambda) = \lambda,
\]

where

\[
g(\lambda) = (n + m) \left[ \sum_{i=1}^{n} \ln(1 + \lambda x_i) \sum_{i=1}^{n} x_i + \sum_{j=1}^{m} \ln(1 + \lambda y_j) \sum_{j=1}^{m} y_j \right]^{-1}.
\]

Since \( \hat{\lambda} \) is a fixed point solution of the non-linear equation (9), therefore, it can be obtained using an iterative scheme as follows:

\[
g(\lambda_{(j)}) = \lambda_{(j+1)}, \tag{10}
\]

where \( \lambda_{(j)} \) is the \( j \)th iterate of \( \hat{\lambda} \). The iteration procedure should be stopped when \( |\lambda_{(j)} - \lambda_{(j+1)}| \) is sufficiently small. Once we obtain \( \hat{\lambda} \), then \( \hat{\alpha} \) and \( \hat{\beta} \) can be obtained from (6) and (7), respectively. Since ML estimators are invariant, so the MLE of \( R \) becomes

\[
\hat{R} = \frac{\hat{\beta}}{\hat{\alpha} + \hat{\beta}}. \tag{11}
\]

3.2. Asymptotic distribution

In this section, the asymptotic distribution of \( \hat{\theta} = (\hat{\alpha}, \hat{\beta}, \hat{\lambda}) \) and the asymptotic distribution of \( \hat{R} \) are obtained. Based on the asymptotic distribution of \( \hat{\theta} \), the asymptotic confidence interval of \( R \) is derived. We denote the expected Fisher information matrix of \( \theta = (\alpha, \beta, \lambda) \) as \( I(\theta) = E[I(\theta)] \), where \( I = [l_{ij}]_{i,j=1,2,3} \) is the observed information matrix. I.e.,

\[
I(\theta) = -\begin{bmatrix} \frac{\partial^2 L}{\partial \alpha^2} & \frac{\partial^2 L}{\partial \alpha \beta} & \frac{\partial^2 L}{\partial \alpha \lambda} \\ \frac{\partial^2 L}{\partial \beta \alpha} & \frac{\partial^2 L}{\partial \beta^2} & \frac{\partial^2 L}{\partial \beta \lambda} \\ \frac{\partial^2 L}{\partial \lambda \alpha} & \frac{\partial^2 L}{\partial \lambda \beta} & \frac{\partial^2 L}{\partial \lambda^2} \end{bmatrix}.
\]

It is easy to see that

\[
l_{11} = \frac{n}{\alpha^2},
\]

\[
l_{12} = l_{21} = 0.
\]
\[ I_{13} = I_{31} = \sum_{i=1}^{n} \frac{x_i}{1 + \lambda x_i}, \]

\[ I_{22} = \frac{m}{\beta^2}, \]

\[ I_{23} = I_{32} = \sum_{j=1}^{m} \frac{y_j}{1 + \lambda y_j}, \]

\[ I_{33} = \frac{n + m}{\lambda^2} - (\alpha + 1) \sum_{i=1}^{n} \frac{x_i^2}{(1 + \lambda x_i)^2} - (\beta + 1) \sum_{j=1}^{m} \frac{y_j^2}{(1 + \lambda y_j)^2}. \]

Using the integrals of the form

\[ \int_{0}^{\infty} x^{\alpha-1} (1 + \lambda x)^{-\nu} dx = \lambda^{-\nu} B(r, v - r) \]

for \( 0 < r < v \), where \( B(x, y) \) is beta function, we have

\[ J_{11} = \frac{n}{\lambda^2}, \]

\[ J_{22} = \frac{m}{\beta^2}, \]

\[ J_{12} = J_{21} = 0, \]

\[ J_{13} = J_{31} = \frac{n \alpha}{\lambda} B(2, \alpha), \]

\[ J_{23} = J_{32} = \frac{m \beta}{\lambda} B(2, \beta), \]

\[ J_{33} = \frac{1}{\lambda} [(n + m) - n(\alpha + 1)B(3, \alpha) - m(\beta + 1)B(3, \beta)]. \]

**Theorem 1.** As \( n \to \infty \) and \( m \to \infty \) and \( n/m \to p \) then

\[ [\sqrt{\hat{x} - \alpha}, \sqrt{\hat{\beta} - \beta}, \sqrt{\hat{\lambda} - \lambda}] \to N_3(0, \mathbf{U}^{-1}(\alpha, \beta, \lambda)), \]

where

\[ \mathbf{U}(\alpha, \beta, \lambda) = \begin{bmatrix} u_{11} & 0 & u_{13} \\ 0 & u_{22} & u_{23} \\ u_{31} & u_{32} & u_{33} \end{bmatrix} \]

and

\[ u_{11} = \frac{1}{n} J_{11} = \frac{1}{\lambda^2}, \]

\[ u_{13} = u_{31} = \frac{1}{n} J_{13} = \frac{\alpha}{\lambda} B(2, \alpha), \]

\[ u_{22} = \frac{1}{m} J_{22} = \frac{1}{\beta^2}, \]

\[ u_{23} = u_{32} = \frac{\sqrt{p} J_{23}}{n} = \frac{1}{\sqrt{p}} \times \frac{\beta}{\lambda} B(2, \beta), \]

\[ u_{33} = \frac{1}{n} J_{33} = \frac{p + 1}{p \lambda^2} - \frac{\alpha(\alpha + 1)}{\lambda^2} B(3, \alpha) - \frac{\beta(\beta + 1)}{p \lambda^2} B(3, \beta). \]

**Proof.** The proof follows from the asymptotic normality of MLE. \(\square\)
Theorem 2. As \( n \to \infty \) and \( m \to \infty \) so that \( n/m \to p \) then

\[
\sqrt{m}(\hat{R} - R) \to N(0, B),
\]

(13)

where

\[
B = \frac{1}{k(x + \beta)^4} \left[ \beta^2(u_{22}u_{33} - u_{23}^2) - 2x\beta \sqrt{p}u_{23}u_{31} + 2^2p(u_{11}u_{33} - u_{13}^2) \right]
\]

and

\[
k = u_{11}u_{22}u_{33} - u_{11}u_{23}u_{32} - u_{13}u_{22}u_{31}.
\]

