

Article

A Penalized Likelihood Approach to Parameter Estimation with Integral Reliability Constraints

Barry Smith ¹, Steven Wang ², Augustine Wong ^{2,*} and Xiaofeng Zhou ²

¹ Department of Economics, York University, Toronto ON M3J 1P3, Canada;

E-Mail: jbsmith@yorku.ca

² Department of Mathematics and Statistics, York University, Toronto ON M3J 1P3, Canada;

E-Mails: stevenw@yorku.ca (S.W.); bluesky@yorku.ca (X.Z.)

* Author to whom correspondence should be addressed; E-Mail: august@yorku.ca;

Tel.: +1-416-736-5250

Academic Editors: Carlos Alberto de Bragança Pereira and Adriano Polpo

Received: 21 April 2015 / Accepted: 10 June 2015 / Published: 12 June 2015

Abstract: Stress-strength reliability problems arise frequently in applied statistics and related fields. Often they involve two independent and possibly small samples of measurements on strength and breakdown pressures (stress). The goal of the researcher is to use the measurements to obtain inference on reliability, which is the probability that stress will exceed strength. This paper addresses the case where reliability is expressed in terms of an integral which has no closed form solution and where the number of observed values on stress and strength is small. We find that the Lagrange approach to estimating constrained likelihood, necessary for inference, often performs poorly. We introduce a penalized likelihood method and it appears to always work well. We use third order likelihood methods to partially offset the issue of small samples. The proposed method is applied to draw inferences on reliability in stress-strength problems with independent exponentiated exponential distributions. Simulation studies are carried out to assess the accuracy of the proposed method and to compare it with some standard asymptotic methods.

Keywords: confidence interval; coverage probability; delta method; exponentiated exponential distribution; penalized likelihood; r^* -formula

1. Introduction

We consider a stress-strength reliability problem, $R = P(Y < X)$, where X and Y are independently distributed as exponentiated exponential random variables. In this case, the parameter of interest, R , is an integral with no closed form solution. Likelihood-based inference for this problem involves, in part, numerical constrained optimization. The likelihood ratio test, for example, uses constrained maximum likelihood parameter estimates where the log-likelihood function is maximized subject to the constraint that the parameter of interest takes on the null-hypothesized value.

A standard approach to solving (equality) constrained problems is to apply the Lagrange method. Detailed discussions can be found in many reference works in optimization, for example, see [1]. Using the Lagrange technique, this problem requires that, in each iteration, one parameter has to be used to guarantee that the integral equality constraint is satisfied up to a given level of numerical accuracy. Satisfying the constraint involves simultaneous numerical integration and the numerical solution of a nonlinear equation for the correct parameter value. Moreover, all of this needs to be repeated for each value of the remaining parameters.

We found that the Lagrange approach led to problems of numerical accuracy and slow convergence. We also discovered that the integral function was homogeneous of degree 0 in two of the parameters of the model. The integral is therefore insensitive to equi-proportionate changes in the two parameters and this complicated optimization.

Penalty function methods are also often used to solve equality and inequality-constrained optimization problems. See Byrne [2] and the extensive list of references therein. The basic idea is to solve a constrained optimization problem by solving a sequence of unconstrained problems where constraint violation is penalized in a successively harsher manner. The sequence of unconstrained optima converges to the constrained optimum. Smith *et al.* (2014) [3] demonstrates the accuracy of the penalty method in a constrained optimization problem. When implementing this approach to the problem considered in this paper, we found it was numerically stable and rapid in the sense that parameter estimates from successively more harshly penalized models quickly converged to optimal values and became effectively insensitive to further penalization.

The remainder of the paper is organized as follows. Section 2 begins with a brief description of our stress-strength reliability problem. We then examine some of the properties of the unconstrained likelihood function. This seems necessary because the likelihood function is not concave and generally cannot be optimized by an algorithm that uses the exact Hessian matrix. We then introduce the integral constraint and some of its properties together with a numerical example to illustrate how the Lagrange approach failed. The section concludes with a discussion of some properties and an implementation of the penalty function approach. Section 3 briefly reviews the likelihood-based third-order method for inference concerning a scalar parameter of interest when samples are small. Section 4 applies the approaches developed in Sections 2 and 3 to study inference for the stress-strength reliability problem with independent exponentiated exponential distributions. Results from simulation studies are presented to illustrate the extreme accuracy of our suggested approach. Some concluding remarks are given in Section 5. Technical details related to some issues raised in Section 2 are presented in the Appendix.

2. Computational Issues and Penalized Likelihood

2.1. Problem

We consider statistical inference for the reliability, $R = P(Y < X)$ where X and Y are independently distributed with exponentiated exponential distributions. The estimation of R is important in the statistical literature and has been widely studied in other areas, such as radiology, mechanical engineering and materials science. In the context of the stress-strength model, reliability refers to the probability that the unit's strength Y is less than the stress X . Following the notation in [4], the cumulative distribution function of the two-parameter exponentiated exponential distribution, $EE(\alpha, \beta)$, is

$$F(x; \alpha, \beta) = (1 - e^{-\beta x})^\alpha; \quad \alpha > 0, \beta > 0, x > 0 \quad (1)$$

where α is the shape parameter and β is the scale parameter. The $EE(\alpha, \beta)$ distribution is also known as the generalized exponential distribution. It is equivalent to the exponentiated Weibull ($\kappa = 1, \alpha, \sigma$) distribution as introduced in [5], where κ is the first shape parameter, α is the second shape parameter and σ is the scale parameter. If $\alpha = 1$, the distribution reduces to the one parameter exponential (β) model. The $EE(\alpha, \beta)$ distribution is recognized as a useful model for lifetime data or skewed data. It has a monotone increasing hazard function when $\alpha > 1$, decreasing when $\alpha < 1$ and constant when $\alpha = 1$. It also has a unimodal and right-skewed density function. For a fixed scale parameter value, the skewness gradually decreases as the shape parameter increases.

Kundu and Gupta [6] and Raqab *et al.* [7] considered maximum likelihood estimation of R when X and Y are independent $EE(\alpha_1, \beta_1)$ and $EE(\alpha_2, \beta_2)$ distributed random variables and where β_1 and β_2 are assumed to be the same. Then $R = \frac{\alpha_1}{\alpha_1 + \alpha_2}$. Likelihood-based inference for R can then be obtained easily. However, without the assumption that $\beta_1 = \beta_2$, then

$$R = P(Y < X) = \int_0^\infty \alpha_1 \beta_1 (1 - e^{-\beta_1 x})^{\alpha_1 - 1} e^{-\beta_1 x} (1 - e^{-\beta_2 x})^{\alpha_2} dx \quad (2)$$

and this integral does not have a known closed form. To obtain likelihood-based inference for R in this case, it is necessary to solve an optimization problem with an integral constraint. We found that standard macros, functions and subroutines in popular software packages such as R or Matlab experienced problems in converging to constrained maximum likelihood estimates in a Lagrange setting. As noted in the Introduction, a penalized likelihood method is proposed to provide a solution for this constrained maximization problem.

Once the constrained maximum likelihood estimates have been obtained, likelihood methods can be used to draw inferences about R . In this paper, the third-order method discussed in Fraser and Reid [8] is applied. This is known to be very accurate even when the sample sizes are small. See, for example, Chang and Wong [9] and She *et al.* [10] where the accuracy of third order methods has been assessed.

2.2. Unconstrained Likelihood and Its Properties

Let $x = (x_1, \dots, x_n)'$ and $y = (y_1, \dots, y_m)'$ be independent random samples from $EE(\alpha_1, \beta_1)$ and $EE(\alpha_2, \beta_2)$ respectively. Then the log-likelihood function is:

$$\begin{aligned} l(\theta; x, y) &= l(\alpha_1, \beta_1, \alpha_2, \beta_2; x, y) \\ &= n \log \alpha_1 + n \log \beta_1 + (\alpha_1 - 1) \sum_{i=1}^n \log(1 - e^{-\beta_1 x_i}) - \beta_1 \sum_{i=1}^n x_i \\ &\quad + m \log \alpha_2 + m \log \beta_2 + (\alpha_2 - 1) \sum_{j=1}^m \log(1 - e^{-\beta_2 y_j}) - \beta_2 \sum_{j=1}^m y_j \end{aligned} \quad (3)$$

where $\theta = (\alpha_1, \alpha_2, \beta_1, \beta_2)'$.

The log-likelihood function is infinitely differentiable and has some interesting properties. In the first place it is not concave. Nor is it quasiconcave or pseudoconcave or a member of any of the other classes of generalized concave functions that appear regularly in the literature on optimization. However, the function is strictly concave in each of its parameters individually. For fixed values of any 3 parameters, the function reaches a unique maximum with respect to the fourth. The determinant of the Hessian matrix changes sign throughout the domain of the function. But, the gradient of the log-likelihood function vanishes at a unique vector of parameter values and at that point and in a neighbourhood around it the Hessian matrix is negative definite. All of these results are derived in the Appendix. So, when the function is restricted to an open region about the global optimum, the function is concave. These results have some important implications regarding how to maximize log-likelihood for any given sample. In particular, all parameters cannot typically be estimated simultaneously using any algorithm that uses exact Hessian information. Quasi-Newton techniques (sometimes called variable metric) work by building an increasingly accurate quadratic approximation to the log-likelihood surface. These will work in this setting as long as exact Hessians are not used building the surface approximation and updating the parameters. Hessian matrices from the exponentiated exponential model usually won't be negative definite and will lead to inappropriate updating. The Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm can work because some versions of it start with negative definite and symmetric "approximate" Hessians and update them until they converge to the true Hessian at the optimum. Given the independence of X and Y it is also possible to concentrate the log-likelihood using the derivative information on α_1 and α_2 and then apply simple line searches to find the β parameters. The Appendix shows why this latter, perhaps more cumbersome, approach will work but only in the unconstrained case.

