Interval estimation of the stress-strength reliability with independent normal random variables

Pierre Nguimkeu∗ Marie Rekkas† Augustine Wong‡

Abstract

This paper develops a procedure to obtain highly accurate confidence interval estimates for the stress-strength reliability \( R = P(X > Y) \) where \( X \) and \( Y \) are data from independent normal distributions of unknown means and variances. Our method is based on third-order likelihood analysis and is compared to the conventional first-order likelihood ratio procedure as well as the approximate methods of Reiser & Guttman (1986) and Guo & Krishnamoorthy (2004). The use of our proposed method is illustrated by an empirical example and its superior accuracy in terms of coverage probability and error rate are examined through Monte Carlo simulation studies.

Keywords: Reliability; Likelihood analysis; Interval estimation; Normal distribution.

∗Department of Economics, Simon Fraser University, 8888 University Drive, Burnaby, British Columbia, Canada V5A 1S6, email: npn1@sfu.ca.
†Corresponding Author. Department of Economics, Simon Fraser University, 8888 University Drive, Burnaby, British Columbia, Canada V5A 1S6, email: mrekkas@sfu.ca, phone: (778) 782-6793, fax: (778) 782-5944
‡Department of Mathematics and Statistics, York University, 4700 Keele Street, Toronto, Ontario, Canada M3J 1P3, email: august@yorku.ca
1 Introduction

Consider two independent random variables $X$ and $Y$. The functional $R = P(X > Y)$, usually referred to as the stress-strength reliability, is of practical importance in many fields including medicine, biostatistics, genetics (Schwartz & Wearden 1959), engineering, reliability (Harris & Soms 1983), psychology, and quality control (Kotz, Lumelskii & Pensky 2003). For example, in clinical studies, if $X$ is a response of a treatment group and $Y$ is the response of the control group, then $R$ measures the effectiveness of the treatment. In a reliability study, where $X$ is the strength of a system and $Y$ is the stress applied to the system, $1 - R$ measures the chance of failure of that system.

This paper considers statistical inference for $R = P(X > Y)$ when the observed data $X$ and $Y$ are drawn from independent and unknown normal distributions. Simonoff, Hochberg & Reiser (1986) explained that this is one of the most common situations encountered in parametric estimation of reliability. Inference about $R$ in this setup has been considered by many authors. The approximate method of Reiser & Guttman (1987) is a generalization of the methods proposed by Enis & Geisser (1971), Church & Harris (1970) and Mazumbar (1970) in several respects. Weerahandi & Johnson (1992) proposed inferential procedures for $R$ based on the generalized p-value approach while Guo & Krishnamoorthy (2004) constructed lower confidence limits for $R$ based on the approximate procedure of Hall (1984) for the difference of two normal variates. Unlike the above authors, the method we propose is based on the recent advances in likelihood based asymptotics which possess third-order accuracy.

The rest of paper is organized as follows. Section 2 discusses existing methods including the approximate method of Reiser & Guttman (1986), the approach of Guo & Krishnamoorthy (2004) and some first-order likelihood based methods. Section 3 develops our proposed third-order procedure. An empirical example using clinical data as well as a Monte Carlo simulation study are examined in Section 4. Section 5 summarizes this paper.

2 Existing methods for interval estimation of $P(X > Y)$

Let $X$ and $Y$ be two independent random variables such that $X \sim N(\mu_x, \sigma_x^2)$ and $Y \sim N(\mu_y, \sigma_y^2)$. With observed samples $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_m)$, the log likelihood function of $\theta = (\mu_x, \sigma_x^2, \mu_y, \sigma_y^2)$ is given as

$$l(\theta) = l(\mu_x, \sigma_x^2, \mu_y, \sigma_y^2) = -\frac{n}{2} \log \sigma_x^2 - \frac{m}{2} \log \sigma_y^2 - \frac{1}{2} \left( \frac{\sigma_x^2}{\sigma_x^2} \right)^2 \sum_{i=1}^{n} (x_i - \mu_x)^2 - \frac{1}{2} \left( \frac{\sigma_y^2}{\sigma_y^2} \right)^2 \sum_{j=1}^{m} (y_j - \mu_y)^2. \quad (1)$$

By solving for the first-order conditions, $\partial l(\theta) / \partial \theta = 0$, we can obtain the overall maximum likelihood estimator (MLE) of $\theta$ denoted by $\hat{\theta} = (\hat{\mu}_x, \hat{\sigma}_x^2, \hat{\mu}_y, \hat{\sigma}_y^2)'$, with
\[
\hat{\mu}_x = \frac{1}{n} \sum_{i=1}^{n} x_i \quad \hat{\sigma}^2_x = \frac{1}{n} \sum_{i=1}^{n} (x_i - \hat{\mu}_x)^2 \\
\hat{\mu}_y = \frac{1}{m} \sum_{j=1}^{m} y_j \quad \hat{\sigma}^2_y = \frac{1}{m} \sum_{j=1}^{m} (y_j - \hat{\mu}_y)^2.
\]