Proof. It is clear that

Using Theorem 1, the consistency and asymptotic normality of MLE, the proof is complete. \( \Box \)

Note that Theorem 2 can be used to construct asymptotic confidence intervals. To compute the confidence interval of \( R \), the variance \( B \) needs to be estimated. To estimate it, the empirical Fisher information matrix and the MLE estimates of \( \alpha, \beta \) and \( \lambda \) are used, as follows:

\[
\hat{u}_{11} = \frac{1}{\hat{x}};
\]

\[
\hat{u}_{13} = \hat{u}_{31} = \frac{\hat{\lambda}}{\hat{x}} B(3, \hat{\lambda}),
\]

\[
\hat{u}_{22} = \frac{1}{\hat{\beta}};
\]

\[
\hat{u}_{23} = \hat{u}_{32} = \frac{1}{\sqrt{p}} \times \frac{\hat{\beta}}{\lambda} B(2, \hat{\beta}),
\]

\[
\hat{u}_{33} = \frac{p + 1}{p \hat{x}} - \frac{\hat{\lambda}(\hat{x} + 1)}{\hat{x}} B(3, \hat{\lambda}) - \frac{1}{p \hat{x}} B(3, \hat{\beta}).
\]

3.3. Bootstrap confidence intervals

It is clear that the confidence intervals based on the asymptotic results do not perform very well for small sample size. So, two confidence intervals based on the parametric bootstrap methods are proposed: (i) percentile bootstrap method (Efron, 1982) and (ii) bootstrap-t method (Hall, 1988).

The algorithms for estimating the confidence intervals of \( R \) using both methods are illustrated below.

**Algorithm of the percentile bootstrap method:** Step 1: From the sample \( \{x_1, x_2, \ldots, x_n\} \) and \( \{y_1, y_2, \ldots, y_m\} \), compute \( \hat{x}, \hat{\beta} \) and \( \hat{\lambda} \).

Step 2: Use \( \hat{x} \) and \( \hat{\lambda} \) to generate a bootstrap sample \( \{x_1^*, x_2^*, \ldots, x_n^*\} \) and similarly use \( \hat{\beta} \) and \( \hat{\lambda} \) to generate a sample \( \{y_1^*, y_2^*, \ldots, y_m^*\} \).

Based on \( \{x_1^*, x_2^*, \ldots, x_n^*\} \) and \( \{y_1^*, y_2^*, \ldots, y_m^*\} \), compute the bootstrap sample estimate of \( R \) using (11), say \( \hat{R}^* \).

Step 3: Repeat step 2, \( N \) boot times.

Step 4: Let \( G(x) = P(\hat{R}^* < x) \) be the cumulative distribution of \( \hat{R}^* \). Define \( \hat{R}_{boot} = G^{-1}(x) \) for a given \( x \). The approximate \( 100(1 - \gamma)\% \) confidence interval of \( R \) is given by

\[
\left( \hat{R}_{boot} \left( \frac{\gamma}{2} \right), \hat{R}_{boot} \left( \frac{1 - \gamma}{2} \right) \right).
\]

**Algorithm of the bootstrap-t method:** Step 1: From the sample \( \{x_1, x_2, \ldots, x_n\} \) and \( \{y_1, y_2, \ldots, y_m\} \), compute \( \hat{x}, \hat{\beta}, \hat{\lambda} \).
Step 2: Using \( \hat{\lambda} \) and \( \hat{\beta} \) generate a bootstrap sample \( \{x_1^*, x_2^*, \ldots, x_n^*\} \) and similarly using \( \hat{\beta} \) and \( \hat{\lambda} \) generate a sample \( \{y_1^*, y_2^*, \ldots, y_m^*\} \). Based on \( \{x_1^*, x_2^*, \ldots, x_n^*\} \) and \( \{y_1^*, y_2^*, \ldots, y_m^*\} \) compute the bootstrap sample estimate of \( R \) using (11), say \( R^* \) and following statistic

\[
T^* = \frac{\sqrt{n}(R^* - \hat{R})}{\sqrt{\text{Var}(R^*)}}
\]

where \( \text{Var}(R^*) \) is obtained using the observed or expected Fisher information matrix.

Step 3: Repeat step 2, \( N \) boot times.

Step 4: For the \( T^* \) values obtained in step 2, determine the upper and lower bounds of the 100(1 - \( \gamma \))% confidence interval of \( R \) as follows: let \( H(x) = P(T^* \leq x) \) be the cumulative distribution function of \( T^* \). For a given \( x \), define

\[
\hat{R}_{\text{boot-t}}(x) = \hat{R} + n^{-1/2} \sqrt{\text{Var}(\hat{R}) H^{-1}(x)}.
\]

Here also, \( \text{Var}(\hat{R}) \) can be computed as same as computing the \( \text{Var}(\hat{R}) \). The approximate 100(1 - \( \gamma \))% confidence interval of \( R \) is given by

\[
\left( \hat{R}_{\text{boot-l}} \left( \frac{\hat{y}}{2} \right), \hat{R}_{\text{boot-t}} \left( 1 - \frac{\hat{y}}{2} \right) \right).
\]

4. Estimation of \( R \) if \( \lambda \) is known

In this section, we consider the estimation of \( R \) when \( \lambda \) is known. Without loss of generality, we can assume that \( \lambda = 1 \).

4.1. MLE of \( R \)

Let \( X_1, X_2, \ldots, X_n \) be a random sample from \( \text{GP}(\alpha, 1) \) and \( Y_1, Y_2, \ldots, Y_m \) be a random sample from \( \text{GP}(\beta, 1) \). Based on Section 2, it is clear that the MLE of \( R \) will be

\[
\hat{R} = \frac{\hat{\beta}}{\hat{\alpha} + \hat{\beta}}.
\]

(14)

where

\[
\hat{\alpha} = \frac{1}{n} \sum_{i=1}^{n} \ln(1 + \lambda x_i),
\]

(15)

\[
\hat{\beta} = \frac{1}{m} \sum_{j=1}^{m} \ln(1 + \lambda y_j).
\]

(16)

Therefore,

\[
\hat{R} = R \frac{m \sum_{i=1}^{n} \ln(1 + \lambda x_i)}{n \sum_{j=1}^{m} \ln(1 + \lambda y_j) + m \sum_{i=1}^{n} \ln(1 + \lambda x_i)}.
\]