In summary, obtaining the unconstrained parameter estimates is possible if done carefully. As well, we found it is often helpful to restrict large changes in parameters during estimation. We recommend box constraints be placed on parameters during estimation. Preset upper and lower bounds on the parameters require that the parameters must be chosen from a four dimensional hyper-cube. A version of BFGS that implants constraints of this type is available in the statistical package, *R*.

2.3. The Integral Reliability Constraint

In this paper, the parameter of interest is stress-strength reliability, $R = P(Y < X)$, where X and Y are independently distributed as $EE(\alpha_1, \beta_1)$ and $EE(\alpha_2, \beta_2)$ distributions respectively. The constraint function, R , is given in (2) and has no known closed form solution. Straightforward differentiation shows that R is an increasing function of α_1 and β_2 and a decreasing function of α_2 and β_1 . Introducing the change of variables: $z = \beta_1 x$, the integral constraint can be expressed equivalently as:

$$R = \int_0^\infty \alpha_1 (1 - e^{-z})^{\alpha_1-1} e^{-z} (1 - e^{-\frac{\beta_2}{\beta_1} z})^{\alpha_2} dz$$

This establishes that R is homogeneous of degree 0 in (β_1, β_2) . The contours of R are all constant along the line in parameter space satisfying $\beta_2 = c\beta_1$. This can complicate the numerical optimization process. It also provides a good reason to introduce the box constraints on the parameters as discussed in Section 2.2.

2.4. Constrained Optimization: Lagrange

We need to find the constrained maximum likelihood parameter estimates $\hat{\theta}_\psi = (\tilde{\alpha}_1, \tilde{\alpha}_2, \tilde{\beta}_1, \tilde{\beta}_2)'$ for a given ψ_0 . This requires that we maximize the log-likelihood function in (3) subject to the constraint $R = \psi(\theta) = \psi_0$ in (2). Using the Lagrange approach, for given values of ψ_0 and three of the likelihood parameters, the constraint $R = \psi(\theta) = \psi_0$ can be numerically integrated and the fourth parameter simultaneously chosen so that the integral constraint holds. The value of the fourth parameter and ψ_0 are held fixed while the log-likelihood function is then maximized with respect to the three “free” likelihood parameters. This process is iterated until convergence.

There is much that can go wrong in the above iterative process. The likelihood function is not concave, the equality constraint is highly nonlinear in the parameters, the constraint integral must be evaluated numerically and then solved numerically and the integral R is homogeneous of degree 0 with respect to two of the parameters.

We discovered in simulation that this iterative Lagrange method sometimes gave relatively satisfactory results for some values of ψ_0 but it often went astray. Typically it took a long time for the estimation process to converge. Performance degraded quickly when R was constrained to be closer to the boundary values of 0 or 1.

Table 1 records one of these situations for a simulated data set from $EE(2, 3)$ and $EE(5, 1.6015)$ distributions each with a sample size of 10 and ψ_0 was set equal to 0.1. Initial attempts to program the search algorithm in Matlab resulted in floating point overflow and division by zero errors.

Table 1. Dataset1: Two Simulated Datasets.

Data Set	Observations					Sample Size
X	0.4977	0.0781	0.3827	0.2694	0.4125	10
	0.6414	0.2669	0.1978	0.1968	0.2397	
Y	1.7057	1.0191	0.5899	0.9031	0.9207	10
	1.9481	2.1290	0.8109	1.6463	1.9842	

We next tried Matlab’s built-in macro, “fminsearch”, which uses a simplex search method. This macro provided the correct unconstrained parameter estimates but failed when constraints were introduced. The unconstrained parameter estimates are recorded in Table 2. Not surprisingly, these unconstrained parameter estimates do not solve the constrained optimization problem. They yield a constraint value of 0.0141 as opposed 0.1. Correct results were found using a penalized likelihood method (appearing as Penalty in the table). This method is introduced below.

Table 2. Parameter Estimates for Dataset1.

Method	α_1	α_2	β_1	β_2	R	Loglikelihood
Unconstrained	4.4239	8.5227	6.7793	2.0262	0.0142	−3.3191
Constrained: Penalty	3.6028	3.2018	5.2707	1.5700	0.0989	−5.1659

We encountered similar problems with small samples and when sample sizes were unequal. We reprogrammed the Lagrange constrained optimization algorithm in the statistical package R. In addition to slow, inaccurate and sometimes failed convergence, we also encountered cases of singular matrices during the optimization process.

2.5. The Penalized Likelihood Approach

As noted in the previous section, we encountered a wide variety of serious numerical problems when we tried to implement the classical Lagrange approach to our integral equality constrained optimization problem. We now present a penalized likelihood method, which was also discussed in Smith *et al.* (2014) [3], that reliably solves the constrained optimization problems addressed in this paper. We begin by defining the penalized likelihood function, PL , as:

$$PL(\theta, \psi_0; x, y) = l(\theta; x, y) - k(R(\theta) - \psi_0)^2, \quad k > 0 \quad (4)$$

The function consists of the log-likelihood function $l(\theta; x, y)$ and a positive function, $k(R(\theta) - \psi_0)^2$, called the penalty function, that is subtracted from the the likelihood function. Here we consider a penalty function that is the square of the divergence of R from the constraint value ψ_0 . The greater is k , the more the likelihood is penalized (decreased) for given divergences of R from ψ_0 . Next, consider the unconstrained problem of maximizing $PL(\theta, \psi_0; x, y)$ with respect to θ for a given value of k . Formally, we obtain an optimal vector θ_k^* . When $k = 0$, we obtain the parameter vector that maximizes the

unconstrained log-likelihood. As k increases and we successively solve the optimization problems, we obtain a sequence of optimal parameter vectors that converges to the solution of the constrained problem of choosing θ to maximize $l(\theta; x, y)$ subject to $R(\theta) = \psi_0$. In contrast to the Lagrange approach, we now have a sequence of unconstrained optimization problems instead of one constrained problem. This means the penalty approach may be theoretically more complicated. But, in practice, we found it is faster and more reliable than the Lagrange approach to our problem. First, the penalty approach does not require that the equality $R(\theta) = \psi_0$ hold at any stage of the optimization process. Instead, divergence of $R(\theta)$ from ψ_0 is (increasingly) discouraged as k increases. Many of the numerical problems we encountered in applying the Lagrange technique arose from trying to impose $R(\theta) = \psi_0$ exactly. Second, in practice as k increases we found that the estimated parameters quickly converged to optimal values and it was not necessary to continue to increase k and solve additional optimization problems. Finally, the Lagrange and Penalty approaches are formally dual so that the Lagrange multiplier can be recovered as the limiting value of the partial derivative of PL with respect to the parameter ψ_0 .

A recent analysis of the rigorous foundations of the penalty (and other) approaches to constrained optimization can be found in [2] and the discussion that follows draws upon Byrne's presentation. We need very few technical conditions to hold in order for the sequence of PL -maximizing vectors, θ_k^* , to converge to the solution of the constrained optimization problem. First, we need the permissible parameter values to be drawn from a compact set. We satisfy this condition by imposing box constraints on the four likelihood parameters. Second, there is a straightforward restriction that l and R are continuous functions of the parameters. These conditions are satisfied. Finally, we require that the sequence of optimal vectors, θ_k^* , correspond to the global maximum of penalized likelihood for each k . Our experience is that, for each k , there is only ever one optimum and that occurs where the gradient of PL vanishes. In all cases the Hessian matrix at the optimum was negative definite.

When implementing the penalty approach we found that when $0.1 < \psi_0 < 0.9$, there was little gain from increasing k beyond 10,000. For the extreme values of ψ_0 , k did not have to exceed 80,000. As well, short sequences were sufficient. It was not necessary solve separate optimization problems for all possible values of k .

We obtained parameter estimates from unconstrained optimization of PL using the *optim* function in the statistical software package, R. Within *optim*, we adopted the *L-BFGS-B* algorithm. *L-BFGS-B* was developed by Byrd *et al.* [11] and is a quasi-Newton optimization variant of the Broydon-Fletcher-Goldfarb-Shanno (*BFGS*) algorithm. *L-BFGS-B* conveniently allows for simple upper and lower search bounds (box constraints) on the parameters. As noted earlier, *L-BFGS-B* does not use an exact Hessian. Instead it updates an approximate Hessian that is guaranteed to be negative definite and symmetric. The approximate Hessian converges to the true one. Our code is available upon request.