An estimated variance of the maximum likelihood estimator is \( \hat{\text{Var}}(\hat{\theta}) = j_{\theta\theta}(\hat{\theta}) \) where

\[
j_{\theta\theta}(\hat{\theta}) = -\frac{\partial^2 l(\theta)}{\partial \theta \partial \theta'} \bigg|_{\theta = \hat{\theta}} = \begin{pmatrix} n/\hat{\sigma}^2_x & 0 & 0 & 0 \\
0 & n/2\hat{\sigma}^4_x & 0 & 0 \\
0 & 0 & m/\hat{\sigma}^2_y & 0 \\
0 & 0 & 0 & m/2\hat{\sigma}^4_y \end{pmatrix}
\]

is the observed information matrix evaluated at \( \hat{\theta} \).

### 2.1 An approximate confidence interval estimation of \( R = P(X > Y) \)

The stress-strength reliability \( R = P(X > Y) \) when \( X \) and \( Y \) are independent normal random variables is given by

\[
R = P(X > Y) = \Phi(\psi),
\]

where

\[
\psi = \psi(\theta) = \frac{\mu_x - \mu_y}{\sqrt{\sigma^2_x + \sigma^2_y}}
\]

and \( \Phi(\cdot) \) is the cumulative distribution function of the standard normal distribution.

By the Invariance Principle, the MLE of \( R \) can be obtained from \( \hat{\theta} \) by

\[
\hat{R} = \Phi(\hat{\psi}) = \Phi(\psi(\hat{\theta})) = \Phi\left(\frac{\hat{\mu}_x - \hat{\mu}_y}{\sqrt{\hat{\sigma}^2_x + \hat{\sigma}^2_y}}\right).
\]

Since \( \Phi(\cdot) \) is a monotonically increasing function of \( \psi \), by finding a confidence bound for \( \psi \), the confidence bound for \( R = \Phi(\psi) \) is trivially obtained. Thus, in the rest of the paper, we focus on \( \psi \) as our main parameter of interest.

With the variances not being assumed equal, as in the Behrens-Fisher problem, there exists no exact distribution for \( \hat{\psi} \); only approximate distributions have been proposed in the literature. However, it is well-known that

\[
\begin{align*}
\hat{\mu}_x - \hat{\mu}_y &\sim N\left(\mu_x - \mu_y, \frac{\sigma^2_x}{n} + \frac{\sigma^2_y}{m}\right) \\
n\frac{\hat{\sigma}^2_x}{\sigma^2_x} &\sim \chi^2_{n-1} \\
(n-1)\frac{s^2_x}{\hat{\sigma}^2_x} &\sim \chi^2_{n-1} \\
m\frac{\hat{\sigma}^2_y}{\sigma^2_y} &\sim \chi^2_{m-1} \\
(m-1)\frac{s^2_y}{\hat{\sigma}^2_y} &\sim \chi^2_{m-1},
\end{align*}
\]

3
where \( s^2_x \) and \( s^2_y \) are unbiased estimators of \( \sigma^2_x \) and \( \sigma^2_y \) defined by:

\[
s^2_x = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \hat{\mu}_x)^2 \quad \text{and} \quad s^2_y = \frac{1}{m-1} \sum_{j=1}^{m} (y_j - \hat{\mu}_y)^2.
\]

As shown by Reiser & Guttman (1986), the distribution of \( \hat{\psi} \) can be derived as a Behrens-Fisher type approximation that takes the form of a noncentral t-distribution:

\[
\sqrt{M} \hat{\psi} \sim t_{d}(\sqrt{M}\hat{\psi}), \tag{10}
\]

where \( \sqrt{M} \psi \) is the noncentrality parameter with

\[
M = \frac{\sigma^2_x + \sigma^2_y}{\frac{\sigma^4_x}{n-1} + \frac{\sigma^4_y}{m-1}} \tag{11}
\]

and the degrees of freedom \( d \) is given by the relation

\[
d = \frac{\sigma^2_x + \sigma^2_y}{\frac{\sigma^4_x}{n-1} + \frac{\sigma^4_y}{m-1}} \left( \frac{\frac{\sigma^4_x}{n-1} + \frac{\sigma^4_y}{m-1}}{\frac{\sigma^4_x}{n-1} + \frac{\sigma^4_y}{m-1}} \right) \tag{12}
\]