(17)

It is easy to see that \( \ln(1 + \lambda x_i) \) follows an exponential distribution with mean \( \alpha^{-1} \). Therefore, \( 2\alpha \sum_{i=1}^{n} \ln(1 + x_i) \sim \chi^2(2n) \) and \( 2\beta \sum_{j=1}^{m} \ln(1 + y_j) \sim \chi^2(2m) \). So,

\[
\hat{R} \sim \frac{1}{1 + \frac{x}{\beta}} F
\]

or

\[
\frac{R}{1-R} \sim \frac{1 - \hat{R}}{\hat{R}} F,
\]

where the random variable \( F \) follows a \( F_{2m,2n} \) distribution with \( 2m \) and \( 2n \) degrees of freedom. So, the probability density function (pdf) of \( R \) is as follows:

\[
f_R(x) = \frac{1}{x^d B(m,n)} \left( \frac{m\beta}{n\alpha} \right)^m \times \frac{(\frac{1-x}{x})^{m-1}}{\left(1 + \frac{m\beta}{n\alpha} \left(\frac{1-x}{x}\right)\right)^{n+m}},
\]

where \( 0 < x < 1 \) and \( \alpha, \beta > 0 \).
The 100(1 − γ)% confidence interval of $\hat{R}$ can be obtained as
\[
\left[ \frac{1}{1 + F_{(1/\gamma)/2;2n,2m}} \times (1/\hat{R} - 1), \frac{1}{1 + F_{(1/\gamma)/2;2n,2m}} \times (1/\hat{R} - 1) \right],
\]
where $F_{(1/\gamma)/2;2n,2m}$ and $F_{(1/\gamma)/2;2n,2m}$ are the lower and upper $\gamma/2$th percentile points of a $F$ distribution.

4.2. UMVUE of $R$

Using the results of Tong (1974, 1977), the UMVUE of $R$ is obtained. This method is also used by Kundu and Gupta (2005) and Kakade et al. (2008). When the common scale parameter $\lambda$ is known ($\sum_{i=1}^{n} \ln(1 + X_i), \sum_{i=1}^{m} \ln(1 + Y_i)$) is a jointly sufficient statistics for $(\alpha, \beta)$. Let us define
\[
\phi(X_i, Y_i) = \begin{cases} 
1 & \text{if } \ln(1 + X_i) < \ln(1 + Y_i), \\
0 & \text{otherwise}
\end{cases}
\]
Clearly, $1 - \phi(X_i, Y_i)$ is an unbiased estimator of $R$, since $E(1 - \phi(X_i, Y_i)) = P(Y < X) = R$. Therefore, the UMVUE of $R$, say $\hat{R}$, can be obtained as
\[
\hat{R} = E \left[ 1 - \phi(X_i, Y_i) \left( \sum_{i=1}^{n} \ln(1 + X_i), \sum_{i=1}^{m} \ln(1 + Y_i) \right) \right].
\]

Since $\ln(1 + X_i)$ and $\ln(1 + Y_i)$ are exponential random variables, so using the results of Tong (1974,1977) it follows that
\[
\hat{R} = \begin{cases} 
1 - \sum_{i=0}^{m-1} (-1)^i \frac{(n-1)!}{(m-i-1)!} \left( \frac{T_1}{T_2} \right)^i & \text{if } T_1 \leq T_2, \\
\sum_{i=0}^{n-1} (-1)^i \frac{(n-1)!}{(m+i-1)!} \left( \frac{T_1}{T_2} \right)^i & \text{if } T_2 \leq T_1,
\end{cases}
\]
where $T_1 = \sum_{i=1}^{n} \ln(1 + X_i)$ and $T_2 = \sum_{i=1}^{m} \ln(1 + Y_i)$.

4.3. Bayes estimation of $R$

In this section, we obtain the Bayes estimation of $R$ under assumption that the shape parameters $\alpha$ and $\beta$ are random variables. It is assumed that $\alpha$ and $\beta$ have independent gamma priors with the pdfs
\[
\pi(\alpha) = \frac{b_1^\alpha \Gamma(a_1)}{\Gamma(\alpha)} \alpha^{a_1-1} e^{-b_1 \alpha}, \quad \alpha > 0
\]
and
\[
\pi(\beta) = \frac{b_2^\beta \Gamma(a_2)}{\Gamma(\beta)} \beta^{a_2-1} e^{-b_2 \beta}, \quad \beta > 0,
\]
with the parameters $\alpha \sim \text{Gamma}(a_1, b_1)$ and $\beta \sim \text{Gamma}(a_2, b_2)$. The posterior pdfs of $\alpha$ and $\beta$ are as follows:
\[
\begin{align*}
\alpha|\text{data} & \sim \text{Gamma}(a_1 + n, b_1 + T_1), \quad \beta|\text{data} \sim \text{Gamma}(a_2 + m, b_2 + T_2),
\end{align*}
\]
where $T_1 = \sum_{i=1}^{n} \ln(1 + X_i)$ and $T_2 = \sum_{i=1}^{m} \ln(1 + Y_i)$.

Since a priori $\alpha$ and $\beta$ are independent, using (23) and (24), the posterior pdf of $R$ becomes
\[
f_R(r) = C \frac{r^{a_2 - 1}(1 - r)^{a_1 - 1} - 2r^2(B_2 - B_1) + r(2B_1 - 2B_2 - A_1B_2 - A_2B_1) + A_2B_1}{(B_1(1 - r) + B_2r)^{a_1 + a_2 + 3}}
\]
for $0 < r < 1$

and 0 otherwise, where
\[
C = \frac{\Gamma(n + m + a_1 + a_2)}{\Gamma(n + a_1)\Gamma(m + a_2)}(b_1 + T_1)^{a_1 + n}(b_2 + T_2)^{a_2 + m}.
\]

