MJFAS
Malaysian Journal of Fundamental and Applied Sciences

RESEARCH ARTICLE

Confidence Intervals for Ratio of Percentiles of Birnbaum-Saunders Distributions and Its Application to PM2.5 in Northern Thailand

Warisa Thangjai[a], Sa-Aat Niwitpong[b]*, Suparat Niwitpong[b]

[a]Department of Statistics, Faculty of Science, Ramkhamhaeng University, Bangkok, 10240, Thailand; [b]Department of Applied Statistics, Faculty of Applied Science, King Mongkut's University of Technology North Bangkok, Bangkok, 10800, Thailand

Abstract In the statistical literature, the Birnbaum-Saunders distribution has garnered considerable attention as an important fatigue life distribution for engineers. This distribution finds applications in medicine, engineering, and science. A percentile denotes a threshold below which a designated percentage of scores or data points are situated. In particular cases, percentiles are the preferred choice over mean and variance, especially when dealing with skewed data. This paper investigates confidence intervals for the ratio of percentiles of Birnbaum-Saunders distributions. It discusses the generalized fiducial confidence interval, Bayesian credible interval, and the highest posterior density intervals using a prior distribution with partial information and a proper prior with known hyperparameters. To compare the performance of these confidence intervals, the study employs coverage probability and average length measurements. Monte Carlo simulation via the R package is used to calculate the coverage probability and average length. The results indicate that the highest posterior density interval using a prior distribution with partial information outperforms the other confidence intervals. Finally, the paper presents the results of the simulation study and applies them in the field of environmental sciences.

Keywords: Bayesian credible interval, Birnbaum-Saunders distribution, generalized fiducial confidence interval, highest posterior density interval, percentile.

Introduction

The well-known Weibull distribution is a commonly utilized model for lifetimes. Nonetheless, the Weibull distribution may not accurately represent lifetime data in certain cases. For instance, Durham and Padgett [8] suggested that the Birnbaum-Saunders (BS) distribution provides a much better fit for some carbon fiber or carbon composite tensile strength data compared to the Weibull distribution. Furthermore, the BS distribution can serve as an approximation of the inverse Gaussian distribution [2] and can be seen as an equal mixture of the inverse Gaussian distribution [7]. Additionally, the properties of the BS distribution bear resemblance to those of the log-normal distribution [21]. Birnbaum and Saunders [3] introduced the BS distribution as a model for failure times resulting from cyclic loading-induced fatigue failure. The BS distribution is widely adopted as a lifetime distribution in various reliability theory models. Therefore, statistical inference for the BS distribution holds practical value in reliability applications. Numerous researchers have developed and discussed methods for inference based on the BS distribution. For example, Birnbaum and Saunders [4] presented the maximum likelihood estimator for its parameters, while Ng *et al*. [17] introduced the modified moment estimator. Wu and Wong [25] provided approximated interval estimation for the BS distribution based on higher-order likelihood asymptotic procedures. Li and Xu [15] presented the fiducial estimator for the distribution's parameters and compared it with maximum likelihood and Bayesian estimators. Guo *et al*. [11] introduced methods for both interval estimation and hypothesis testing regarding the common mean of BS distributions. Their approach blends aspects of generalized inference with principles from large sample theory. Jayalath [13] used a flexible Gibbs sampler for parameter inference for the BS distribution. Puggard *et al*. [20] introduced confidence intervals (CIs) for the variance and the difference of variances for BS distributions. The BS distribution, also referred to as the fatigue life distribution, is a widely employed probability

***For correspondence:**
sa-aat.n@sci.kmutnb.ac.th

Received: 18 Dec. 2023
Accepted: 6 May 2024

distribution in the fields of reliability engineering and statistics. It is primarily utilized for modeling the lifespan or fatigue life of materials or products exposed to cyclical or repetitive stress. This distribution proves instrumental in gaining insights into the degradation or wear and tear of products and materials over time. This distribution provides a robust framework for the modeling and analysis of data pertaining to fatigue life and failure rates, serving as a valuable tool for engineers and researchers to make well-informed decisions concerning the reliability and durability of products and materials. Widely embraced in the realm of reliability analysis, particularly within industries like engineering, material science, and product design, the BS distribution is employed to discern the likely points in time when products or materials may fail. It considers factors such as cumulative damage, wear and tear, and fatigue in its assessments. Furthermore, this distribution plays a pivotal role in risk assessment by aiding in the evaluation of the likelihood of rare or extreme events associated with failures. Its utility in risk management and the analysis of product reliability is particularly noteworthy.

While the mean and variance are fundamental statistical concepts, percentiles provide unique advantages in certain scenarios, especially when data deviate from normal distributions, incorporate outliers, or when assessing risk and making decisions relies on evaluating event probabilities concerning specific benchmarks. The selection of the most suitable measure depends on both the characteristics of the data and the specific goals of the analysis. Percentiles are frequently used in various fields such as medicine, engineering, and environmental sciences. In medicine, comparing a new drug to a standard one may involve examining the percentile response of a majority of patients, which can be more significant than the average response. In engineering, the percentile of a lifetime distribution is commonly employed for maintenance or structural design.

There has been extensive research into the parameters of the BS distribution. Several researchers have explored inferences based on these parameters, including Achcar [1], Lu and Chang [16], Ng *et al*. [17], Wu and Wong [25], Leiva *et al*. [14], Wang [23], Niu *et al*. [18], Wang *et al*. [24], and Guo *et al*. [11]. This paper's objective is to provide CIs for the ratio of two percentiles of the BS distributions. It employs four different approaches: the generalized fiducial confidence interval (GFCI) approach, the Bayesian approach, and the highest posterior density (HPD) approaches using a prior distribution with partial information (HPD-PI) and a proper prior with known hyperparameters (HPD-KH) to estimate interval estimates for the population ratio of percentiles. These approaches rely on simulated data for constructing the CIs. To facilitate practical usage, a computer program is developed in the R language for calculating coverage probability (CP) and average length (AL). A numerical example is provided to demonstrate the application of this program.

Materials and Methods

Let X_{ij} be non-negative random variable drawn from a BS distribution with shape parameter α_i and scale parameter β_i, where $i = 1,2$ and $j = 1,2,...,n_i$. The probability density function is

$$f(x_{ij};\alpha_i,\beta_i) = \frac{1}{2\alpha_i\beta_i\sqrt{2\pi}}\left[\left(\frac{\beta_i}{x_{ij}}\right)^{\frac{1}{2}} + \left(\frac{\beta_i}{x_{ij}}\right)^{\frac{3}{2}}\right]\exp\left[-\frac{1}{2(\alpha_i)^2}\left(\frac{x_{ij}}{\beta_i} + \frac{\beta_i}{x_{ij}} - 2\right)\right], \quad (1)$$

where $x_{ij} > 0$, $\alpha_i > 0$, and $\beta_i > 0$.