The Reiser & Guttman (1986) procedure is a generalization of several methods that exist in the statistical literature. For example, the procedure extends the method of Enis & Geisser (1971), who only considered the case where \( m = n \) and \( \sigma_x = \sigma_y \). It also generalizes Church & Harris (1970) and Mazumdar (1970), who treated the case where \( \mu_y \) and \( \sigma_y \) are known. \( M \) and \( d \) are unknown but are estimated by \( s^2_x \) and \( s^2_y \) in their formulas. An approximate \((1 - \alpha)100^{th}\) percentile of \( \psi \) can therefore be obtained by numerically solving for \( \psi \) in

\[
P[t_{d_1} \left( \sqrt{M_1} \psi \right) \leq \sqrt{M} \hat{\psi}] = 1 - \alpha.
\]

Hence, the \((1 - \alpha)100\%\) confidence interval for \( \psi \) can be obtained.

Likewise, the approach proposed by Guo & Krishnamoorthy (2004) also requires the computation of noncentral \( t \) distributions, noncentrality parameters and non-integer degrees of freedom. An approximate \( 1 - \alpha \) percentile of \( \psi \) can be obtained from the Guo & Krishnamoorthy (2004) method by solving

\[
\min \left\{ P \left[ t_{d_1} \left( \sqrt{M_1} \psi \right) \leq \sqrt{M_1} \psi \right] , P \left[ t_{d_2} \left( \sqrt{M_2} \psi \right) \leq \sqrt{M_2} \psi \right] \right\} = 1 - \alpha
\]

where

\[
\hat{M}_1 = \frac{s^2_x + s^2_y}{s^2_x/n + s^2_y/m}, \quad \hat{d}_1 = \left( s^2_x + s^2_y \right)^2 \left( \frac{s^4_x}{n-1} + \frac{s^4_y}{m-1} \right)
\]

\[
\hat{M}_2 = \frac{(n-1)s^2_x + (n-3)s^2_y}{(n-1)s^2_x/n + (n-3)s^2_y/m}, \quad \hat{d}_2 = \left( (n-1)s^2_x + (n-3)s^2_y \right)^2 \left( \frac{(n-1)^2 s^4_x}{n-1} + \frac{(n-3)^2 s^4_y}{m-1} \right).
\]

\(^1\)Note that the Reiser & Guttman (1986) and the Guo & Krishnamoorthy (2004) methods use \( s^2_x \) and \( s^2_y \) in their estimation of \( \sigma^2_x \) and \( \sigma^2_y \) instead of \( \hat{\sigma}^2_x \) and \( \hat{\sigma}^2_y \) like the MLE.
2.2 Likelihood-based first-order approximations methods

Based on the MLE of $\psi$ and the log likelihood function, two familiar methods can be used for confidence interval estimation of the parameter $\psi$. The first method, the standardized maximum likelihood estimate method (also known as the Wald method), is based on the statistic ($q$) defined by

$$q = q(\psi) = (\hat{\psi} - \psi)[\text{Var}(\hat{\psi})]^{-1/2}. \quad (13)$$

One can apply the delta method to estimate the variance of $\hat{\psi}$ by

$$\hat{\text{Var}}(\hat{\psi}) = \frac{\partial \psi(\hat{\theta})}{\partial \theta'} \frac{\partial \psi(\hat{\theta})}{\partial \theta} \hat{\text{Var}}(\hat{\theta}) \frac{\partial \psi(\hat{\theta})}{\partial \theta'} = \frac{1}{\hat{M}} + \frac{\hat{d}}{4\hat{d}}. \quad (14)$$

where $\hat{M}$ and $\hat{d}$ are maximum likelihood estimates of $M$ and $d$ obtained by plugging $\hat{\sigma}_x^2$ and $\hat{\sigma}_y^2$ in Formulas (11) and (12).

Since $q$ is asymptotically distributed as standard normal, a $(1 - \alpha)100\%$ confidence interval for $\psi$ can be approximated by

$$\left( \hat{\psi} - z_{\alpha/2} \sqrt{\hat{\text{Var}}(\hat{\psi})}, \hat{\psi} + z_{\alpha/2} \sqrt{\hat{\text{Var}}(\hat{\psi})} \right), \quad (15)$$

where $z_{\alpha}$ is the $(1 - \alpha)100^{th}$ percentile of the standard normal. Although the Wald method is simple, it is not invariant to parameterization.