There is no explicit expression for the posterior mean or median. On the other hand the posterior mode can be easily obtained:
\[
\frac{d}{dr} f_R(r) = \frac{r^{a_2 - 1}(1 - r)^{a_1 - 1} - 2r^2(B_2 - B_1) + r(2B_1 - 2B_2 - A_1B_2 - A_2B_1) + A_2B_1}{(B_1(1 - r) + B_2r)^{a_1 + a_2 + 3}}.
\]
where $B_1 = b_1 + T_1$, $B_2 = b_2 + T_2$, $A_1 = a_1 + n - 1$ and $A_2 = a_2 + m - 1$. 
Note that, for \( r \in (0, 1) \), \((d/dr)f_R(r) = 0\) has only two roots. Using the fact that \( \lim_{r \to 0} (d/dr)f_R(r) > 0\) and \( \lim_{r \to 1} (d/dr)f_R(r) < 0\), it easily follows that the density function \( f_R(r) \) has a unique mode. The posterior mode can be obtained as the unique root of which lies in between 0 and 1 of the following quadratic:

\[
2r^2(B_1 - B_2) + r(2B_2 - 2B_1 + A_1B_2 + A_2B_1) - A_2B_1 = 0.
\]

Now, consider the following loss function:

\[
I(a, b) = \begin{cases} 
1 & \text{if } |a - b| > c, \\
0 & \text{if } |a - b| \leq c. 
\end{cases}
\] (26)

It is known that Bayes estimates with respect to the above equation is the midpoint of the `modal interval' of length 2c of the posterior distribution (see Ferguson, 1967). Therefore, the posterior mode is an approximate Bayes estimator of \( R \) under squared error loss function cannot be computed analytically. Alternatively, using the approximate method of Lindley (1980) and Ahmad et al. (1997), it can be easily seen that the approximate Bayes estimate of \( R \), say \( \tilde{R}_{\text{Bayes}} \), under squared error loss function is

\[
\tilde{R}_{\text{Bayes}} = \tilde{\beta}
\]

where

\[
\tilde{R} = \frac{\tilde{\beta}}{\hat{\alpha} + \hat{\beta}}, \quad \hat{\alpha} = \frac{n + a_1 - 1}{b_1 + t_1} \quad \text{and} \quad \hat{\beta} = \frac{m + a_2 - 1}{b_2 + t_2}.
\]

5. Estimation of \( R \) in general case

Computing the \( R \) when the scale parameter is different is considered, in this section. Surles and Padgett (1998, 2001) considered this case, also. In Surles and Padgett (2001) there is no exact expression for \( R \), but they presented a bound for it.

5.1. Maximum likelihood estimator of \( R \)

Let \( X \sim GP(\lambda, \lambda_1) \) and \( Y \sim GP(\beta, \lambda_2) \), where \( X \) and \( Y \) are independent random variables. Therefore,

\[
\begin{align*}
R &= \int_0^\infty P(Y < X|X = x)P(X = x) \, dx \\
&= \int_0^\infty \lambda_1 x^{\lambda_1 - 1}(1 + \lambda_1 x)^{-\lambda_1}(1 + \lambda_2 x)^{-\beta} \, dx \\
&= 1 - \int_0^\infty \lambda_1 x^{\lambda_1 - 1}(1 + \lambda_1 x)^{-\lambda_1}(1 + \lambda_2 x)^{-\beta} \, dx \\
&= 1 - x \left( \frac{\lambda_1}{\lambda_2} \right)^{-\beta} \int_0^\infty (1 + t)^{-\lambda_1} \left( \frac{\lambda_2}{\lambda_1} + t \right)^{-\beta} \, dt.
\end{align*}
\]

Considering integral of the form,

\[
\int_0^\infty x^{\lambda - 1}(\beta + x)^{-\mu}e^{-Q} \, dx = \Gamma(\nu + \mu - \nu - Q) F_2 \left( \mu, \nu + Q; 1 - \frac{\nu}{\beta} \right),
\]

where \( \nu > 0, \mu > -\nu \) and \( F_2(\cdot) \) is Gauss’ hypergeometric function \([\cdot]\). Then,

\[
R = 1 - \frac{x}{x + \beta} \left( \frac{\lambda_2}{\lambda_1} \right)^{-\beta} F_2 \left( \alpha + 1, 1; \alpha + \beta + 1; 1 - \frac{\lambda_2}{\lambda_1} \right).
\] (28)

To compute the MLE of \( R \), suppose \( X_1, X_2, \ldots, X_n \) is a random sample from \( GP(\lambda, \lambda_1) \) and \( Y_1, Y_2, \ldots, Y_m \) is a random sample from \( GP(\beta, \lambda_2) \). Therefore, the log-likelihood function of the observed sample is

\[
L(\lambda, \beta, \lambda_1, \lambda_2) = n \ln x + n \ln \lambda_1 - (\alpha + 1) \sum_{i=1}^n \ln(1 + \lambda_1 x_i) + m \ln \beta + m \ln \lambda_2 - (\beta + 1) \sum_{j=1}^m \ln(1 + \lambda_2 y_j).
\] (29)
The MLEs of $\alpha$, $\beta$, $\lambda_1$ and $\lambda_2$ say $\hat{\alpha}$, $\hat{\beta}$, $\hat{\lambda}_1$ and $\hat{\lambda}_2$, respectively, can be obtained as the solutions of

\[
\frac{\partial L}{\partial \alpha} = \frac{n}{\hat{\alpha}} - \sum_{i=1}^{n} \ln(1 + \hat{\lambda}_1 x_i), \\
\frac{\partial L}{\partial \beta} = \frac{m}{\hat{\beta}} - \sum_{j=1}^{m} \ln(1 + \hat{\lambda}_2 y_j), \\
\frac{\partial L}{\partial \hat{\lambda}_1} = \frac{n}{\hat{\lambda}_1} - (\hat{\alpha} + 1) \sum_{i=1}^{n} \frac{x_i}{1 + \hat{\lambda}_1 x_i}, \\
\frac{\partial L}{\partial \hat{\lambda}_2} = \frac{m}{\hat{\lambda}_2} - (\hat{\beta} + 1) \sum_{j=1}^{m} \frac{y_j}{1 + \hat{\lambda}_2 y_j}.
\]  