3. Likelihood-based Inference for Any Scalar Parameter of Interest

In this section, we provide a brief review of likelihood-based inference for any scalar parameter of interest. Let $y = (y_1, \dots, y_n)'$ be a random sample from a distribution with log-likelihood function

$\ell(\theta) = \ell(\theta; y)$, where θ is a vector parameter with $\dim(\theta) = p$. Also let $\psi = \psi(\theta)$ be a scalar parameter of interest. Denote $\hat{\theta}$ to be the overall maximum likelihood estimate, which satisfies

$$\ell_{\theta}(\hat{\theta}) = \left. \frac{\partial \ell(\theta)}{\partial \theta} \right|_{\theta=\hat{\theta}} = 0$$

Moreover, denote $\hat{\theta}_{\psi}$ as the constrained maximum likelihood estimate of θ for a given $\psi(\theta) = \psi_0$. $\hat{\theta}$ and $\hat{\theta}_{\psi}$ can be obtained from the penalized likelihood method discussed in Section 2. Moreover, with $\hat{\theta}_{\psi}$, the corresponding Lagrange multiplier, \hat{K} , can be obtained. Define the tilted log-likelihood function as

$$\tilde{\ell}(\theta) = \ell(\theta) + \hat{K}[\psi(\theta) - \psi_0] \quad (5)$$

which has the property $\tilde{\ell}(\hat{\theta}_{\psi}) = \ell(\hat{\theta}_{\psi})$.

Two widely used likelihood-based methods for obtaining asymptotic confidence interval for ψ are based on the maximum likelihood estimator of θ and the signed log-likelihood ratio statistic. It is well-known that $(\hat{\theta} - \theta)'[var(\hat{\theta})]^{-1}(\hat{\theta} - \theta)$ is asymptotically distributed as χ^2 -distribution with p degrees of freedom, and $var(\hat{\theta})$ can be estimated by the inverse of either the expected Fisher information matrix or the observed information matrix evaluated at $\hat{\theta}$. In practice, the latter,

$$\widehat{var}(\hat{\theta}) \approx j_{\theta\theta'}^{-1}(\hat{\theta}) = \left[-\ell_{\theta\theta'}(\hat{\theta}) \right]^{-1} = \left[-\frac{\partial^2 \ell(\theta)}{\partial \theta \partial \theta'} \right]_{\theta=\hat{\theta}}^{-1}$$

is preferred because of the simplicity in calculation. By applying the Delta method, we have

$$\widehat{var}(\hat{\psi}) = \widehat{var}(\psi(\hat{\theta})) \approx \psi'_{\theta}(\hat{\theta}) \widehat{var}(\hat{\theta}) \psi_{\theta}(\hat{\theta}) \quad (6)$$

where $\psi_{\theta}(\hat{\theta}) = \left. \frac{\partial \psi(\theta)}{\partial \theta} \right|_{\theta=\hat{\theta}}$. Hence $\frac{\hat{\psi} - \psi}{\sqrt{\widehat{var}(\hat{\psi})}}$ is asymptotically distributed as $N(0, 1)$. Thus a $(1 - \gamma)100\%$ confidence interval of ψ based on the maximum likelihood estimator is

$$\left(\hat{\psi} - z_{\gamma/2} \sqrt{\widehat{var}(\hat{\psi})}, \hat{\psi} + z_{\gamma/2} \sqrt{\widehat{var}(\hat{\psi})} \right) \quad (7)$$

where $z_{\gamma/2}$ is the $(1 - \gamma/2)100^{th}$ percentile of the standard normal distribution.

Alternatively, with regularity conditions stated in [12], the signed log-likelihood ratio statistic is

$$\begin{aligned} r(\psi) &= \text{sgn}(\hat{\psi} - \psi) \{2[\ell(\hat{\theta}) - \ell(\hat{\theta}_{\psi})]\}^{1/2} \\ &= \text{sgn}(\hat{\psi} - \psi) \{2[\ell(\hat{\theta}) - \tilde{\ell}(\hat{\theta}_{\psi})]\}^{1/2} \end{aligned} \quad (8)$$

is asymptotically distributed as $N(0, 1)$. Therefore, a $(1 - \gamma)100\%$ confidence interval of ψ based on the signed log-likelihood ratio statistic is

$$\{\psi : |r(\psi)| \leq z_{\gamma/2}\} \quad (9)$$

It should be noted that both methods have rates of convergence only $O(n^{-1/2})$. While the maximum likelihood estimator based interval is often preferred because of the simplicity in calculation, the signed log-likelihood ratio method is invariant to reparametrization and the maximum likelihood estimator based method is not. Doganaksoy and Schmee [13], showed that the signed log-likelihood ratio statistic

based interval has better coverage property than the maximum likelihood estimator based interval in cases they considered.

In recent years, various adjustments to the signed log-likelihood ratio statistic have been proposed to improve the accuracy of the signed log-likelihood ratio statistic. Reid [14] and Severini [15] gave detail overview of this development. In this paper, we consider the modified signed log-likelihood ratio statistic, which was introduced by [16,17] and has the form

$$r^*(\psi) = r(\psi) + r(\psi)^{-1} \log \left\{ \frac{Q(\psi)}{r(\psi)} \right\} \quad (10)$$

where $r(\psi)$ is the signed log-likelihood ratio statistic, and $Q(\psi)$ is a statistic that based on the log-likelihood function and an ancillary statistic. Barndorff-Nielsen [16,17] showed that $r^*(\psi)$ is asymptotically distributed as $N(0,1)$ with rate of convergence $O(n^{-3/2})$. Hence a $100(1 - \gamma)\%$ confidence interval for $\psi(\theta)$ is given by

$$\{\psi : |r^*(\psi)| \leq z_{\gamma/2}\} \quad (11)$$

However, for a general model, an ancillary statistic might not exist, and even when it does, it might not be unique. Fraser and Reid [8] showed that $Q(\psi)$ is a standardized maximum likelihood departure in the canonical parameter scale and Fraser *et al.* [18] derived the general formula for obtaining the statistic $Q(\psi)$. More specifically, Fraser *et al.* [18] obtained the locally defined canonical parameter

$$\varphi(\theta) = \frac{\partial \ell(\theta)}{\partial y} V \quad (12)$$

where

$$V = - \left(\frac{\partial z(y; \theta)}{\partial y} \right)^{-1} \left(\frac{\partial z(y; \theta)}{\partial \theta} \right) \Big|_{\theta=\hat{\theta}} \quad (13)$$

is the ancillary direction and $z(y; \theta) = (z_1(y; \theta), \dots, z_n(y; \theta))'$ is a vector of pivotal quantities. Then the standardized maximum likelihood departure in $\varphi(\theta)$ scale takes the form:

$$Q(\psi) = \text{sgn}(\hat{\psi} - \psi) \frac{|\chi(\hat{\theta}) - \chi(\hat{\theta}_\psi)|}{\sqrt{\widehat{\text{var}}(\chi(\hat{\theta}) - \chi(\hat{\theta}_\psi))}} \quad (14)$$

where

$$\chi(\theta) = \psi_\theta(\hat{\theta}_\psi) \varphi_\theta^{-1}(\hat{\theta}_\psi) \varphi(\theta) \quad (15)$$

is the recalibrated parameter of interest in the $\varphi(\theta)$ scale, and $|\chi(\hat{\theta}) - \chi(\hat{\theta}_\psi)|$ is a measure of maximum likelihood departure of $|\hat{\psi} - \psi|$ in $\varphi(\theta)$ scale,

$$\psi_\theta(\theta) = \frac{\partial \psi(\theta)}{\partial \theta} \quad \text{and} \quad \varphi_\theta(\theta) = \frac{\partial \varphi(\theta)}{\partial \theta}$$

Since the exact $\text{var}(\chi(\hat{\theta}) - \chi(\hat{\theta}_\psi))$ is difficult to obtain, Fraser *et al.* [18] showed that it can be approximated by

$$\widehat{\text{var}}(\chi(\hat{\theta}) - \chi(\hat{\theta}_\psi)) \approx \psi_\theta(\hat{\theta}_\psi) \tilde{j}_{\theta\theta}^{-1}(\hat{\theta}_\psi) \psi'_\theta(\hat{\theta}_\psi) \frac{|\tilde{j}_{(\theta\theta)}(\hat{\theta}_\psi)|}{|\tilde{j}_{(\theta\theta)}(\hat{\theta})|} \quad (16)$$

where

$$|j_{\theta\theta}(\hat{\theta})| = |j_{\theta\theta}(\hat{\theta})| |\varphi_{\theta}(\hat{\theta})|^{-2} \quad \text{and} \quad |\tilde{j}_{\theta\theta}(\hat{\theta}_{\psi})| = |\tilde{j}_{\theta\theta}(\hat{\theta}_{\psi})| |\varphi_{\theta}(\hat{\theta}_{\psi})|^{-2}$$

with $j_{\theta\theta}(\theta)$ and $\tilde{j}_{\theta\theta}(\theta)$ being the observed information matrix obtained from the log-likelihood function and the tilted log-likelihood function respectively. Hence, the confidence interval of ψ based on $r^*(\psi)$ can be obtained from Equation (11).

4. Application to Stress-Strength Reliability with Independent EE Distributions

In Section 2 the EE distribution was introduced as was the reliability constraint. In Section 4.1, we consider the case where the scale parameters are different. The proposed penalized likelihood method is used to obtain the constrained maximum likelihood estimate. The likelihood-based third order method is then applied to obtain inference for $R = P(Y < X)$. Section 4.2 presents results from some numerical studies. The special case where the scale parameters are assumed to be equal, and hence, $R = P(Y < X)$ has a closed-form, is also examined.