Based on the research conducted by Puggard *et al*. [20], they employed this transformation to produce sample data from the BS distribution, streamlining the extraction of several of its supplementary properties, including various statistical moments.

The normal transformation was utilized to generate samples from the BS distributions and to facilitate the derivation of several other properties, such as various moments. Therefore, suppose that X_{ij} is a random variable following the BS distribution with parameters α_i and β_i, and it is

$$X_{ij} = \beta_i\left(1 + 2(Z_{ij})^2 + 2Z_{ij}\sqrt{1 + (Z_{ij})^2}\right), \quad (2)$$

where

$$Z_{ij} = \frac{1}{2}\left(\sqrt{\frac{X_{ij}}{\beta_i}} - \sqrt{\frac{\beta_i}{X_{ij}}}\right) \sim N\left(0, \frac{(\alpha_i)^2}{4}\right). \quad (3)$$

Following Chang and Tang [6] and Padgett and Tomlinson [19], the p -th percentile of X_{ij} is

$$\theta_i = \frac{\beta_i}{4}\left[\alpha_i z_p + \sqrt{(\alpha_i)^2 z_p^2 + 4}\right]^2 , \tag{4}$$

where $z_p = \Phi^{-1}(p)$ is the p -th quantile of the standard normal distribution.

Therefore, the ratio of percentiles is

$$\eta = \frac{\theta_1}{\theta_2} , \tag{5}$$

where

$$\theta_1 = \frac{\beta_1}{4}\left[\alpha_1 z_p + \sqrt{(\alpha_1)^2 z_p^2 + 4}\right]^2 \tag{6}$$

and

$$\theta_2 = \frac{\beta_2}{4}\left[\alpha_2 z_p + \sqrt{(\alpha_2)^2 z_p^2 + 4}\right]^2 . \tag{7}$$

Generalized Fiducial Confidence Interval (GFCI) Approach

Generalized fiducial inference offers a means to convert the initial dataset into different distributions with established properties. Following the guidelines of these alternative distributions, adjustments are made to the transformed data, and the outcomes are then reconverted to their original form using an inverse transformation, as described by Hannig [12].

Suppose that $\propto$ means "is proportional to". Li and Xu [15] derived the generalized fiducial distribution, which is

$$f(\alpha_i, \beta_i \mid x_{ij}) \propto J(x_{ij}, (\alpha_i, \beta_i)) L(x_{ij} \mid \alpha_i, \beta_i) , \tag{8}$$

where

$$L(x_{ij} \mid \alpha_i, \beta_i) \propto \frac{1}{(\alpha_i)^{n_i} (\beta_i)^{n_i}} \prod_{j=1}^{n_i} \left[\left(\frac{\beta_i}{x_{ij}}\right)^{\frac{1}{2}} + \left(\frac{\beta_i}{x_{ij}}\right)^{\frac{3}{2}}\right] \exp\left[-\sum_{j=1}^{n_i} \frac{1}{2(\alpha_i)^2}\left(\frac{x_{ij}}{\beta_i} + \frac{\beta_i}{x_{ij}} - 2\right)\right] \tag{9}$$

and

$$J(x_{ij}, (\alpha_i, \beta_i)) = \sum_{1 \le j < k \le n_i} \frac{4\left|x_{ij} - x_{ik}\right|}{\alpha_i \left(1 + \frac{\beta_i}{x_{ij}}\right)\left(1 + \frac{\beta_i}{x_{ik}}\right)} . \tag{10}$$

According to Li and Xu [15] and Puggard *et al*. [20], the priors of α_i and β_i are the special case of the prior with partial information which are

$$\pi(\alpha_i) \propto \frac{1}{\alpha_i} \tag{11}$$

and

$$\pi(\beta_i) \propto \sum_{1 \le j < k \le n_i} \frac{\left|x_{ij} - x_{ik}\right|}{(1 + \beta_i / x_{ij})(1 + \beta_i / x_{ik})} . \tag{12}$$

Hence, $f(\alpha_i, \beta_i \mid x_{ij})$ is suitable for the specific scenario involving a prior informed by the priors of α_i in Equation (11) and β_i in Equation (12), both representing partial information. Let $\hat{\alpha}_i$ and $\hat{\beta}_i$ be the generalized fiducial samples of α_i and β_i, respectively. Because $\hat{\alpha}_i$ and $\hat{\beta}_i$ can be derived from the generalized fiducial distribution using a procedure similar to that of the Bayesian posterior, the process employed to generate fiducial samples from the generalized fiducial distribution specified in Equation (8) involved the utilization of adaptive rejection Metropolis sampling (ARMS), an extension of adaptive rejection sampling (ARS). Gilks and Wild [9] introduced the concept of ARS, which was primarily designed for handling target densities that exhibit log-concavity. However, recognizing the limitations of ARS, Gilks *et al*. [10] enhanced the technique to make it more versatile. Their modifications allowed ARS to accommodate multivariate distributions and non-log-concave density functions by permitting the proposal distribution to be lower than the target in specific regions. They also introduced a Metropolis-Hastings step to ensure that the accepted samples conform to the desired distribution. This improved technique was named ARMS. It is easy to implement ARMS using the 'arms' function within the R software suite's 'dlm' package. Please be aware that $\hat{\alpha}_i$ and $\hat{\beta}_i$ are random variables. Hence, α_i is substituted by $\hat{\alpha}_i$ and β_i is substituted by $\hat{\beta}_i$.

Therefore, the generalized fiducial estimates of η is

$$\hat{\eta} = \frac{\hat{\theta}_1}{\hat{\theta}_2}, \tag{13}$$

where

$$\hat{\theta}_1 = \frac{\hat{\beta}_1}{4}\left[\hat{\alpha}_1 z_p + \sqrt{(\hat{\alpha}_1)^2 z_p^2 + 4}\right]^2 \tag{14}$$

and

$$\hat{\theta}_2 = \frac{\hat{\beta}_2}{4}\left[\hat{\alpha}_2 z_p + \sqrt{(\hat{\alpha}_2)^2 z_p^2 + 4}\right]^2. \tag{15}$$

Therefore, the $100(1-\gamma)\%$ two-sided CI based on the GFCI approach is

$$CI_{GFCI} = [L_{GFCI}, U_{GFCI}] = [\hat{\eta}(\gamma/2), \hat{\eta}(1-\gamma/2)], \tag{16}$$

where $\hat{\eta}(\gamma/2)$ and $\hat{\eta}(1-\gamma/2)$ are the $100(\gamma/2)$-th and $100(1-\gamma/2)$-th percentiles of $\hat{\eta}$, respectively, and $\hat{\eta}$ is defined in Equation (13).