From the above equations, we obtain

\[
\hat{\alpha} = \frac{1}{n} \sum_{i=1}^{n} \ln(1 + \hat{\lambda}_1 x_i), \\
\hat{\beta} = \frac{1}{m} \sum_{j=1}^{m} \ln(1 + \hat{\lambda}_2 y_j)
\]

and $\hat{\lambda}_1$ and $\hat{\lambda}_2$ can be obtained as the solution of the non-linear equations

\[
f(\hat{\lambda}_1) = \frac{n}{\hat{\lambda}_1} - \frac{\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} 1 + \hat{\lambda}_1 x_i} - \frac{\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} 1 + \hat{\lambda}_1 x_i},
\]

and

\[
f(\hat{\lambda}_2) = \frac{m}{\hat{\lambda}_2} - \frac{\sum_{j=1}^{m} y_j}{\sum_{j=1}^{m} 1 + \hat{\lambda}_2 y_j} - \frac{\sum_{j=1}^{m} y_j}{\sum_{j=1}^{m} 1 + \hat{\lambda}_2 y_j}.
\]

By invariance property of the ML estimators, the MLE of $R$ becomes

\[
\hat{R} = 1 - \frac{\hat{\alpha}}{\hat{\alpha} + \hat{\beta}} \frac{\hat{\lambda}_2}{\hat{\lambda}_1} _2F_1 \left( \hat{\alpha} + 1, 1; \hat{\alpha} + \hat{\beta} + 1; 1 - \frac{\hat{\lambda}_2}{\hat{\lambda}_1} \right).
\]

5.2. Asymptotic distribution

The asymptotic distribution of $\hat{\theta} = (\hat{\alpha}, \hat{\beta}, \hat{\lambda}_1, \hat{\lambda}_2)$ is obtained and similar to Theorems 1 and 2, the asymptotic distribution of $\hat{R}$ could be obtained. We denote the expected Fisher information matrix of $\theta = (\alpha, \beta, \lambda_1, \lambda_2)$ as $I(\theta) = E(I; \theta)$, where $I = [I_{ij}]_{j=1,2,3,4}$ is the observed information matrix. I.e., $I_{ij} = -\partial^2 /\partial \theta_i \partial \theta_j$.

It is easy to see that

\[
I_{11} = \frac{n}{\hat{\alpha}^2}, \\
I_{12} = 0, \\
I_{13} = \sum_{i=1}^{n} \frac{x_i}{1 + \hat{\lambda}_1 x_i}, \\
I_{14} = 0, \\
I_{22} = \frac{m}{\hat{\beta}^2}, \\
I_{23} = 0, \\
I_{24} = \sum_{j=1}^{m} \frac{y_j}{1 + \hat{\lambda}_2 y_j}, \\
I_{33} = \frac{n}{\hat{\lambda}_1} - (\hat{\alpha} + 1) \sum_{i=1}^{n} \frac{x_i^2}{(1 + \hat{\lambda}_1 x_i)^2}, \\
I_{34} = 0, \\
I_{44} = \frac{m}{\hat{\lambda}_2} - (\hat{\beta} + 1) \sum_{j=1}^{m} \frac{y_j^2}{(1 + \hat{\lambda}_2 y_j)^2}.
\]
Therefore,

\[ I_{34} = 0, \]
\[ I_{44} = \frac{m}{s_2} - \frac{(\beta + 1)}{s_2} \sum_{j=1}^{m} \frac{y_j^2}{(1 + \lambda_j y_j)^\beta}. \]

Thus,

\[ J_{11} = \frac{n}{s_1^2}, \]
\[ J_{12} = 0, \]
\[ J_{13} = \frac{m}{s_1} \beta B(2, M), \]
\[ J_{14} = 0, \]
\[ J_{22} = m \beta, \]
\[ J_{23} = 0, \]
\[ J_{24} = \frac{m}{s_2} \beta B(2, \beta), \]
\[ J_{33} = \frac{n}{s_1} - \frac{m}{s_1} \beta B(3, M), \]
\[ J_{34} = 0, \]
\[ J_{44} = \frac{m}{s_2} - \frac{n}{s_2} \beta B(3, \beta). \]

Based on the above Fisher information matrix, it is possible to present confidence intervals of \( R \) based on the percentile bootstrap and bootstrap-t methods. But, since they are similar to those of mentioned in Section 3.3 and for saving space, we omit them.

5.3. Bayes estimation of \( R \)

In this section, we obtain the Bayes estimation of \( R \) under assumption that the shape parameters \( x \) and \( \beta \) and scale parameters \( \lambda_1 \) and \( \lambda_2 \) are random variables. We mainly obtain the Bayes estimate of \( R \) under the squared error loss, and the corresponding credible interval by the Gibbs sampling technique. It is assumed that \( x, \beta, \lambda_1 \) and \( \lambda_2 \) have independent gamma priors with the parameters \( x \sim \text{Gamma}(a_1, b_1), \beta \sim \text{Gamma}(a_2, b_2), \lambda_1 \sim \text{Gamma}(a_3, b_3) \) and \( \lambda_2 \sim \text{Gamma}(a_4, b_4) \). Based on the above assumptions, we have the likelihood function of the observed data as

\[ L(\text{data}; x, \beta, \lambda_1, \lambda_2) = x^n \lambda_1^n \prod_{i=1}^{n} (1 + \lambda_1 y_i)^{-x} \lambda_2^m \prod_{i=1}^{m} (1 + \lambda_2 y_i)^{-\beta}. \]

Therefore, the joint density of the data, \( x, \beta, \lambda_1 \) and \( \lambda_2 \) can be obtained as

\[ L(\text{data}; x, \beta, \lambda_1, \lambda_2) = L(\text{data}; x, \beta, \lambda_1, \lambda_2) \pi(\lambda_2) \pi(\lambda_1) \pi(\beta) \pi(x), \]

where \( \pi(\cdot) \) is the prior distribution. Therefore the joint posterior density of \( x, \beta, \lambda_1 \) and \( \lambda_2 \) given the data is

\[ L(x, \beta, \lambda_1, \lambda_2; \text{data}) = \int_{0}^{\infty} \int_{0}^{\infty} \int_{0}^{\infty} L(x, \beta, \lambda_1, \lambda_2; \text{data}) \pi(x) \pi(\lambda_1) \pi(\lambda_2) \pi(\beta) \pi(\lambda_2). \]

Since (40) cannot be obtained analytically, we adopt the Gibbs sampling technique to compute the Bayes estimate of \( R \), and the corresponding credible interval of \( R \).