4.1. Stress-Strength Reliability with Unequal Scale Parameters

Let X and Y be independently distributed as $EE(\alpha_1, \beta_1)$ and $EE(\alpha_2, \beta_2)$ respectively. $R(\theta)$ is given in (2) which does not have a closed form when $\beta_1 \neq \beta_2$. Following the steps in Section 3, we first obtain the overall maximum likelihood estimate $\hat{\theta} = (\hat{\alpha}_1, \hat{\alpha}_2, \hat{\beta}_1, \hat{\beta}_2)'$ and the constrained maximum likelihood estimate $\hat{\theta}_{\psi} = (\tilde{\alpha}_1, \tilde{\alpha}_2, \tilde{\beta}_1, \tilde{\beta}_2)'$ by the penalized likelihood method discussed in Section 2.5. Then observed information matrix $j_{\theta\theta}(\hat{\theta})$ can be obtained accordingly as follows.

$$j_{\theta\theta}(\hat{\theta}) = -\ell_{\theta\theta}(\hat{\theta}) = \begin{pmatrix} \frac{n}{\hat{\alpha}_1^2} & -\sum_{i=1}^n \frac{x_i e^{-\hat{\beta}_1 x_i}}{1 - e^{-\hat{\beta}_1 x_i}} & 0 & 0 \\ -\sum_{i=1}^n \frac{x_i e^{-\hat{\beta}_1 x_i}}{1 - e^{-\hat{\beta}_1 x_i}} & \frac{n}{\hat{\beta}_1^2} + A & 0 & 0 \\ 0 & 0 & \frac{m}{\hat{\alpha}_2^2} & -\sum_{j=1}^m \frac{y_j e^{-\hat{\beta}_2 y_j}}{1 - e^{-\hat{\beta}_2 y_j}} \\ 0 & 0 & -\sum_{j=1}^m \frac{y_j e^{-\hat{\beta}_2 y_j}}{1 - e^{-\hat{\beta}_2 y_j}} & \frac{m}{\hat{\beta}_2^2} + B \end{pmatrix}$$

where $A = (\hat{\alpha}_1 - 1) \sum_{i=1}^n \frac{x_i^2 e^{-\hat{\beta}_1 x_i}}{(1 - e^{-\hat{\beta}_1 x_i})^2}$ and $B = (\hat{\alpha}_2 - 1) \sum_{j=1}^m \frac{y_j^2 e^{-\hat{\beta}_2 y_j}}{(1 - e^{-\hat{\beta}_2 y_j})^2}$.

The tilted log-likelihood function $\tilde{l}(\theta)$ is defined as

$$\tilde{l}(\theta) = l(x, y; \alpha_1, \alpha_2, \beta_1, \beta_2) + \hat{K}[\psi(\theta) - \psi]$$

where $\psi(\theta) = R$ defined by (2). Similarly, we can obtain the constrained maximum likelihood estimate $\hat{\theta}_\psi = (\tilde{\alpha}_1, \tilde{\alpha}_2, \tilde{\beta}_1, \tilde{\beta}_2)'$ by the penalized likelihood method and \hat{K} is the Lagrange multiplier, then constrained observed information matrix $\tilde{j}_{\theta\theta}(\hat{\theta}_\psi)$ can be written as

$$\tilde{j}_{\theta\theta}(\hat{\theta}_\psi) = -\tilde{l}_{\theta\theta}(\hat{\theta}_\psi) = \begin{pmatrix} \tilde{j}_{\alpha_1\alpha_1}(\hat{\theta}_\psi) & \tilde{j}_{\alpha_1\alpha_2}(\hat{\theta}_\psi) & \tilde{j}_{\alpha_1\beta_1}(\hat{\theta}_\psi) & \tilde{j}_{\alpha_1\beta_2}(\hat{\theta}_\psi) \\ \tilde{j}_{\beta_1\alpha_1}(\hat{\theta}_\psi) & \tilde{j}_{\beta_1\alpha_2}(\hat{\theta}_\psi) & \tilde{j}_{\beta_1\beta_1}(\hat{\theta}_\psi) & \tilde{j}_{\beta_1\beta_2}(\hat{\theta}_\psi) \\ \tilde{j}_{\alpha_2\alpha_1}(\hat{\theta}_\psi) & \tilde{j}_{\alpha_2\alpha_2}(\hat{\theta}_\psi) & \tilde{j}_{\alpha_2\beta_1}(\hat{\theta}_\psi) & \tilde{j}_{\alpha_2\beta_2}(\hat{\theta}_\psi) \\ \tilde{j}_{\beta_2\alpha_1}(\hat{\theta}_\psi) & \tilde{j}_{\beta_2\alpha_2}(\hat{\theta}_\psi) & \tilde{j}_{\beta_2\beta_1}(\hat{\theta}_\psi) & \tilde{j}_{\beta_2\beta_2}(\hat{\theta}_\psi) \end{pmatrix}$$

where

- $\tilde{j}_{\alpha_1\alpha_1}(\hat{\theta}_\psi) = \frac{n}{\tilde{\alpha}_1^2} - \hat{K} R_{\alpha_1\alpha_1}(\hat{\theta}_\psi)$, where $R_{\alpha_1\alpha_1}(\hat{\theta}_\psi) = \frac{\partial^2 R(\theta)}{\partial \alpha_1^2} \Big|_{\theta=\hat{\theta}_\psi}$.
- $\tilde{j}_{\alpha_1\alpha_2}(\hat{\theta}_\psi) = -\hat{K} R_{\alpha_1\alpha_2}(\hat{\theta}_\psi)$, where $R_{\alpha_1\alpha_2}(\hat{\theta}_\psi) = \frac{\partial^2 R(\theta)}{\partial \alpha_1 \partial \alpha_2} \Big|_{\theta=\hat{\theta}_\psi}$.
- $\tilde{j}_{\alpha_1\beta_1}(\hat{\theta}_\psi) = -\sum_{i=1}^n \frac{x_i e^{-\tilde{\beta}_1 x_i}}{1 - e^{-\tilde{\beta}_1 x_i}} - \hat{K} R_{\alpha_1\beta_1}(\hat{\theta}_\psi)$, where $R_{\alpha_1\beta_1}(\hat{\theta}_\psi) = \frac{\partial^2 R(\theta)}{\partial \alpha_1 \partial \beta_1} \Big|_{\theta=\hat{\theta}_\psi}$.
- $\tilde{j}_{\alpha_1\beta_2}(\hat{\theta}_\psi) = -\hat{K} R_{\alpha_1\beta_2}(\hat{\theta}_\psi)$, where $R_{\alpha_1\beta_2}(\hat{\theta}_\psi) = \frac{\partial^2 R(\theta)}{\partial \alpha_1 \partial \beta_2} \Big|_{\theta=\hat{\theta}_\psi}$.
- $\tilde{j}_{\alpha_2\alpha_2}(\hat{\theta}_\psi) = \frac{m}{\tilde{\alpha}_2^2} - \hat{K} R_{\alpha_2\alpha_2}(\hat{\theta}_\psi)$, where $R_{\alpha_2\alpha_2}(\hat{\theta}_\psi) = \frac{\partial^2 R(\theta)}{\partial \alpha_2^2} \Big|_{\theta=\hat{\theta}_\psi}$.
- $\tilde{j}_{\alpha_2\beta_1}(\hat{\theta}_\psi) = -\hat{K} R_{\alpha_2\beta_1}(\hat{\theta}_\psi)$, where $R_{\alpha_2\beta_1}(\hat{\theta}_\psi) = \frac{\partial^2 R(\theta)}{\partial \alpha_2 \partial \beta_1} \Big|_{\theta=\hat{\theta}_\psi}$.
- $\tilde{j}_{\alpha_2\beta_2}(\hat{\theta}_\psi) = -\sum_{j=1}^m \frac{y_j e^{-\tilde{\beta}_2 y_j}}{1 - e^{-\tilde{\beta}_2 y_j}} - \hat{K} R_{\alpha_2\beta_2}(\hat{\theta}_\psi)$, where $R_{\alpha_2\beta_2}(\hat{\theta}_\psi) = \frac{\partial^2 R(\theta)}{\partial \alpha_2 \partial \beta_2} \Big|_{\theta=\hat{\theta}_\psi}$.
- $\tilde{j}_{\beta_1\beta_1}(\hat{\theta}_\psi) = \frac{m}{\tilde{\beta}_1^2} + (\tilde{\alpha}_1 - 1) \sum_{i=1}^n \frac{x_i^2 e^{-\tilde{\beta}_1 x_i}}{(1 - e^{-\tilde{\beta}_1 x_i})^2} - \hat{K} R_{\beta_1\beta_1}(\hat{\theta}_\psi)$, where $R_{\beta_1\beta_1}(\hat{\theta}_\psi) = \frac{\partial^2 R(\theta)}{\partial \beta_1^2} \Big|_{\theta=\hat{\theta}_\psi}$.
- $\tilde{j}_{\beta_1\beta_2}(\hat{\theta}_\psi) = -\hat{K} R_{\beta_1\beta_2}(\hat{\theta}_\psi)$, where $R_{\beta_1\beta_2}(\hat{\theta}_\psi) = \frac{\partial^2 R(\theta)}{\partial \beta_1 \partial \beta_2} \Big|_{\theta=\hat{\theta}_\psi}$.
- $\tilde{j}_{\beta_2\beta_2}(\hat{\theta}_\psi) = \frac{m}{\tilde{\beta}_2^2} + (\tilde{\alpha}_2 - 1) \sum_{j=1}^m \frac{y_j^2 e^{-\tilde{\beta}_2 y_j}}{(1 - e^{-\tilde{\beta}_2 y_j})^2} - \hat{K} R_{\beta_2\beta_2}(\hat{\theta}_\psi)$, where $R_{\beta_2\beta_2}(\hat{\theta}_\psi) = \frac{\partial^2 R(\theta)}{\partial \beta_2^2} \Big|_{\theta=\hat{\theta}_\psi}$.