Algorithm 1: CI based on the GFCI approach

Step 1: Generate x_{ij} from BS distribution, where $i = 1, 2$ and $j = 1, 2, ..., n_i$

Step 2: Generate K samples of α_i and β_i using the 'arms' function

Step 3: Burn-in B samples

Step 4: Reduce the number of samples by using sampling lag $L > 1$ and the final number of samples is $K' = (K - B)/L$

Step 5: Calculate $\hat{\eta}$ using Equation (13) and obtain $\hat{\eta}_{(1)}, \hat{\eta}_{(2)}, ..., \hat{\eta}_{(K')}$

Step 6: Calculate L_{GFCI} and U_{GFCI}

Bayesian Approach

The Bayesian approach is commonly expressed using Bayes' theorem, a mathematical representation that elucidates the conversion of prior probability into posterior probability, accounting for the data's likelihood.

Xu and Tang [26] pointed out that utilizing the reference prior for the BS distribution leads to an improper posterior distribution. Wang *et al*. [24] subsequently confirmed this impropriety. To address this issue, proper priors with known hyperparameters are established by assuming that β_i follows an inverse-gamma distribution with parameters a_i and b_i, denoted as $IG(\beta_i | a_i, b_i)$, and that $\lambda_i = (\alpha_i)^2$ follows an inverse-gamma distribution with parameters c_i and d_i, denoted as $IG((\alpha_i)^2 | c_i, d_i)$.

The joint posterior density function of (α_i, β_i) is

$$\begin{aligned} p(\lambda_i, \beta_i | x_{ij}) &\propto L(x_{ij} | \alpha_i, \beta_i)\pi(\beta_i | a_i, b_i)\pi(\lambda_i | c_i, d_i) \\ &\propto \frac{1}{(\lambda_i)^{\frac{n_i}{2}} \beta_i^{n_i}} \prod_{j=1}^{n_i}\left[\left(\frac{\beta_i}{x_{ij}}\right)^{\frac{1}{2}} + \left(\frac{\beta_i}{x_{ij}}\right)^{\frac{3}{2}}\right] \exp\left[-\sum_{j=1}^{n_i}\frac{1}{2\lambda_i}\left(\frac{x_{ij}}{\beta_i} + \frac{\beta_i}{x_{ij}} - 2\right)\right] \\ &\times \beta_i^{-a_i-1}\exp\left(-\frac{b_i}{\beta_i}\right)(\lambda_i)^{-c_i-1}\exp\left(-\frac{d_i}{\lambda_i}\right). \end{aligned} \tag{17}$$

The marginal distribution of β_i is

$$\pi(\beta_i \mid x_{ij}) \propto \beta_i^{-(n_i+a_i+1)} \exp\left(-\frac{b_i}{\beta_i}\right) \prod_{j=1}^{n_i} \left[\left(\frac{\beta_i}{x_{ij}}\right)^{\frac{1}{2}} + \left(\frac{\beta_i}{x_{ij}}\right)^{\frac{3}{2}}\right] \left[\sum_{j=1}^{n_i} \frac{1}{2}\left(\frac{x_{ij}}{\beta_i} + \frac{\beta_i}{x_{ij}} - 2\right) + d_i\right]^{\frac{-(n_i+1)}{2-c_i}}. \quad (18)$$

The conditional posterior distribution of λ_i given β_i is

$$\pi(\lambda_i \mid x_{ij}, \beta_i) \propto IG\left(\frac{n_i}{2} + c_i, \frac{1}{2} \sum_{j=1}^{n_i} \left(\frac{x_{ij}}{\beta_i} + \frac{\beta_i}{x_{ij}} - 2\right) + d_i\right). \quad (19)$$

Posterior samples are generated using the Markov Chain Monte Carlo technique. The marginal distribution of β_i, as defined in Equation (18), cannot be easily expressed, making it impossible to use standard methods to generate posterior samples for β_i. In such cases, three common approaches come into play: the random-walk Metropolis procedure, the Metropolis-Hastings algorithm, and the slice sampler. These methods introduce auxiliary variables to simplify the sampling problem when dealing with the marginal posterior distribution defined in Equation (18). However, all three approaches are prone to producing serially correlated draws, often necessitating a very large sample size to obtain accurate estimates of the desired attributes of the posterior distribution. To overcome these potential challenges in generating posterior samples, the generalized ratio-of-uniforms method, as outlined by Wakefield *et al*. [22], is employed to generate posterior samples for β_i, which is defined as $\tilde{\beta}_i$. The generalized ratio-of-uniforms method operates based on the following concept.

A pair of random variables (u_i, v_i) is uniformly distributed inside region which is

$$A(r_i) = \left\{(u_i, v_i) : 0 < u_i \le \left[\pi\left(\frac{v_i}{u_i^{r_i}} \middle| x_{ij}\right)\right]^{1/(r_i+1)}\right\}, \quad (20)$$

where $r_i \ge 0$ is a constant term and $\pi(\cdot \mid x_{ij})$ is the marginal posterior distribution in Equation (18). Therefore, the probability density function of $\beta_i = \frac{v_i}{u_i^{r_i}}$ becomes $\frac{\pi(\beta_i \mid x_{ij})}{\int \pi(\beta_i \mid x_{ij}) d\beta_i}$. To generate random samples that are uniformly distributed in $A(r_i)$, the accept-reject method is applied from a suitable enveloping region. Following Wakefield *et al*. [22], the minimal bounding rectangle for $A(r_i)$ is

$$[0, a(r_i)] \times [b^-(r_i), b^+(r_i)], \quad (21)$$

where

$$a(r_i) = \sup_{\beta_i > 0} \left\{\left[\pi(\beta_i \mid x_{ij})\right]^{1/(r_i+1)}\right\}, \quad (22)$$

$$b^-(r_i) = \inf_{\beta_i > 0} \left\{\beta_i \left[\pi(\beta_i \mid x_{ij})\right]^{r_i/(r_i+1)}\right\}, \quad (23)$$

and

$$b^+(r_i) = \sup_{\beta_i > 0} \left\{\beta_i \left[\pi\left(\beta_i \mid x_{ij}\right)\right]^{r_i/(r_i+1)}\right\}. \quad (24)$$

Wang *et al*. [24] proposed that $b^-(r_i) = 0$, $a(r_i)$ is finite, and $b^+(r_i)$ is finite when choosing an appropriate value for r_i. The generalized ratio-of-uniforms method comprises the following three sequential stages:

1. Compute $a(r_i)$ and $b^+(r_i)$
2. Compute u_i from $U(0, a(r_i))$, compute v_i from $U(0, b^+(r_i))$, and compute $\rho = v_i / u_i^{r_i}$
3. If $u_i \le \left[\pi(\rho_i \mid x_{ij})\right]^{1/(r_i+1)}$ is true, assign $\tilde{\beta} = \rho_i$; otherwise, repeat the process

In the meantime, $\tilde{\lambda}_i$, representing the posterior samples of λ_i, can be obtained by applying the conditional posterior distribution described in Equation (19) using the LearnBayes package in the R software suite. Let $\tilde{\alpha}_i$ be the posterior samples of α_i which is computed by the square roots of $\tilde{\lambda}_i$. Let $\tilde{\alpha}_i$ and $\tilde{\beta}_i$ be random variables. Consequently, the Bayesian estimate for η is

$$\tilde{\eta} = \frac{\tilde{\theta}_1}{\tilde{\theta}_2}, \tag{25}$$

where

$$\tilde{\theta}_1 = \frac{\tilde{\beta}_1}{4}\left[\tilde{\alpha}_1 z_p + \sqrt{(\tilde{\alpha}_1)^2 z_p^2 + 4}\right]^2 \tag{26}$$

and

$$\tilde{\theta}_2 = \frac{\tilde{\beta}_2}{4}\left[\tilde{\alpha}_2 z_p + \sqrt{(\tilde{\alpha}_2)^2 z_p^2 + 4}\right]^2. \tag{27}$$

Therefore, the $100(1-\gamma)\%$ two-sided credible interval based on the Bayesian approach is

$$CI_{Baye} = [L_{Baye}, U_{Baye}] = [\tilde{\eta}(\gamma/2), \tilde{\eta}(1-\gamma/2)], \tag{28}$$

where $\tilde{\eta}(\gamma/2)$ and $\tilde{\eta}(1-\gamma/2)$ are the $100(\gamma/2)$-th and $100(1-\gamma/2)$-th percentiles of $\tilde{\eta}$, respectively.

Algorithm 2: Credible interval based on Bayesian approach

Step 1: Set a_i, b_i, c_i, and d_i, where $i = 1,2$

Step 2: Compute $a(r_i)$ and $b^+(r_i)$

Step 3: Repeat m times

(a) Generate u_i from $U(0, a(r_i))$ and generate v_i from $U(0, b^+(r_i))$, and then compute $\rho_i = v_i / u_i^{r_i}$

(b) If $u_i \le \left[\pi(\rho_i \mid x_{ij})\right]^{1/(r_i+1)}$, set $\tilde{\beta}_{i,(m)} = \rho_i$; otherwise, repeat step (a)

(c) Generate $\tilde{\lambda}_{i,(m)}$ from $IG\left(\frac{n_i}{2} + c_i, \sum_{j=1}^{n_i}\left(\frac{x_{ij}}{\beta_{i,(m)}} + \frac{\beta_{i,(m)}}{x_{ij}} - 2\right) + d_i\right)$, and then $\tilde{\alpha}_{i,(m)} = \sqrt{\tilde{\lambda}_{i,(m)}}$

(d) Compute the Bayesian estimates for η using Equation (25)

Step 4: Repeat step 3, a total M times

Step 5: Calculate L_{Baye} and U_{Baye}

Highest Posterior Density (HPD) Approach

The HPD approach is a foundational concept in Bayesian statistics, allowing the determination of credible intervals for parameters by focusing on the range of values where the posterior probability density is most densely clustered. This method is highly valuable for expressing the degree of uncertainty associated with parameter estimates and for enabling probabilistic reasoning in a Bayesian context.

The HPD interval encompasses values where the posterior density at every point within the interval surpasses the posterior densities at points outside of it. This indicates that the interval includes the parameter values that are more probable while excluding the less probable ones. Box and Tiao [5] outlined two key characteristics of the HPD interval as follows:

1. Within the interval, the probability density is higher compared to points outside of it.
2. At a specified probability level $(1-\gamma)$, the interval possesses the smallest possible length.

Using Equation (10), Li and Xu [15] demonstrated that $J(x_{ij}, (\alpha_i, \beta_i))$ corresponds to a specific instance of a prior informed by partial information, and the generalized fiducial estimates for α_i and β_i can be derived using a method analogous to that for the Bayesian posterior.

Therefore, the $100(1-\gamma)\%$ two-sided CI based on the HPD-PI approach is

$$CI_{HPD-PI} = [L_{HPD-PI}, U_{HPD-PI}], \tag{29}$$

where L_{HPD-PI} and U_{HPD-PI} are obtained by using the HDInterval package from the R software suite.

Algorithm 3: CI based on the HPD-PI approach
Step 1: Generate x_{ij} from BS distribution, where $i=1,2$ and $j=1,2,...,n_i$
Step 2: Generate K samples of α_i and β_i using the 'arms' function
Step 3: Burn-in B samples
Step 4: Reduce the number of samples by applying sampling lag $L>1$ and the final number of samples is $K'=(K-B)/L$
Step 5: Calculate $\hat{\eta}$ using Equation (13) and obtain $\hat{\eta}_{(1)},\hat{\eta}_{(2)},...,\hat{\eta}_{(K')}$
Step 6: Calculate L_{HPD-PI} and U_{HPD-PI}

Moreover, the $100(1-\gamma)\%$ two-sided CI based on the HPD-KH approach is

$$CI_{HPD-KH}=[L_{HPD-KH},U_{HPD-KH}], \tag{30}$$

where L_{HPD-KH} and U_{HPD-KH} are obtained by using the HDInterval package from the R software suite.

Algorithm 4: CI based on the HPD-KH approach
Step 1: Set a_i, b_i, c_i, and d_i, where $i=1,2$
Step 2: Compute $a(r_i)$ and $b^{+}(r_i)$
Step 3: Repeat m times
(a) Generate u_i from $U(0,a(r_i))$ and generate v_i from $U(0,b^{+}(r_i))$, and then compute $\rho_i=v_i/u_i^{r_i}$
(b) If $u_i\le\left[\pi(\rho_i\,|\,x_{ij})\right]^{1/(r_i+1)}$, set $\tilde{\beta}_{i,(m)}=\rho_i$; otherwise, repeat step (a)
(c) Generate $\tilde{\lambda}_{i,(m)}$ from $IG\left(\frac{n_i}{2}+c_i,\sum_{j=1}^{n_i}\left(\frac{x_{ij}}{\beta_{i,(m)}}+\frac{\beta_{i,(m)}}{x_{ij}}-2\right)+d_i\right)$, and then $\tilde{\alpha}_{i,(m)}=\sqrt{\tilde{\lambda}_{i,(m)}}$
(d) Compute the Bayesian estimates for η using Equation (25)
Step 4: Repeat step 3, a total M times
Step 5: Calculate L_{HPD-KH} and U_{HPD-KH}

Results

A Monte Carlo simulation was undertaken to assess the performance of the proposed CIs for the ratio of percentiles in the BS distribution, employing the R package. The evaluation of these CIs focused on their CPs and ALs, maintaining a fixed nominal confidence level of 0.95. An optimal CI was defined by its CP of 0.95 or higher and the shortest AL.