The posterior pdfs of \( x \) and \( \beta \) are as follows:

\[ x, \beta, \lambda_1, \lambda_2, \text{data} \sim \text{Gamma} \left( a_1 + n, b_1 + \sum_{i=1}^{n} \ln(1 + \lambda_1 y_i) \right), \]

\[ \beta, \lambda_1, \lambda_2, \text{data} \sim \text{Gamma} \left( a_2 + m, b_2 + \sum_{i=1}^{m} \ln(1 + \lambda_2 y_i) \right). \]
and

\[
f_{\lambda_1}(\hat{\lambda}_1|\beta, \lambda_2, \text{data}) \propto \lambda_1^{n+a-1} e^{-b_1 \lambda_1} \prod_{i=1}^{n}(1 + \lambda_1 x_i)^{-1},
\]

(43)

\[
f_{\lambda_2}(\hat{\lambda}_2|\beta, \lambda_1, \text{data}) \propto \lambda_2^{m+a-1} e^{-b_2 \lambda_2} \prod_{i=1}^{m}(1 + \lambda_2 y_i)^{-1}.
\]

(44)

The posterior pdfs of \(\hat{\lambda}_1\) and \(\hat{\lambda}_2\) are not known, but the plots of them show that they are similar to normal distribution. So to generate random numbers from these distributions, we use the Metropolis–Hastings method with normal proposal distribution. Therefore the algorithm of Gibbs sampling is as follows:

- **Step 1:** Start with an initial guess \((x^{(0)}, \beta^{(0)}, \lambda_1^{(0)}, \lambda_2^{(0)})\).
- **Step 2:** Set \(t = 1\).
- **Step 3:** Generate \(x^{(t)}\) from Gamma\((a_1 + n, b_1 + \sum_{i=1}^{n} \ln(1 + \hat{\lambda}_1^{(t-1)} x_i))\).
- **Step 4:** Generate \(\beta^{(t)}\) from Gamma\((a_2 + m, b_2 + \sum_{i=1}^{m} \ln(1 + \hat{\lambda}_2^{(t-1)} y_i))\).
- **Step 5:** Using Metropolis–Hastings, generate \(\hat{\lambda}_1^{(t)}\) from \(f_{\lambda_1}\) with the \(N(\hat{\lambda}_1^{(t-1)}, 1)\) proposal distribution.
- **Step 6:** Using Metropolis–Hastings, generate \(\hat{\lambda}_2^{(t)}\) from \(f_{\lambda_2}\) with the \(N(\hat{\lambda}_2^{(t-1)}, 1)\) proposal distribution.
- **Step 7:** Compute \(R^{(t)}\) from (38).
- **Step 8:** Set \(t = t + 1\).
- **Step 9:** Repeat steps 3–8, \(T\) times.

Note that in steps 5 and 6, we use the Metropolis–Hastings algorithm with \(q \sim N(\hat{\lambda}^{(t-1)}, \sigma^2)\) proposal distribution as follows:

a. Let \(x = \hat{\lambda}^{(t-1)}\).

b. Generate \(y\) from the proposal distribution \(q\).

c. Let \(p(x, y) = \min\{1, f_{\lambda_1}(y)/f_{\lambda_1}(x)q(x)/q(y)\}\).

d. Accept \(x\) with probability \(p(x, y)\) or accept \(x\) with probability \(1 - p(x, y)\).

Now the approximate posterior mean, and posterior variance of \(R\) become

\[
\hat{E}(R|\text{data}) = \frac{1}{T} \sum_{t=1}^{T} R^{(t)}
\]

(45)

and

\[
\hat{V}(R|\text{data}) = \frac{1}{T} \sum_{t=1}^{T} (R^{(t)} - \hat{E}(R|\text{data}))^2,
\]

(46)

respectively. Based on \(T\), and \(R\) values, using the method proposed by Chen and Shao (1999), the approximate highest posterior density (HPD) credible interval of \(R\) can be easily constructed. This method is also used by Kundu and Gupta (2006), but for generating random numbers from an unknown distribution, they used the method proposed by Devroye (1984) to generate a sample from a log-concave density function.

6. Simulation results

In this section, we present some results based on Monte Carlo simulations to compare the performance of the different methods mainly for small sample sizes.

We consider three cases separately to draw inference on \(R\), namely when (i) the common scale parameter \(\lambda\) is unknown, (ii) the common scale parameter \(\lambda\) is known and (iii) the scale parameter \(\lambda_1\) and \(\lambda_2\) are unknown. In the first two cases, we consider the following small sample sizes: \(m, n = 15, 25\) and \(50\). We take different values for \(x\), \(\beta\) and \(\lambda\), also.

For the first case, we assumed that the common scale parameter \(\lambda\) is unknown. From the sample, we compute the estimate of \(\lambda\), using the iterative algorithm (10). We have used the initial estimate to be 1 and the iterative process stops when the difference between the two consecutive iterates are less than \(10^{-6}\). Once we estimate \(\lambda\), we estimate \(x\) and \(\beta\) using (6) and (7), respectively. Finally, we obtain the MLE of \(R\) using (11). We report the average biases and mean square error (MSE) of \(R\) over 1000 replications. We also compute the 95% confidence intervals based on the asymptotic distribution of \(R\) through the estimation of \(B\) using two different methods: using the expected information matrix (Theorem 2) and replacing the parameters by their estimates (denoted by \(C_{\lambda}\)), and estimating \(B\) from the observed information matrix (denoted by \(C_{\lambda}\)). We report the average confidence intervals and the coverage percentages, \(cp\), based on 1000 replications. All the results are reported in Table 1.

Some of the points are clear from this experiment. Even for small sample sizes, the performance of the MLEs are quite satisfactory in terms of biases and MSE\((R)\). It is observed that when \(m = n\) and \(n\) increase, then MSE\((R)\) and biases decrease. It
In this case as expected, for all the methods when it is observed that estimating verifies the consistency property of the MLE estimators of Table 2. On the other hand, the non-informative prior, i.e., MLE method. The ML estimation of Bayes estimates, $(25, 25)$ $(1,1,1)$ $(1.06,1.08,0.97)$ 0.0010 0.0056 $(0.362,0.639)$ $(0.362,0.639)$ 0.937,
$(15, 15)$ $(1.18,1.13,0.95)$ 0.0001 0.0091 $(0.321,0.679)$ $(0.321,0.679)$ 0.941;
$(2,1.5,0.5)$ $(2.70,1.58,0.43)$ 0.0008 0.0094 $(0.253,0.606)$ $(0.255,0.604)$ 0.930;
$(2,1.5,1.5)$ $(2.60,1.92,1.31)$ 0.0025 0.0089 $(0.250,0.603)$ $(0.251,0.601)$ 0.934;
$(0.5,0.5,1.5)$ $(0.55,0.54,1.02)$ 0.0016 0.0096 $(0.320,0.677)$ $(0.320,0.677)$ 0.934;
$(0.5,0.5,1.5,3)$ $(0.57,1.82,2.82)$ 0.0082 0.0064 $(0.612,0.904)$ $(0.634,0.882)$ 0.880.