Thus $r(\psi)$ can be obtained accordingly.

For the problem we are considering, the natural choice of the pivotal quantity is

$$\begin{aligned} z &= (z_1, \dots, z_n, z_{n+1}, \dots, z_{n+m})' \\ &= (\log F(x_1; \alpha_1, \beta), \dots, \log F(x_n; \alpha_1, \beta), \log F(y_1; \alpha_2, \beta), \dots, \log F(y_m; \alpha_2, \beta))'. \end{aligned}$$

Hence the ancillary direction V is

$$V = (V_1, V_2, V_3, V_4)$$

$$= \begin{pmatrix} -\log(1 - e^{-\hat{\beta}_1 x_1}) \frac{1 - e^{-\hat{\beta}_1 x_1}}{\hat{\alpha}_1 \hat{\beta}_1 e^{-\hat{\beta}_1 x_1}} & -\frac{x_1}{\hat{\beta}_1} & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ -\log(1 - e^{-\hat{\beta}_1 x_n}) \frac{1 - e^{-\hat{\beta}_1 x_n}}{\hat{\alpha}_1 \hat{\beta}_1 e^{-\hat{\beta}_1 x_n}} & -\frac{x_n}{\hat{\beta}_1} & 0 & 0 \\ 0 & 0 & -\log(1 - e^{-\hat{\beta}_2 y_1}) \frac{1 - e^{-\hat{\beta}_2 y_1}}{\hat{\alpha}_2 \hat{\beta}_2 e^{-\hat{\beta}_2 y_1}} & -\frac{y_1}{\hat{\beta}_2} \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & -\log(1 - e^{-\hat{\beta}_2 y_m}) \frac{1 - e^{-\hat{\beta}_2 y_m}}{\hat{\alpha}_2 \hat{\beta}_2 e^{-\hat{\beta}_2 y_m}} & -\frac{y_m}{\hat{\beta}_2} \end{pmatrix}$$

Then we can calculate the locally defined canonical parameter $\varphi(\theta)$ as

$$\varphi(\theta) = \left(\sum_{i=1}^{n+m} \frac{\partial l(\theta)}{\partial w_i} V_{1i}, \sum_{i=1}^{n+m} \frac{\partial l(\theta)}{\partial w_i} V_{2i}, \sum_{i=1}^{n+m} \frac{\partial l(\theta)}{\partial w_i} V_{3i}, \sum_{i=1}^{n+m} \frac{\partial l(\theta)}{\partial w_i} V_{4i} \right)'$$

where $w = (x_1, \dots, x_n, y_1, \dots, y_m)'$ be the observed data. Hence, we also have $\varphi_\theta(\theta)$. Therefore, for this unequal scale parameter case, $\chi(\theta)$, $\widehat{var}(\chi(\hat{\theta}) - \chi(\hat{\theta}_\psi))$, $Q(\psi)$ and $r^*(\psi)$ can be obtained accordingly. Hence, the $(1 - \gamma)100\%$ confidence interval can be obtained from the modified signed log-likelihood ratio statistics.

4.2. Numerical Examples

To illustrate the proposed third-order method for interval estimation, the following two data sets with sample size of 11 and 9 were used: $x = (2.1828, 0.5911, 1.0711, 0.9007, 1.7814, 1.3616, 0.8629, 0.2301, 1.5183, 0.8481, 1.0845)$ and $y = (0.8874, 1.1482, 0.8227, 0.4086, 0.5596, 1.1978, 1.1324, 0.5625, 1.0679)$. By assumption, X and Y are independently distributed as $EE(\alpha_1, \beta_1)$ and $EE(\alpha_2, \beta_2)$ distributions respectively. We are interested in testing $H_0 : R = 0.5$ vs. $H_1 : R > 0.5$, where $R = P(Y < X)$ is given in (2). Table 3 present the 90% and 95% confidence intervals (CI) for R based on the maximum likelihood method (MLE), the signed log-likelihood ratio statistic method (r) and the proposed third-order method (Proposed). In examining Table 3, we observe that all three methods give different interval estimates.

Table 3. Interval Estimates of ψ for Example.

	90% Confidence Interval	95% Confidence Interval
MLE	(0.4223, 0.8179)	(0.3843, 0.8557)
r	(0.4151, 0.7966)	(0.3767, 0.8241)
Proposed	(0.4080, 0.7910)	(0.3698, 0.8188)

When the scale parameters is the same, the value of R can be expressed in closed form as $\frac{\alpha_1}{\alpha_1 + \alpha_2}$. This simplifies the analysis. Following the same procedures introduced in Section 3, $\widehat{var}(\chi(\hat{\theta}) - \chi(\hat{\theta}_\psi))$, $Q(\psi)$ and $r^*(\psi)$ can be obtained.

To compare the accuracy of the three methods discussed in this paper, Monte Carlo simulation studies were conducted. The cases of unequal and equal scale parameters are both examined. For each parameter configuration and for each sample size, we generate 10,000 random samples from the EE distributions by using the following transformation:

$$T = -\frac{1}{\beta} \log(1 - U^{1/\alpha})$$

where U is a uniform variate between 0 and 1.

For each simulated setting, we report the proportion of ψ that fall outside the lower bound of the confidence interval (lower error), the proportion of ψ that fall outside the upper bound of the confidence interval (upper error), the proportion of ψ that fall within the confidence interval (central coverage), and the average bias (Average Bias), which is defined as

$$\text{Average Bias} = \frac{|\text{lower error} - 0.025| + |\text{upper error} - 0.025|}{2}$$

The nominal values for the lower and the upper errors, the central coverage and the average bias are 0.025, 0.025, 0.95 and 0 respectively. These values reflect the desired properties of the accuracy and symmetry of the interval estimates of ψ .

Tables 4–6 present simulation results for the unequal scale parameters case, *i.e.*, $\alpha_1 = 2$, $\alpha_2 = 5$, $\beta_1 = 3$, and $R = 0.1(0.1)0.9$ with $(n, m) = (10, 10)$, $(10, 50)$ and $(50, 10)$. Note that, we fixed R and β_2 is determined uniquely by Equation (2). It is clear that the coverage probabilities for R are poor and the two-tail error probabilities are extremely asymmetric from the MLE method. Results from the signed log-likelihood method are not satisfactory especially when the two sample sizes are small or sample sizes are unequal, and it also shows some evidence of asymmetry of two-tail error probabilities. However, the proposed method gives not only an almost exact coverage probability but also it has symmetric two-tail error probabilities even for small or uneven sample sizes.

Note that if we had used the built-in macros / subroutines from Matlab or R, our simulation study could not have been completed because of difficulties in obtaining the constrained maximum likelihood estimates. Fortunately, we did not encounter any difficulties when the penalized likelihood method was used.

Tables 7–9 present simulation results for the equal scale parameter case, *i.e.*, $\alpha_1 = 4$, $\beta = 8$, and $R = 0.1(0.1)0.9$ with $(n, m) = (10, 10)$, $(10, 50)$ and $(50, 10)$. In estimation, we fixed R and α_2 is then determined uniquely by $R = \alpha_1/(\alpha_1 + \alpha_2)$. Again, the proposed method outperformed the other two methods even when the sample sizes were small. Other simulation results, though not reported here, essentially corroborate those of Tables 7–9, and are available upon request. In the equal scale parameter case, when sample sizes were small, or when R was close to the boundary values, we used the penalized likelihood method to obtain the constrained maximum likelihood estimates. More simulations have been performed with the same pattern of results. They are not reported here, but are available on request.

Table 4. $\beta_1 \neq \beta_2$, $\alpha_1 = 2, \alpha_2 = 5, \beta_1 = 3$ and β_2 is obtained by Equation (2), $(n, m) = (10, 10)$.