According to a study by Puggard *et al*. [20], they set the shape parameters as $(\alpha_1,\alpha_2)=$ (0.25, 0.25), (0.25, 0.50), (0.25, 1.00), (0.50, 0.50), (0.50, 1.00), and (1.00, 1.00), while the scale parameters remained fixed at $(\beta_1,\beta_2)=$ (1.00, 1.00). The study involved 1000 replications with $K=$ 3000 for the GFCI, $B=$ 1000 for HPD-PI interval, and $M=$ 1000 for Bayesian credible interval and HPD-KH interval. Additionally, according to Wang *et al*. [24], Bayesian credible interval and HPD-KH interval were considered with $(r_1,r_2)=$ (2.00, 2.00) and set the hyperparameters a_1, a_2, b_1, b_2, c_1, c_2, d_1, and d_2 to 10^{-4}.

The CPs and ALs of the CIs can be computed based on the following methodology.
Step 1: Use Algorithm 1 - Algorithm 4 to construct the CIs
Step 2: If $L\le\eta\le U$, set $P=$ 1; else set $P=$ 0
Step 3: Compute $U-L$
Step 4: Repeat step 1 - step 3, a total 1000 times
Step 5: Compute average of P defined as the CP
Step 6: Compute average of $U-L$ defined as the AL

The proposed CIs for the ratio of percentiles of BS distributions were evaluated in Table 1 and illustrated in Figures 1 and 2. In Table 1, the findings suggest that, in certain instances, the CPs of all approaches exceeded the nominal confidence level of 0.95. The HPD-PI interval exhibited shorter ALs compared to the other CIs. However, the GFCI and HPD-KH intervals were the shortest in some cases.

Figures 1 and 2 illustrate the CPs and ALs of the CIs, respectively, concerning different sample sizes and shape parameters. As depicted in Figure 1, when the sample sizes were set to 50, the CPs of all approaches matched the nominal confidence level of 0.95, but they decreased for sample sizes greater than 50. Furthermore, as the sample size increased, the AL of all approaches decreased. The CPs of all CIs were smaller when the sample sizes are larger because its ALs were shorter which was effect to the proportion of the time that the interval contains the true value. In Figure 2, the CPs of all approaches closely aligned with the nominal confidence level of 0.95 when the shape parameters were set to 0.50. Additionally, the AL of all approaches increased with the shape parameter.

Table 1. The CPs and ALs of 95% two-sided CIs for the ratio of percentiles of BS distributions

(n_1,n_2)	(α_1,α_2)	CP (AL)			
		CI_{GFCI}	CI_{Baye}	CI_{HPD-PI}	CI_{HPD-KH}
(10,10)	(0.25,0.25)	0.9540 (0.4590)	0.9530 (0.4595)	0.9460 (0.4518)	**0.9520 (0.4519)**
	(0.25,0.50)	0.9410 (0.7343)	0.9390 (0.7435)	0.9400 (0.7159)	0.9400 (0.7225)
	(0.25,1.00)	0.9300 (1.2780)	0.9360 (1.3524)	0.9200 (1.2117)	0.9310 (1.2703)
	(0.50,0.50)	0.9440 (0.9537)	0.9470 (0.9630)	0.9350 (0.9203)	0.9410 (0.9285)
	(0.50,1.00)	**0.9500 (1.4376)**	0.9530 (1.5076)	0.9450 (1.3485)	0.9470 (1.4070)
	(1.00,1.00)	0.9430 (1.9955)	0.9430 (2.0722)	**0.9500 (1.8183)**	0.9560 (1.8775)
(10,30)	(0.25,0.25)	0.9440 (0.3695)	0.9490 (0.3706)	0.9440 (0.3645)	0.9440 (0.3654)
	(0.25,0.50)	0.9350 (0.4861)	0.9420 (0.4881)	0.9340 (0.4783)	0.9430 (0.4801)
	(0.25,1.00)	0.9440 (0.7391)	0.9430 (0.7460)	0.9430 (0.7197)	0.9420 (0.7278)
	(0.50,0.50)	0.9510 (0.7521)	0.9470 (0.7562)	**0.9560 (0.7317)**	0.9560 (0.7352)
	(0.50,1.00)	**0.9510 (0.9442)**	0.9550 (0.9571)	0.9480 (0.9114)	0.9480 (0.9236)
	(1.00,1.00)	0.9380 (1.5330)	0.9400 (1.5343)	0.9310 (1.4279)	0.9350 (1.4329)
(30,30)	(0.25,0.25)	0.9470 (0.2544)	0.9450 (0.2541)	0.9480 (0.2515)	0.9440 (0.2513)
	(0.25,0.50)	0.9410 (0.3977)	0.9450 (0.3992)	0.9390 (0.3925)	0.9450 (0.3936)
	(0.25,1.00)	0.9440 (0.6845)	0.9420 (0.6927)	0.9440 (0.6693)	0.9430 (0.6774)
	(0.50,0.50)	0.9480 (0.5055)	0.9520 (0.5073)	0.9440 (0.4975)	**0.9520 (0.4992)**
	(0.50,1.00)	0.9450 (0.7490)	0.9420 (0.7566)	0.9380 (0.7305)	0.9380 (0.7374)
	(1.00,1.00)	0.9460 (0.9628)	0.9450 (0.9671)	**0.9510 (0.9302)**	0.9470 (0.9352)
(30,50)	(0.25,0.25)	0.9370 (0.2269)	0.9400 (0.2269)	0.9320 (0.2247)	0.9320 (0.2247)
	(0.25,0.50)	0.9340 (0.3261)	0.9380 (0.3266)	0.9320 (0.3224)	0.9340 (0.3226)
	(0.25,1.00)	**0.9510 (0.5331)**	0.9480 (0.5369)	0.9470 (0.5241)	0.9450 (0.5274)
	(0.50,0.50)	0.9480 (0.4479)	0.9490 (0.4479)	0.9450 (0.4411)	0.9470 (0.4411)
	(0.50,1.00)	0.9480	0.9450	0.9400	0.9420