Table 2
Simulation results and estimation of the parameters when $\lambda$ is known from 1000 samples.

<table>
<thead>
<tr>
<th>($n, m$)</th>
<th>$(\alpha, \beta, \lambda)$</th>
<th>$(\hat{\alpha}, \hat{\beta}, \hat{\lambda})$</th>
<th>Bias(R)</th>
<th>MSE(R)</th>
<th>$Cl_\alpha$</th>
<th>$Cl_\beta$</th>
<th>$cp$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(15, 15)</td>
<td>$(1.11,1.11,1.11)$</td>
<td>$(1.11,1.11,1.11)$</td>
<td>0.0001</td>
<td>0.0091</td>
<td>$(0.321,0.679)$</td>
<td>$(0.321,0.679)$</td>
<td>0.941</td>
</tr>
<tr>
<td>(50, 25)</td>
<td>$(1.11,1.10,0.97)$</td>
<td>$(1.10,1.10,0.97)$</td>
<td>0.0010</td>
<td>0.0056</td>
<td>$(0.362,0.639)$</td>
<td>$(0.362,0.639)$</td>
<td>0.942</td>
</tr>
</tbody>
</table>

Considering the case when the common scale parameter is known, we obtain the estimates of $\hat{R}$ by substituting the parameters by their estimates in Theorem 2 or using observed Fisher information matrix, leads to very similar estimates. The confidence intervals based on the MLEs work quite well unless the sample size is very small, say $(15, 15)$. It is observed that when the sample size is increase, then the coverage percentages of the confidence intervals based on the asymptotic results reach the nominal level, 95%.

To compute different Bayes estimates, we prefer to use the non-informative prior, because we do not have any prior information on $R$. On the other hand, the non-informative prior, i.e., $\alpha_1 = \alpha_2 = \beta_1 = \beta_2 = 0$ provides prior distributions which are not proper, we adopt the suggestion of Congdon (2001, p. 20) and Kundu and Gupta (2005), i.e., choose $\alpha_1 = \alpha_2 = \beta_1 = \beta_2 = 0.0001$. The average estimation of Bayes estimates, $\hat{R}_{\text{Bayes}}$, is presented in Table 2, under the same prior distributions and based on 1000 replications. In this case as expected, for all the methods when $m$ and $n$ increase, then the average biases decrease.

In the third case when the scale parameters $\lambda_1$ and $\lambda_2$ are different and unknown, we obtain the estimates of $\hat{R}$ using the MLE method. The ML estimation of $\hat{R}$, $\hat{R}$ are reported in Table 3. Since for the MLE, the exact distribution is also known therefore it can be used to construct confidence intervals, also. We denote this confidence interval by $CI_{\text{MLE}}$. To compute different Bayes estimates, we prefer to use the non-informative prior, because we do not have any prior information on $R$. On the other hand, the non-informative prior, i.e., $\alpha_1 = \alpha_2 = \beta_1 = \beta_2 = 0$ provides prior distributions which are not proper, we adopt the suggestion of Congdon (2001, p. 20) and Kundu and Gupta (2005), i.e., choose $\alpha_1 = \alpha_2 = \beta_1 = \beta_2 = 0.0001$. The average estimation of Bayes estimates, $\hat{R}_{\text{Bayes}}$, is presented in Table 2, under the same prior distributions and based on 1000 replications. In this case as expected, for all the methods when $m$ and $n$ increase, then the average biases decrease.

In the third case when the scale parameters $\lambda_1$ and $\lambda_2$ are different and unknown, we obtain the estimates of $\hat{R}$ using the MLE method. The ML estimation of $\hat{R}$, $\hat{R}$ are reported in Table 3. To compute different Bayes estimates, we prefer to use the
Table 3
Simulation results and estimation of the parameters when \( \lambda_1 \) and \( \lambda_2 \) are different from 1000 samples. \( \langle \lambda_1 = 1 \rangle \).

<table>
<thead>
<tr>
<th>(n, m)</th>
<th>( \beta )</th>
<th>( \lambda_2 )</th>
<th>( R )</th>
<th>( \tilde{R} )</th>
<th>( \tilde{E}(R) )</th>
<th>( \tilde{V}(R) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>(25,25)</td>
<td>1</td>
<td>1.5</td>
<td>0.5</td>
<td>0.7124</td>
<td>0.7082</td>
<td>0.7119</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>1.5</td>
<td>2</td>
<td>0.4787</td>
<td>0.4787</td>
<td>0.4784</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>1.5</td>
<td>3</td>
<td>0.4086</td>
<td>0.4087</td>
<td>0.4091</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1.5</td>
<td>0.5</td>
<td>0.5619</td>
<td>0.5599</td>
<td>0.5607</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1.5</td>
<td>2</td>
<td>0.3032</td>
<td>0.3000</td>
<td>0.3034</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1.5</td>
<td>3</td>
<td>0.2393</td>
<td>0.2356</td>
<td>0.2393</td>
</tr>
<tr>
<td>(25,50)</td>
<td>1</td>
<td>1.5</td>
<td>0.5</td>
<td>0.7124</td>
<td>0.7119</td>
<td>0.7116</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>1.5</td>
<td>2</td>
<td>0.4787</td>
<td>0.4805</td>
<td>0.4798</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>1.5</td>
<td>3</td>
<td>0.4086</td>
<td>0.4076</td>
<td>0.4096</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1.5</td>
<td>0.5</td>
<td>0.5619</td>
<td>0.5603</td>
<td>0.5622</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1.5</td>
<td>2</td>
<td>0.3032</td>
<td>0.3011</td>
<td>0.3029</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1.5</td>
<td>3</td>
<td>0.2393</td>
<td>0.2363</td>
<td>0.2398</td>
</tr>
<tr>
<td>(50,50)</td>
<td>1</td>
<td>1.5</td>
<td>0.5</td>
<td>0.7124</td>
<td>0.7129</td>
<td>0.7122</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>1.5</td>
<td>2</td>
<td>0.4787</td>
<td>0.4780</td>
<td>0.4784</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>1.5</td>
<td>3</td>
<td>0.4086</td>
<td>0.4089</td>
<td>0.4083</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1.5</td>
<td>0.5</td>
<td>0.5619</td>
<td>0.5614</td>
<td>0.5613</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1.5</td>
<td>2</td>
<td>0.3032</td>
<td>0.3036</td>
<td>0.3030</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1.5</td>
<td>3</td>
<td>0.2393</td>
<td>0.2440</td>
<td>0.2390</td>
</tr>
</tbody>
</table>