<i>R</i>	Method	Lower Error	Upper Error	Central Coverage	Average Bias
0.1	MLE	0.1602	0.0033	0.8365	0.07845
	<i>r</i>	0.0401	0.0177	0.9422	0.01120
	Proposed	0.0207	0.0252	0.9541	0.00255
0.2	MLE	0.1138	0.0125	0.8737	0.05065
	<i>r</i>	0.0388	0.0235	0.9377	0.00765
	Proposed	0.0218	0.0258	0.9524	0.00200
0.3	MLE	0.0857	0.0225	0.8918	0.03160
	<i>r</i>	0.0372	0.0262	0.9366	0.00670
	Proposed	0.0230	0.0259	0.9527	0.00135
0.4	MLE	0.0656	0.0362	0.8982	0.02590
	<i>r</i>	0.0352	0.0294	0.9354	0.00730
	Proposed	0.0244	0.0259	0.9497	0.00075
0.5	MLE	0.0505	0.0506	0.8989	0.02555
	<i>r</i>	0.0317	0.0328	0.9355	0.00725
	Proposed	0.0249	0.0255	0.9496	0.00030
0.6	MLE	0.0353	0.0670	0.8977	0.02615
	<i>r</i>	0.0290	0.0359	0.9351	0.00745
	Proposed	0.0244	0.0246	0.9510	0.00050
0.7	MLE	0.0235	0.0900	0.8865	0.03325
	<i>r</i>	0.0257	0.0394	0.9349	0.00755
	Proposed	0.0246	0.0238	0.9516	0.00080
0.8	MLE	0.0142	0.1234	0.8624	0.05460
	<i>r</i>	0.0239	0.0419	0.9342	0.00900
	Proposed	0.0261	0.0239	0.9500	0.00110
0.9	MLE	0.0035	0.1763	0.8202	0.08640
	<i>r</i>	0.0198	0.0465	0.9337	0.01335
	Proposed	0.0262	0.0240	0.9498	0.00110

Table 5. $\beta_1 \neq \beta_2$, $\alpha_1 = 2, \alpha_2 = 5, \beta_1 = 3$ and β_2 is obtained by Equation (2), $(n, m) = (10, 50)$.

<i>R</i>	Method	Lower Error	Upper Error	Central Coverage	Average Bias
0.1	MLE	0.1341	0.0046	0.8613	0.06475
	<i>r</i>	0.0399	0.0186	0.9415	0.01065
	Proposed	0.0231	0.0246	0.9523	0.00115
0.2	MLE	0.1007	0.0131	0.8862	0.04380
	<i>r</i>	0.0370	0.0243	0.9387	0.00635
	Proposed	0.0239	0.0251	0.9510	0.00060
0.3	MLE	0.0774	0.0249	0.8977	0.02625
	<i>r</i>	0.0349	0.0289	0.9362	0.00690
	Proposed	0.0228	0.0270	0.9502	0.00210
0.4	MLE	0.0615	0.0353	0.9032	0.02340
	<i>r</i>	0.0327	0.0311	0.9362	0.00690
	Proposed	0.0220	0.0260	0.9520	0.00200
0.5	MLE	0.0475	0.0488	0.9037	0.02315
	<i>r</i>	0.0294	0.0323	0.9383	0.00585
	Proposed	0.0229	0.0240	0.9531	0.00155
0.6	MLE	0.0351	0.0682	0.8967	0.02665
	<i>r</i>	0.0278	0.0348	0.9374	0.00630
	Proposed	0.0222	0.0222	0.9556	0.00280
0.7	MLE	0.0225	0.0962	0.8813	0.03685
	<i>r</i>	0.0256	0.0388	0.9356	0.00720
	Proposed	0.0226	0.0227	0.9547	0.00235
0.8	MLE	0.0126	0.1249	0.8625	0.05615
	<i>r</i>	0.0217	0.0431	0.9352	0.01070
	Proposed	0.0224	0.0242	0.9534	0.00170
0.9	MLE	0.0063	0.1779	0.8158	0.08580
	<i>r</i>	0.0168	0.0473	0.9359	0.01525
	Proposed	0.0215	0.0238	0.9547	0.00235

Table 6. $\beta_1 \neq \beta_2$, $\alpha_1 = 2, \alpha_2 = 5, \beta_1 = 3$ and β_2 is obtained by Equation (2), $(n, m) = (50, 10)$.

<i>R</i>	Method	Lower Error	Upper Error	Central Coverage	Average Bias
0.1	MLE	0.1257	0.0046	0.8697	0.06055
	<i>r</i>	0.0392	0.0167	0.944115	0.01125
	Proposed	0.0207	0.0203	0.9590	0.00450
0.2	MLE	0.0885	0.0154	0.8961	0.03655
	<i>r</i>	0.0347	0.0223	0.9430	0.00620
	Proposed	0.0226	0.0226	0.9548	0.00240
0.3	MLE	0.0657	0.0234	0.9109	0.02115
	<i>r</i>	0.0332	0.0252	0.9416	0.00420
	Proposed	0.0221	0.0233	0.9546	0.002130
0.4	MLE	0.0497	0.0335	0.9168	0.01660
	<i>r</i>	0.0317	0.0286	0.9397	0.00515
	Proposed	0.0228	0.0241	0.9531	0.00155
0.5	MLE	0.0368	0.0428	0.9204	0.01480
	<i>r</i>	0.0285	0.0309	0.9405	0.00475
	Proposed	0.0222	0.0236	0.9542	0.00210
0.6	MLE	0.0264	0.0499	0.9237	0.01315
	<i>r</i>	0.0248	0.034830	0.9422	0.00410
	Proposed	0.0215	0.0231	0.9554	0.00270
0.7	MLE	0.0184	0.0595	0.9221	0.02055
	<i>r</i>	0.0222	0.0332	0.9446	0.00550
	Proposed	0.0206	0.0235	0.9559	0.00295
0.8	MLE	0.0112	0.0715	0.9173	0.03015
	<i>r</i>	0.0186	0.0324	0.9490	0.00690
	Proposed	0.0196	0.0212	0.9592	0.00460
0.9	MLE	0.0055	0.0927	0.9018	0.04360
	<i>r</i>	0.0149	0.0291	0.9560	0.00710
	Proposed	0.0183	0.0209	0.9608	0.00540

Table 7. $\alpha_1 = 4, \beta = 8$ and α_2 satisfies $R = \alpha_1/(\alpha_1 + \alpha_2)$, $(n, m) = (10, 10)$.

R	Method	Lower Error	Upper Error	Central Coverage	Average Bias
0.1	MLE	0.1608	0.0024	0.8368	0.07920
	r	0.0526	0.0189	0.9285	0.01685
	Proposed	0.0236	0.0267	0.9497	0.00155
0.2	MLE	0.1195	0.0121	0.8684	0.05370
	r	0.0502	0.0236	0.9262	0.01330
	Proposed	0.0269	0.0264	0.9467	0.00165
0.3	MLE	0.0944	0.0224	0.8832	0.03600
	r	0.0428	0.0256	0.9316	0.00920
	Proposed	0.0225	0.0234	0.9541	0.00205
0.4	MLE	0.0724	0.0347	0.8929	0.02855
	r	0.0387	0.0277	0.9336	0.00820
	Proposed	0.0256	0.0224	0.9520	0.00160
0.5	MLE	0.0523	0.0517	0.8960	0.02700
	r	0.0335	0.0328	0.9337	0.00815
	Proposed	0.0244	0.0230	0.9526	0.00013
0.6	MLE	0.0378	0.0753	0.8869	0.03155
	r	0.0295	0.0400	0.9305	0.00975
	Proposed	0.0234	0.0260	0.9506	0.00130
0.7	MLE	0.0245	0.0944	0.8811	0.03495
	r	0.0282	0.0454	0.9264	0.01180
	Proposed	0.0261	0.0262	0.9477	0.00115
0.8	MLE	0.0109	0.1187	0.8704	0.05390
	r	0.0226	0.0467	0.9307	0.01205
	Proposed	0.0239	0.0262	0.9499	0.00115
0.9	MLE	0.0027	0.1464	0.8509	0.07185
	r	0.0211	0.0499	0.9290	0.01440
	Proposed	0.0268	0.0252	0.9480	0.00100

Table 8. $\alpha_1 = 4, \beta = 8$ and α_2 satisfies $R = \alpha_1/(\alpha_1 + \alpha_2)$, $(n, m) = (10, 50)$.

R	Method	Lower Error	Upper Error	Central Coverage	Average Bias
0.1	MLE	0.0722	0.0093	0.9185	0.03145
	r	0.0302	0.0273	0.9425	0.00375
	Proposed	0.0250	0.0261	0.9489	0.00055
0.2	MLE	0.0577	0.0206	0.9217	0.01855
	r	0.0287	0.0268	0.9445	0.00275
	Proposed	0.0260	0.0244	0.9496	0.00080
0.3	MLE	0.0431	0.0327	0.9242	0.01290
	r	0.0276	0.0306	0.9418	0.00410
	Proposed	0.0261	0.0257	0.9482	0.00090
0.4	MLE	0.0299	0.0412	0.9289	0.01055
	r	0.0216	0.0290	0.9494	0.00370
	Proposed	0.0218	0.0227	0.9555	0.00275
0.5	MLE	0.0475	0.0488	0.9037	0.02315
	r	0.0234	0.0358	0.9408	0.00620
	Proposed	0.0243	0.0269	0.9488	0.00130
0.6	MLE	0.0166	0.0688	0.9146	0.02610
	r	0.0221	0.0350	0.9429	0.00645
	Proposed	0.0249	0.0254	0.9497	0.00250
0.7	MLE	0.0105	0.0789	0.9106	0.03420
	r	0.0206	0.0366	0.9428	0.00800
	Proposed	0.0276	0.0229	0.9495	0.00235
0.8	MLE	0.0059	0.1021	0.8920	0.04810
	r	0.0195	0.0426	0.9379	0.01155
	Proposed	0.0259	0.0269	0.9472	0.00140
0.9	MLE	0.0016	0.1210	0.8774	0.05970
	r	0.0160	0.0456	0.9384	0.01480
	Proposed	0.0239	0.0249	0.9512	0.00060

Table 9. $\alpha_1 = 4, \beta = 8$ and α_2 satisfies $R = \alpha_1/(\alpha_1 + \alpha_2)$, $(n, m) = (50, 10)$.