(n_1,n_2)	(α_1,α_2)	CP (AL)			
		CI_{GFCI}	CI_{Baye}	CI_{HPD-PI}	CI_{HPD-KH}
		(0.6133)	(0.6151)	(0.6011)	(0.6031)
	(1.00,1.00)	0.9490	0.9480	**0.9500**	0.9560
		(0.8480)	(0.8494)	**(0.8243)**	(0.8247)
(50,50)	(0.25,0.25)	0.9420	0.9420	0.9390	0.9410
		(0.1958)	(0.1962)	(0.1939)	(0.1943)
	(0.25,0.50)	0.9460	0.9420	0.9390	0.9420
		(0.3054)	(0.3058)	(0.3020)	(0.3024)
	(0.25,1.00)	0.9380	0.9380	0.9350	0.9320
		(0.5186)	(0.5201)	(0.5101)	(0.5118)
	(0.50,0.50)	**0.9560**	0.9520	0.9480	0.9460
		(0.3878)	(0.3879)	(0.3828)	(0.3829)
	(0.50,1.00)	0.9460	0.9480	0.9430	0.9490
		(0.5661)	(0.5696)	(0.5557)	(0.5594)
	(1.00,1.00)	0.9610	0.9620	**0.9530**	0.9590
		(0.7222)	(0.7235)	**(0.7051)**	(0.7065)
(50,100)	(0.25,0.25)	0.9500	0.9520	**0.9510**	0.9470
		(0.1698)	(0.1699)	**(0.1682)**	(0.1684)
	(0.25,0.50)	0.9470	0.9500	0.9440	**0.9530**
		(0.2363)	(0.2366)	(0.2339)	**(0.2342)**
	(0.25,1.00)	0.9320	0.9320	0.9310	0.9330
		(0.3752)	(0.3765)	(0.3705)	(0.3718)
	(0.50,0.50)	0.9430	**0.9500**	0.9390	0.9490
		(0.3339)	**(0.3340)**	(0.3301)	(0.3300)
	(0.50,1.00)	0.9360	0.9350	0.9330	0.9290
		(0.4431)	(0.4440)	(0.4366)	(0.4376)
	(1.00,1.00)	0.9530	0.9510	**0.9530**	0.9520
		(0.6165)	(0.6166)	**(0.6042)**	(0.6048)
(100,100)	(0.25,0.25)	0.9390	0.9430	0.9350	0.9400
		(0.1375)	(0.1377)	(0.1363)	(0.1364)
	(0.25,0.50)	0.9470	0.9460	0.9450	0.9440
		(0.2150)	(0.2153)	(0.2129)	(0.2132)
	(0.25,1.00)	0.9430	0.9410	0.9340	0.9390
		(0.3609)	(0.3626)	(0.3563)	(0.3580)
	(0.50,0.50)	0.9570	0.9560	0.9490	**0.9530**
		(0.2710)	(0.2715)	(0.2682)	**(0.2687)**
	(0.50,1.00)	0.9320	0.9370	0.9280	0.9290
		(0.3990)	(0.4001)	(0.3935)	(0.3947)
	(1.00,1.00)	0.9370	0.9390	0.9380	0.9440
		(0.5000)	(0.5010)	(0.4921)	(0.4929)

Note: Bold font means the CI with CP greater than or equal to 0.95 and the shortest AL