Table 4
The data generated with \( m = n = 20 \), \( \beta = 1.4 \) and \( \lambda = 1 \).

<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
<th>xy</th>
<th>x</th>
<th>y</th>
<th>xy</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0024</td>
<td>0.0359</td>
<td>4.9743</td>
<td>5.0290</td>
<td>0.1250</td>
<td>0.5266</td>
</tr>
<tr>
<td>1.0863</td>
<td>0.7108</td>
<td>0.0648</td>
<td>1.4026</td>
<td>0.0185</td>
<td>0.7632</td>
</tr>
<tr>
<td>19.3212</td>
<td>2.0499</td>
<td>1.7140</td>
<td>0.2223</td>
<td>10.4438</td>
<td>0.7383</td>
</tr>
<tr>
<td>6.5356</td>
<td>4.7506</td>
<td>4.9126</td>
<td>0.3367</td>
<td>1.3550</td>
<td>0.0296</td>
</tr>
<tr>
<td>6.1724</td>
<td>2.4195</td>
<td>1.8520</td>
<td>0.0538</td>
<td>0.0088</td>
<td>0.0206</td>
</tr>
</tbody>
</table>

Table 5
Parameters estimation and bootstrap confidence intervals with \( N = 500 \) boot times for the data presented in Table 3.

<table>
<thead>
<tr>
<th>( \lambda )</th>
<th>( \langle \tilde{z}, \tilde{\beta}, \tilde{\lambda} \rangle )</th>
<th>( \tilde{R} )</th>
<th>( \tilde{R}^c )</th>
<th>( \tilde{B}_e )</th>
<th>( \tilde{B}_o )</th>
<th>( CI_{boot-p} )</th>
<th>( CI_{boot-t} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unknown</td>
<td>(1.13,1.31,1.02)</td>
<td>0.5346</td>
<td>0.5379</td>
<td>0.1242</td>
<td>0.1243</td>
<td>(0.368,0.718)</td>
<td>(0.349,0.686)</td>
</tr>
<tr>
<td>Known</td>
<td>(1.15,1.32,1.00)</td>
<td>0.5353</td>
<td>0.5361</td>
<td>0.1245</td>
<td>0.1247</td>
<td>(0.380,0.691)</td>
<td>(0.359,0.702)</td>
</tr>
</tbody>
</table>

non-informative prior, because we do not have any prior information on \( R \). So, we adopt the suggestion of Congdon (2001, p. 20) and Kundu and Gupta (2005), i.e., choose \( a_1 = a_2 = b_1 = b_2 = 0.0001 \). The average and variance estimations of Bayes estimates, \( \tilde{E}(R) \) and \( \tilde{V}(R) \), are presented in Table 3, under the same prior distributions and based on 1000 replications. In this case as expected, for all the methods when \( m \) and \( n \) increase, then the average biases decrease.

7. Numerical example

Since the confidence intervals based on the asymptotic results for small sample sizes do not perform very well results, so we present an analysis based on two bootstrap methods. To do this, we simulate 20 numbers from \( GE(1, 1) \) and 20 numbers from \( GE(1.4, 1) \), reported in Table 4. A complete analysis of these data are presented in this section. In the first row of Table 5, it is assumed that \( \lambda \) is unknown. The MLE of \( \langle z, \beta, \lambda \rangle \) is obtained using (6)–(8). The iterative procedure is stops whenever two consecutive values are less than \( 10^{-6} \). Using the percentile bootstrap method, we present the mean of 500 bootstrap samples of \( R \) by \( \tilde{R} \) and its 95% confidence interval by \( CI_{boot-p} \). We also present the confidence interval of bootstrap-t method. To compute \( CI_{boot-t} \), the \( \tilde{B}_e \) and \( \tilde{B}_o \) are computed using the expected and observed information matrix, respectively.

In the second row of Table 5, it is assumed that \( \lambda \) is known and is equal to 1. The parameters are estimated using (15)–(17). As same as above, the bootstrap confidence intervals are obtained. The performance of the bootstrap confidence intervals are quite well. But construction of bootstrap confidence intervals are computationally more demanding than the asymptotic confidence intervals. Using (20) the UMVUE of \( R \) becomes \( \tilde{R} = 0.5362 \) and using (27) the Bayes estimation becomes \( \tilde{R}_{\text{Bayes}} = 0.5344 \). The profile likelihood function of \( \lambda \) is plotted in Fig. 1. It is an upside down function and has a unique maximum. The posterior probability density function of \( R \) for the given data set, is plotted in Fig. 2. It is almost symmetric, as expected.
8. Conclusion

In this paper, we have addressed the problem of estimating \( P(Y < X) \) for the generalized Pareto distributions. We consider the cases when the common scale parameter is known or unknown and when the scale parameters are different. When the common scale parameter is unknown, it is observed that the maximum likelihood estimator works quite well. We can use the asymptotic distribution of the maximum likelihood estimator to construct confidence intervals which work well even for small sample sizes. Based on the simulation results, we recommend to use the parametric Bootstrap percentile method, when the sample size is very small.

When the common scale parameter is known we propose maximum likelihood estimator and uniformly minimum variance unbiased estimator. The data analysis indicates that the confidence interval based on the exact distribution of the MLE and the corresponding credible intervals based on the non-informative priors are almost identical.

References