R	Method	Lower Error	Upper Error	Central Coverage	Average Bias
0.1	MLE	0.1120	0.0008	0.8872	0.05560
	r	0.0412	0.0161	0.9427	0.01255
	Proposed	0.0230	0.0247	0.9523	0.00115
0.2	MLE	0.1011	0.0048	0.8941	0.04815
	r	0.0407	0.0202	0.9319	0.01025
	Proposed	0.0261	0.0260	0.9479	0.00105
0.3	MLE	0.0805	0.0108	0.9087	0.03485
	r	0.0366	0.0232	0.9402	0.00670
	Proposed	0.0248	0.0275	0.9477	0.00135
0.4	MLE	0.0648	0.0140	0.9212	0.02540
	r	0.0338	0.0215	0.9447	0.00615
	Proposed	0.0238	0.0247	0.9515	0.00075
0.5	MLE	0.0551	0.0258	0.9191	0.01545
	r	0.0317	0.0256	0.9427	0.00365
	Proposed	0.0242	0.0276	0.9482	0.00170
0.6	MLE	0.0437	0.0296	0.9267	0.01165
	r	0.0329	0.0228	0.9443	0.00505
	Proposed	0.0247	0.0229	0.9524	0.00120
0.7	MLE	0.0332	0.0400	0.9268	0.01160
	r	0.0307	0.0247	0.9446	0.00300
	Proposed	0.0257	0.0241	0.9502	0.00080
0.8	MLE	0.0228	0.0548	0.9224	0.01600
	r	0.0294	0.0266	0.9440	0.00300
	Proposed	0.0256	0.0242	0.9502	0.00070
0.9	MLE	0.0098	0.0670	0.9232	0.02860
	r	0.0247	0.0299	0.9454	0.00260
	Proposed	0.0234	0.0261	0.9505	0.00135

5. Conclusions

A penalized likelihood method is proposed to overcome numerical difficulties that arose in using the Lagrange method to solve an optimization problem with an integral constraint. The proposed method is used to obtain constrained maximum likelihood estimates for the stress-strength reliability with independent exponentiated exponential distributions. In turn, these estimates are used to obtain inference for the stress-strength reliability via a likelihood-based third order asymptotic method. In our simulation studies, the penalized likelihood method did not encounter any numerical difficulties in obtaining the constrained maximum likelihood estimates. Moreover, the likelihood-based third order asymptotic method gives very accurate results even when the sample sizes are small. Although the paper is restricted to independent exponentiated exponential distributions, it can easily be extended to other distributions.

Acknowledgments

The authors would like to thank two anonymous referees for their very helpful comments and suggestions. The corresponding author gratefully acknowledged the support of NSERC through his Discovery Grant.

Author Contributions

The paper is the result of a joint effort by all of the authors, who contributed in equal ways to conceiving of the idea, carrying out all numerical calculations and writing the paper. All authors have read and approved the final manuscript.

Appendix

In this appendix, we derive the properties of the unconstrained likelihood function that we stated in Section 2 of the paper. Some or all of these results may be known in the literature but we have not found such a source.

Let $x = (x_1, \dots, x_n)'$ and $y = (y_1, \dots, y_m)'$ be independent random samples from $EE(\alpha_1, \beta_1)$ and $EE(\alpha_2, \beta_2)$ respectively. Then the log-likelihood function is:

$$\begin{aligned} l(\theta; x, y) &= l(\alpha_1, \beta_1, \alpha_2, \beta_2; x, y) \\ &= n \log \alpha_1 + n \log \beta_1 + (\alpha_1 - 1) \sum_{i=1}^n \log(1 - e^{-\beta_1 x_i}) - \beta_1 \sum_{i=1}^n x_i \\ &\quad + m \log \alpha_2 + m \log \beta_2 + (\alpha_2 - 1) \sum_{j=1}^m \log(1 - e^{-\beta_2 y_j}) - \beta_2 \sum_{j=1}^m y_j \end{aligned} \quad (17)$$

where $\theta = (\alpha_1, \alpha_2, \beta_1, \beta_2)'$. We note that, because of independence of X and Y , the log-likelihood function is the sum of two functions that are identical except for the parameter and variable names.

Parameters of one function do not enter the second function. For that reason, we limit our analysis to the generic log-likelihood function:

$$l(\alpha, \beta; x) = n \log \alpha + n \log \beta + (\alpha - 1) \sum_{i=1}^n \log(1 - e^{-\beta x_i}) - \beta \sum_{i=1}^n x_i \quad (18)$$

The first and second partial derivatives are given, using subscript notation, as:

$$l_{\alpha} = \frac{n}{\alpha} + \sum_{i=1}^n \log(1 - e^{-\beta x_i}) \quad (19)$$

$$l_{\beta} = \frac{n}{\beta} + (\alpha - 1) \sum_{i=1}^n \frac{x_i e^{-\beta x_i}}{(1 - e^{-\beta x_i})} - \sum_{i=1}^n x_i \quad (20)$$

$$l_{\alpha\alpha} = -\frac{n}{\alpha^2} \quad (21)$$

$$l_{\beta\beta} = -\frac{n}{\beta^2} - (\alpha - 1) \sum_{i=1}^n \frac{x_i^2 e^{-\beta x_i}}{(1 - e^{-\beta x_i})^2} \quad (22)$$

$$l_{\alpha\beta} = \sum_{i=1}^n \frac{x_i e^{-\beta x_i}}{(1 - e^{-\beta x_i})} \quad (23)$$

It is clear from (21) that $l_{\alpha\alpha} < 0$ as stated in the Section 2. In order to establish that $l_{\beta\beta} < 0$, it is convenient to rewrite Equation (22) as:

$$l_{\beta\beta} = -\frac{n}{\beta^2} + \sum_{i=1}^n \frac{x_i^2 e^{-\beta x_i}}{(1 - e^{-\beta x_i})^2} - \alpha \sum_{i=1}^n \frac{x_i^2 e^{-\beta x_i}}{(1 - e^{-\beta x_i})^2} \quad (24)$$

The first and third terms in Equation (24) are negative but the second term is positive. However, it is possible to show that the the sum of the first two terms is negative. Note first that $n = \frac{(1 - e^{-\beta x_1})^2}{(1 - e^{-\beta x_1})^2} + \dots + \frac{(1 - e^{-\beta x_n})^2}{(1 - e^{-\beta x_n})^2}$. This allows us to rewrite Equation (24) as:

$$l_{\beta\beta} = \frac{1}{\beta^2} \left(\sum_{i=1}^n \frac{\beta^2 x_i^2 e^{-\beta x_i} - (1 - e^{-\beta x_i})^2}{(1 - e^{-\beta x_i})^2} \right) - \alpha \sum_{i=1}^n \frac{x_i^2 e^{-\beta x_i}}{(1 - e^{-\beta x_i})^2} \quad (25)$$

A representative numerator term from the first expression is: $\beta^2 x_i^2 e^{-\beta x_i} - (1 - e^{-\beta x_i})^2$ which is the difference of squares and can be re-expressed as:

$$\beta^2 x_i^2 e^{-\beta x_i} - (1 - e^{-\beta x_i})^2 = (\beta x_i e^{-.5\beta x_i} - (1 - e^{-\beta x_i}))(\beta x_i e^{-.5\beta x_i} + (1 - e^{-\beta x_i}))$$

The second product term on the right hand side is the sum of two positive terms and is therefore positive. The first product term is always negative when $\beta > 0$ and $x_i > 0$. This is because

$$\beta x_i e^{-.5\beta x_i} - (1 - e^{-\beta x_i}) = e^{-.5\beta x_i} [\beta x_i - (e^{.5\beta x_i} - e^{-.5\beta x_i})]$$

and using a Taylor expansion of $\beta x_i - (e^{.5\beta x_i} - e^{-.5\beta x_i}) < 0$ with $\beta > 0$ and $x_i > 0$, we see that $\beta x_i e^{-.5\beta x_i} - (1 - e^{-\beta x_i}) < 0$. This completes the demonstration that $l_{\beta\beta} < 0$ whenever $\beta > 0$ and $x_i > 0$.

The log-likelihood function, $l(\alpha, \beta; x)$, is not a concave function of (α, β) over the domain of the function. However, it is possible to show that when $n > 2$, there is a unique maximum of $l(\alpha, \beta; x)$ that occurs where $l_\alpha = 0$ and $l_\beta = 0$. We begin by showing that there is a unique solution to $l_\alpha = 0$ and $l_\beta = 0$.