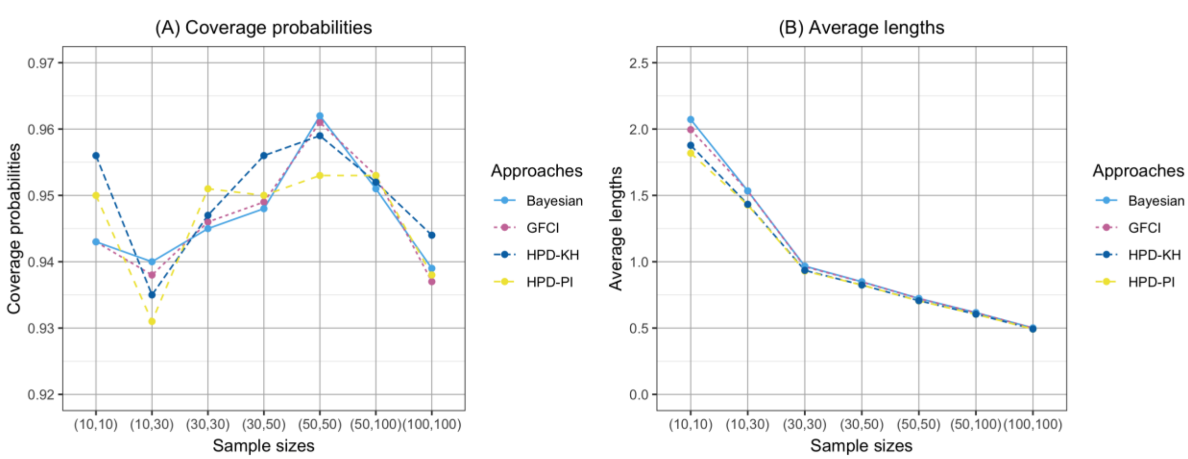

Figure 1. Comparison of the CPs and ALs of the CIs according to sample sizes

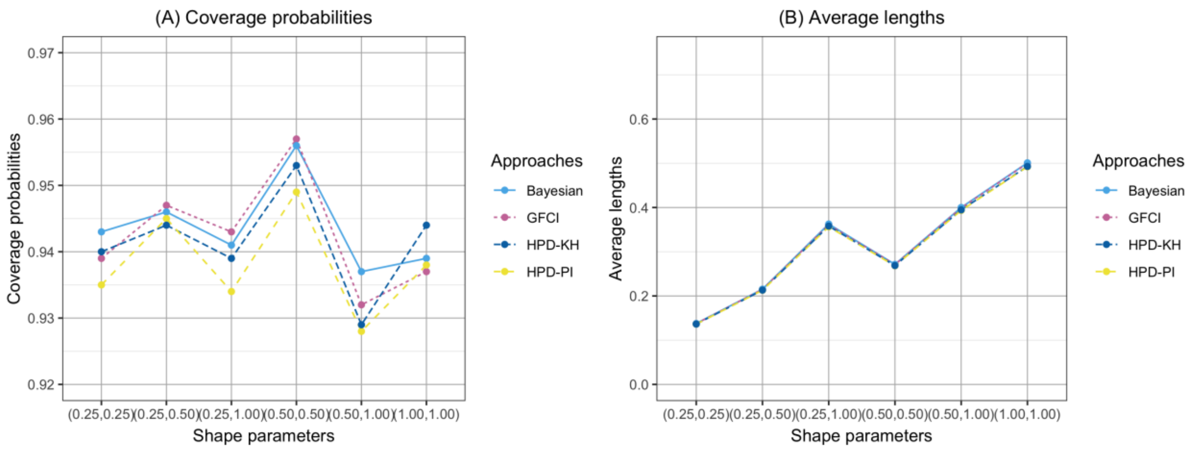

Figure 2. Comparison of the CPs and ALs of the CIs according to shape parameters

Empirical Application

The GFCI, Bayesian, HPD-PI, and HPD-KH approaches were utilized to determine CIs for the percentile ratio. Our focus was on the dataset for particulate matter 2.5 (PM2.5) levels. The daily PM2.5 level data in Phrae and Lampang provinces have become serious air pollution issues, significantly affecting health and visibility for transportation. The percentile ratio was employed to compare the PM2.5 levels in Phrae and Lampang provinces. We examined the CIs for the percentile ratio of PM2.5 levels in Phrae province and Lampang province. The Pollution Control Department collected daily PM2.5 level data in Phrae and Lampang provinces from January 1 to July 31, 2022, as displayed in Table 2. Figure 3 provides histograms compared with normal distribution curves illustrating the distribution of daily PM2.5 levels in Phrae and Lampang provinces. The suitability of the probability models for fitting the daily PM2.5 level data was evaluated through the Akaike Information Criterion (AIC). Table 3 presents the AIC values estimated for different probability models applied to the PM2.5 level data from Phrae and Lampang provinces. As shown in Table 3, the PM2.5 level data from these provinces was best fit by the BS distributions.

Table 2. Daily PM2.5 levels data in Phrae and Lampang provinces

Provinces	**Daily PM2.5 levels ($\mu g / m^3$)**									
Phrae	26	18	26	26	29	31	34	38	33	36
	38	44	39	45	38	42	21	25	29	27
	29	27	21	29	31	31	35	45	43	38
	56	75	34	38	47	22	17	24	21	28
	39	45	29	36	46	36	23	23	20	14
	11	11	19	28	29	34	40	46	50	59
	64	60	52	52	51	37	19	31	40	51
	52	57	59	55	46	34	25	25	46	18
	15	20	19	33	40	48	42	25	31	32
	32	12	12	35	51	64	81	84	102	87
	74	40	55	70	55	48	52	22	16	16
	19	24	29	31	33	35	39	42	24	25
	17	14	19	32	38	32	15	8	10	11
	9	8	11	13	15	15	10	11	12	9
	6	12	17	13	14	21	20	21	20	23
	16	16	22	20	8	7	8	9	8	12
	11	12	11	14	14	15	10	8	9	8
	7	8	7	8	9	9	14	12	8	8
	11	10	8	9	10	10	9	8	7	6
	6	7	11	6	6	6	9	12	11	13
	8	6	6	5	8	9	11	15	22	24
	19	8								
Lampang	19	15	19	22	26	21	26	24	26	24
	31	37	35	33	31	33	15	16	20	17

Provinces	Daily PM2.5 levels ($\mu g / m^3$)									
	17	16	13	17	21	22	21	25	20	24
	27	36	24	27	36	13	16	17	19	27
	29	41	31	34	53	50	16	19	17	8
	8	18	15	24	22	30	41	40	48	62
	54	52	48	42	42	38	14	22	30	30
	37	41	46	48	40	34	26	29	33	17
	11	10	11	26	31	40	43	22	26	31
	27	8	8	21	38	41	57	74	77	73
	59	43	62	69	55	54	57	28	18	13
	16	22	30	33	37	38	41	41	23	19
	15	15	14	22	24	23	13	8	9	10
	9	9	12	13	15	14	7	8	9	7
	5	13	16	12	13	20	19	23	18	20
	16	12	17	16	7	6	6	7	8	10
	13	10	10	12	10	10	7	6	5	6
	5	6	5	7	5	7	9	8	6	6
	5	5	5	5	5	7	6	5	5	6
	5	6	8	5	5	4	7	8	6	8
	5	5	4	6	6	8	8	10	13	14
	10	5								

Source: Pollution Control Department, Thailand http://air4thai.pcd.go.th/webV3/#/History

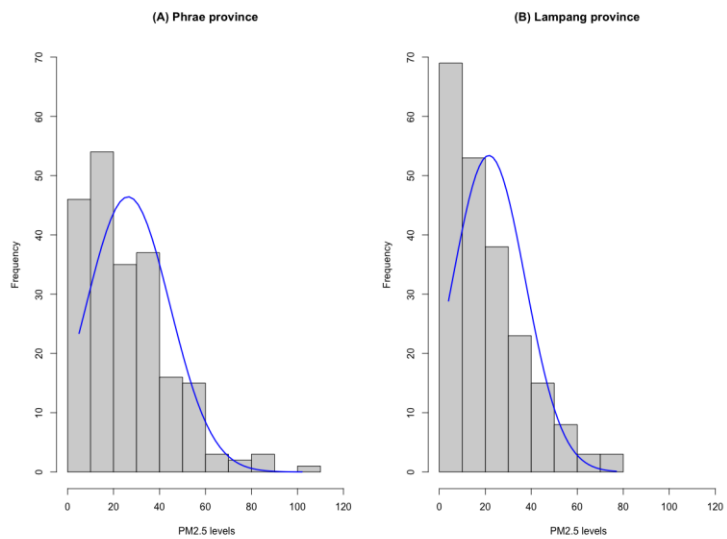

Figure 3. Histograms of the daily PM2.5 level data for Phrae and Lampang provinces

Table 3. The estimated AIC values for the seven probability models, using the PM2.5 level data from Phrae and Lampang provinces

Distributions	Phrae province	Lampang province
Normal	1835.17	1775.88
Log-normal	1744.43	1673.66
Weibull	1758.89	1688.17
Gamma	1750.37	1680.63
Exponential	1812.78	1728.25
Cauchy	1889.10	1815.74
BS	1738.24	1666.50

Table 4. Sample statistics of the daily PM2.5 level data in Phrae and Lampang provinces

Statistics	Phrae province	Lampang province
Minimum	5.00	4.00
Mean	26.33	21.57
Maximum	102.00	77.00
n	212	212
$\hat{\alpha}$	0.74	0.79
$\hat{\beta}$	18.87	15.20
$\hat{\theta}$	18.87	15.20

Table 4 displays the sample statistics of the PM2.5 level data in Phrae and Lampang provinces. The estimated ratio of percentiles of daily PM2.5 level data in Phrae and Lampang provinces was 1.2414. Table 5 displays the 95% CIs for the percentile ratio of daily PM2.5 level data in Phrae and Lampang provinces using four approaches. According to Table 4, the HPD-KH approach provided the shortest interval length.

Table 5. The lower limit (L) and upper limit (U) of the 95% CIs for the ratio of percentiles of the daily PM2.5 level data in Phrae and Lampang provinces

Approaches	$[L, U]$	Interval length
GFCI	[1.1118,1.4493]	0.3375
Bayesian	[1.0994,1.4349]	0.3355
HPD-PI	[1.1163,1.4525]	0.3362
HPD-KH	[1.0884,1.4205]	0.3321

Discussion

CIs for the ratio of percentiles of BS distributions were established through four distinct approaches: the GFCI approach, the Bayesian approach, the HPD-PI approach, and the HPD-KH approach. The HPD-PI approach, specifically, stood out by producing comparatively shorter ALs. All four methods employed simulation data in their CI construction.

Based on the outcomes of the study, it can be inferred that, for constructing CIs for the ratio of percentiles in BS distributions, the HPD-PI approach is the recommended choice. This recommendation is consistent with the findings presented in Puggard *et al*. [20].

Conclusion

CIs for the percentile ratio were created using various approaches, including GFCI, Bayesian, HPD-PI, and HPD-KH approaches. Using example from the PM2.5 dataset in Phrae and Lampang provinces, all approaches were illustrated through real data analysis. The performance of the CIs agreed with our simulation studies, as the ALs of GFCI, HPD-PI, and HPD-KH were shorter in some cases.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgment

This research has received funding support from the National Science, Research and Innovation Fund (NSRF), and King Mongkut’s University of Technology North Bangkok: KMUTNB-FF-67-B-06.

References

[1] Achcar, J. A. (1993). Inferences for the Birnbaum-Saunders fatigue life model using Bayesian methods. *Computational Statistics and Data Analysis, 15*(4), 367-380.

[2] Bhattacharyya, G., & Fries, A. (1982). Fatigue failure Models-Birnbaum-Saunders vs. Inverse Gaussian. *IEEE Transactions on Reliability, 31*(5), 439-441.

[3] Birnbaum, Z. W., & Saunders, S. C. (1969). A new family of life distributions. *Journal of Applied Probability, 6*(2), 319-327.

[4] Birnbaum, Z. W., & Saunders, S. C. (1969). Estimation for a family of life distributions with applications to fatigue. *Journal of Applied Probability, 6(2),* 328-347.

[5] Box, G. E. P., & Tiao, G. C. (1992). *Bayesian inference in statistical analysis*. USA: Wiley.

[6] Chang, D. S., & Tang, L. C. (1994). Percentile bounds and tolerance limits for the Birnbaum-Saunders distribution. *Communications in Statistics-Theory and Methods, 23*(10), 2853-2863.

[7] Desmond, A. F. (1986). On the relationship between two fatigue-life models. *IEEE Transactions on Reliability, 35*(2), 167-169.

[8] Durham, S. D., & Padgett, W. J. (1997). Cumulative damage models for system failure with application to carbon fibers and composites. *Technometrics, 39*(1), 34-44.

[9] Gilks, W. R., & Wild, P. (1992). Adaptive rejection sampling for Gibbs sampling. *Journal of the Royal Statistical Society Series C Applied Statistics, 41*(2), 337-348.

[10] Gilks, W. R., Best, N. G., & Tan, K. K. C. (1995). Adaptive rejection metropolis sampling within Gibbs sampling. *Journal of the Royal Statistical Society Series C Applied Statistics, 44*(4), 455-472.

[11] Guo, X., Wu, H., Li, G., & Li, Q. (2017). Inference for the common mean of several Birnbaum-Saunders populations. *Journal of Applied Statistics, 44*(5), 941-954.

[12] Hannig, J. (2009). On generalized fiducial inference. *Statistica Sinica, 19*(2), 491-544.

[13] Jayalath, K. P. (2021). Fiducial inference on the right censored Birnbaum-Saunders data via Gibbs sampler. *Stats, 4*(2), 385-399.

[14] Leiva, V., Barros, M., Paula, G. A., & Sanhueza, A. (2008). Generalized Birnbaum-Saunders distributions applied to air pollutant concentration. *Environmetrics, 19*(3), 235-249.

[15] Li, Y. L., & Xu, A. C. (2016). Fiducial inference for Birnbaum-Saunders distribution. *Journal of Statistical Computation and Simulation, 86*(9), 1673-1685.

[16] Lu, M. C., & Chang, D. S. (1997). Bootstrap prediction intervals for the Birnbaum-Saunders distribution. *Microelectronics Reliability, 37*(8), 1213-1216.

[17] Ng, H. K. T., Kundu, D., & Balakrishnan, N. (2003). Modified moment estimation for the two-parameter Birnbaum-Saunders distribution. *Computational Statistics and Data Analysis, 43*(3), 283-298.

[18] Niu, C., Guo, X., Xu, W., & Zhu, L. (2014). Comparison of several Birnbaum-Saunders distributions. *Journal of Statistical Computation and Simulation, 84*(12), 2721-2733.

[19] Padgett, W. J., & Tomlinson, M. A. (2003). Lower confidence bounds for percentiles of Weibull and Birnbaum-Saunders distributions. *Journal of Statistical and Simulation, 73*(6), 429-443.

[20] Puggard, W., Niwitpong, S-A., & Niwitpong, S. (2022). Confidence intervals for comparing the variances of two independent Birnbaum-Saunders distributions. *Symmetry, 14*(1492), 1-12.

[21] Vilca, F., Sanhueza, A., Leiva, V., & Christakos, G. (2010). An extended Birnbaum-Saunders model and its application in the study of environmental quality in Santiago, Chile. *Stochastic Environmental Research and Risk Assessment, 24*(5), 771-782.

[22] Wakefield, J. C., Gelfand, A. E., & Smith, A. F. M. (1991). Efficient generation of random variates via the ratio-of-uniforms method. *Statistics and Computing, 1,* 129-133.

[23] Wang, B. X. (2012). Generalized interval estimation for the Birnbaum-Saunders distribution. *Computational Statistics and Data Analysis, 56*(12), 4320-4326.

[24] Wang, M., Sun, X., & Park, C. (2016). Bayesian analysis of Birnbaum-Saunders distribution via the generalized ratio-of-uniforms method. *Computational Statistics, 31,* 207-225.

[25] Wu, J., & Wong, A. C. M. (2004). Improved interval estimation for the two-parameter Birnbaum-Saunders distribution. *Computational Statistics and Data Analysis, 47*(4), 809-821.

[26] Xu, A., & Tang, Y. (2010). Reference analysis for Birnbaum-Saunders distribution. *Computational Statistics and Data Analysis, 54*(1), 185-192.