Equation (19) can be solved for α as:

$$\alpha(\beta) = \frac{-n}{\sum_{i=1}^n \log(1 - e^{-\beta x_i})} \quad (26)$$

From (26) it is clear that $\alpha(\beta)$ is monotone increasing and that $\lim_{\beta \rightarrow 0} \alpha(\beta) = 0$. Substituting $\alpha(\beta)$ into Equation (20) yields:

$$\frac{n}{\beta} + \left(\frac{-n}{\sum_{i=1}^n \log(1 - e^{-\beta x_i})} \right) \sum_{i=1}^n \frac{x_i e^{-\beta x_i}}{(1 - e^{-\beta x_i})} - \sum_{i=1}^n \frac{x_i e^{-\beta x_i}}{(1 - e^{-\beta x_i})} - \sum_{i=1}^n x_i = 0 \quad (27)$$

The first three terms on the left hand side define a function of β , call it $lhs(\beta)$, with the property that $lhs(\beta)$ is monotone decreasing with $\lim_{\beta \rightarrow 0} lhs(\beta) = +\infty$ and $\lim_{\beta \rightarrow \infty} lhs(\beta) = 0$. The fact that $\lim_{\beta \rightarrow \infty} lhs(\beta) = 0$ is clear by inspection. As well, it is clear that the second term on the left hand side approaches $+\infty$ as $\beta \rightarrow 0$. What is less obvious is that the sum of the first and third terms limits to $.5 \sum_{i=1}^n x_i$ as $\beta \rightarrow 0$. However, this can be established by writing the sum of the first and third terms as: $\frac{1}{\beta} (n - \sum_{i=1}^n \frac{\beta x_i e^{-\beta x_i}}{(1 - e^{-\beta x_i})})$. If we make the substitution: $n = \frac{(1 - e^{-\beta x_1})}{(1 - e^{-\beta x_1})} + \dots + \frac{(1 - e^{-\beta x_n})}{(1 - e^{-\beta x_n})}$ we obtain: $\frac{1}{\beta} (n - \sum_{i=1}^n \frac{\beta x_i e^{-\beta x_i}}{(1 - e^{-\beta x_i})}) = \frac{1}{\beta} (\sum_{i=1}^n \frac{1 - e^{-\beta x_i} - \beta x_i e^{-\beta x_i}}{(1 - e^{-\beta x_i})})$. The result above follows when the exponential terms are expressed as series and the limit is taken as $\beta \rightarrow 0$. Alternatively, l'Hopital's rule can be applied to each term of the sum.

Given that $\sum_{i=1}^n x_i > 0$ because X is a positive random variable, we see that there is a unique solution to Equation (27).

The final task in this Appendix is to show that the unique point $(\alpha, \beta)^*$, where the gradient vanishes, is the global maximum of the likelihood function. This can be accomplished by showing that the determinant of the Hessian matrix of $l(\alpha, \beta; x)$ is strictly positive at $(\alpha, \beta)^*$. The determinant of the Hessian matrix, call it D , can be expressed equivalently as:

$$D = -A_1 - A_2 \quad (28)$$

where:

$$A_1 = -\frac{n}{\beta^2} + \sum_{i=1}^n \frac{x_i^2 e^{-\beta x_i}}{(1 - e^{-\beta x_i})^2} \quad (29)$$

and

$$A_2 = B \sum_{i=1}^n \log(1 - e^{-\beta x_i}) \sum_{i=1}^n \frac{x_i^2 e^{-\beta x_i}}{(1 - e^{-\beta x_i})^2} + B \left(\sum_{i=1}^n \frac{x_i e^{-\beta x_i}}{(1 - e^{-\beta x_i})} \right)^2 \quad (30)$$

with $B = \frac{n}{(\sum_{i=1}^n \log(1 - e^{-\beta x_i}))^2} > 0$. To establish that $D > 0$ it is sufficient to show that $A_1 < 0$ and $\frac{A_2}{B} < 0$. Note that the first two terms in Equation (24) are exactly the same as A_1 and we proved the sum of those terms was negative in the discussion following Equation (24). It remains only to prove that:

$$\sum_{i=1}^n \log(1 - e^{-\beta x_i}) \sum_{i=1}^n \frac{x_i^2 e^{-\beta x_i}}{(1 - e^{-\beta x_i})^2} + \left(\sum_{i=1}^n \frac{x_i e^{-\beta x_i}}{(1 - e^{-\beta x_i})} \right)^2 < 0 \quad (31)$$

Recalling the inequality $\log(s) < s - 1$ and therefore that $\log(1 - e^{-t}) < -e^{-t} < 0$ for all $t > 0$, we replace $\sum_{i=1}^n \log(1 - e^{-\beta_1 x_i})$ by $-\sum_{i=1}^n e^{-\beta_1 x_i}$ and obtain:

$$-\sum_{i=1}^n e^{-\beta x_i} \sum_{i=1}^n \frac{x_i^2 e^{-\beta x_i}}{(1 - e^{-\beta x_i})^2} + \left(\sum_{i=1}^n \frac{x_i e^{-\beta x_i}}{(1 - e^{-\beta x_i})} \right)^2 < 0 \quad (32)$$

and note that the inequality in Equation (31) will hold whenever the inequality in Equation (32) holds. The Cauchy-Schwarz inequality establishes that $(\sum_{i=1}^n u_i v_i)^2 \leq \sum_{i=1}^n u_i^2 \sum_{i=1}^n v_i^2$. Equality holds only when $\frac{u_i}{v_i} = K$ for all i where K is a constant different from 0 or where $u_i = 0$ and/or $v_i = 0$ for all i . Setting $u_i = e^{-.5\beta x_i}$ and $v_i = \frac{x_i e^{-.5\beta x_i}}{(1 - e^{-\beta x_i})}$, we see that the Cauchy-Schwarz inequality establishes the strict inequality in Equation (32) and therefore the determinant of the Hessian matrix of the log-likelihood function is positive. In fact, we have proved a stronger result. Suppose that we concentrate the log-likelihood function by substituting $\alpha(\beta)$ from Equation (26) into Equation (18). This function shows how the log-likelihood varies with β assuming that α is always chosen according to $l_\alpha = 0$ as in Equation (26). The second derivative of the concentrated log-likelihood function with respect to β is exactly equal to the negative of the determinant of the Hessian of the log-likelihood function. We have proved that this is negative without requiring that β must satisfy $l_\beta = 0$ as in Equation (27). As such, the concentrated log-likelihood function must be strictly concave as a function of β . We therefore have established a straightforward algorithm to get to the maximum of the log-likelihood function. Use a simple line search to find the unique β^* that satisfies Equation (27). This is guaranteed to converge. The corresponding unique α^* is then obtained from Equation (26). Alternatively, a search algorithm using second derivative information from the log-likelihood function could fail. Yee [19] notes possible problems in estimating EE models using the maximum likelihood function `expexp()` in the VGAM package in R.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Bertsekas, D.P. *Constrained Optimization and Lagrange Multiplier Methods*; Athena Scientific: Belmont, MA, USA, 1996.
2. Byrne, C.L. Sequential Unconstrained Minimization: A Survey. Available online: <http://faculty.uml.edu/cbyrne/SUM.pdf> (accessed on 11 June 2015).
3. Smith, J.B.; Wong, A.; Zhou, X. Higher order inference for stress-strength reliability with independent Burr-type X distributions. *J. Stat. Comput. Simul.* **2014**, doi:10.1080/00949655.2014.951359.
4. Gupta, R.D.; Kundu, D. Exponentiated exponential family: An alternative to gamma and Weibull distributions. *Biom. J.* **2001**, *43*, 117–130.
5. Mudholkar, G.S.; Srivastava, D.K. Exponentiated Weibull family for analyzing bathtub failure-rate data. *IEEE Trans. Reliab.* **1993**, *42*, 299–302.
6. Kundu, D.; Gupta, R.D. Estimation of $P[Y < X]$ for generalized exponential distribution. *Metrika* **2005**, *61*, 291–308.

7. Raqab, M.Z.; Madi, M.T.; Kundu, D. Estimation of $P[Y < X]$ for the three-parameter generalized exponential distribution. *Commun. Stat. -Theory Methods* **2008**, *37*, 2854–2864.
8. Fraser, D.A.S.; Reid, N. Ancillaries and third order significance. *Util. Math.* **1995**, *47*, 33–35.
9. Chang, F.; Wong, A. Improved likelihood-based inference for the stationary AR(2) model. *J. Stat. Plan. Inference* **2010**, *140*, 2099–2110.
10. She, Y.; Wong, A.; Zhou, X. Revisit the two sample t -test with a known ratio of variances. *Open J. Stat.* **2011**, *1*, 151–156.
11. Byrd, R.H.; Lu, P.; Nocedal, J.; Zhu, C. A limited memory algorithm for bound constrained optimization. *SIAM J. Sci. Comput.* **1995**, *16*, 1190–1208.
12. Cox, D.R.; Hinkley, D.V. *Theoretical Statistics*; Chapman and Hall: London, UK, 1974.
13. Doganaksoy, N.; Schmee, J. Comparisons of approximate confidence intervals for distributions used in life-data analysis. *Technometrics* **1993**, *35*, 175–184.
14. Reid, N. Likelihood and higher-order approximations to tail areas: A review and annotated bibliography. *Can. J. Stat.* **1996**, *24*, 141–166.
15. Severeni, T. *Likelihood Methods in Statistics*; Oxford University Press: New York, NY, USA, 2000.
16. Barndorff-Nielsen, O.E. Inference on full or partial parameters based on the standardized signed log likelihood ratio. *Biometrika* **1986**, *73*, 307–322.
17. Barndorff-Nielsen, O.E. Modified signed log likelihood ratio. *Biometrika* **1991**, *78*, 557–563.
18. Fraser, D.A.S.; Reid, N.; Wu, J. A simple general formula for tail probabilities for frequentist and Bayesian inference. *Biometrika* **1999**, *86*, 249–264.
19. Yee, T.W. *Vector Generalized Linear and Additive Models: With an Implementation in R*; Springer: Berlin/Heidelberg, Germany, 2015.

© 2015 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).