In the statistical literature, with a scalar parameter of interest, one of the most common inferential methods that is invariant to parametrization is the signed log likelihood ratio method, which is based on the statistic ($r$) defined by

$$r = r(\psi) = \text{sgn}(\hat{\psi} - \psi)[2\{l(\hat{\theta}) - l(\hat{\theta}_\psi)\}]^{1/2}, \quad (16)$$

where $\hat{\theta}_\psi = (\hat{\mu}_x, \hat{\sigma}_x^2, \hat{\mu}_y, \hat{\sigma}_y^2)'$ is the constrained maximum likelihood estimator of $\theta$ under the constraint that $\psi(\theta) = \psi$. To obtain $l(\hat{\theta}_\psi)$, we maximize $l(\theta)$ subject to the constraint $\psi(\theta) = \psi$. We apply the Lagrange multiplier method. Denoting the multiplier by $\lambda$, the Lagrangian function can be written as

$$H(\theta, \lambda) = l(\theta) + \lambda(\psi(\theta) - \psi)$$

$$= -\frac{n}{2} \log \sigma_x^2 - \frac{m}{2} \log \sigma_y^2 - \sum_{i=1}^n \frac{(x_i - \mu_x)^2}{2\sigma_x^2} - \sum_{i=1}^m \frac{(y_i - \mu_y)^2}{2\sigma_y^2} + \lambda \left[ \frac{\mu_x - \mu_y}{\sqrt{\sigma_x^2 + \sigma_y^2}} - \psi \right]. \quad (17)$$

The constrained MLE $\hat{\theta}_\psi$ and the estimated Lagrange multiplier $\hat{\lambda}$ can then be obtained by numerically
solving the first-order conditions

\[
\frac{\partial H(\theta, \lambda)}{\partial \mu_x} = \sum_{i=1}^{n} \frac{(x_i - \mu_x)}{\sigma_x^2} + \frac{\lambda}{\sqrt{\sigma_x^2 + \sigma_y^2}} = 0
\]

\[
\frac{\partial H(\theta, \lambda)}{\partial \sigma_x^2} = -\frac{1}{\sigma_x^2} + \frac{\lambda}{\sqrt{\sigma_x^2 + \sigma_y^2}} = 0
\]

\[
\frac{\partial H(\theta, \lambda)}{\partial \mu_y} = \sum_{j=1}^{m} \frac{(y_j - \mu_y)}{\sigma_y^2} - \frac{\lambda}{\sqrt{\sigma_x^2 + \sigma_y^2}} = 0
\]

\[
\frac{\partial H(\theta, \lambda)}{\partial \sigma_y^2} = -\frac{1}{\sigma_y^2} + \frac{\lambda}{\sqrt{\sigma_x^2 + \sigma_y^2}} = 0
\]

\[
\frac{\partial H(\theta, \lambda)}{\partial \lambda} = \frac{\mu_x - \mu_y}{\sqrt{\sigma_x^2 + \sigma_y^2}} - \psi = 0.
\]

We define the tilted log likelihood function by \( \tilde{l}(\theta) = l(\theta) + \tilde{\lambda}[\psi(\theta) - \psi] \). Note that the log likelihood function and the tilted log likelihood function are the same when evaluated at the constrained MLE of \( \hat{\psi} \), \( l(\hat{\theta}_\psi) = \tilde{l}(\hat{\theta}_\psi) \). The observed information matrix of the tilted log likelihood function evaluated at \( \hat{\theta}_\psi \), \( \tilde{\mathbf{j}}_{\psi\psi}(\hat{\theta}_\psi) \) is then defined by

\[
\tilde{\mathbf{j}}_{\psi\psi}(\hat{\theta}_\psi) = -\frac{\partial^2 \tilde{l}(\theta)}{\partial \theta \partial \theta'} \bigg|_{\hat{\theta}_\psi} = \mathbf{j}_{\theta\theta}(\hat{\theta}_\psi) + \frac{\tilde{\lambda}}{(\sigma_x^2 + \sigma_y^2)^{3/2}} \begin{pmatrix}
0 & -\frac{1}{2} & 0 & -\frac{1}{2} \\
-\frac{1}{2} & 3 \bar{\mu}_x - \bar{\mu}_y & 1 & 3 \bar{\mu}_x - \bar{\mu}_y \\
0 & 1 & 0 & \frac{1}{2} \\
-\frac{1}{2} & 3 \bar{\mu}_x - \bar{\mu}_y & 1 & 3 \bar{\mu}_x - \bar{\mu}_y \\
\end{pmatrix}.
\]

The statistic \( r(\psi) \) is asymptotically distributed as standard normal. Hence, a \((1-\alpha)100\%\) confidence interval for \( \psi \) based on \( r(\psi) \) is given by

\[
\{ \psi : |r(\psi)| \leq z_{\alpha/2} \}.
\]

It is well-known that both the Wald method and the signed log likelihood ratio method are first-order methods; that is, both \( q(\psi) \) and \( r(\psi) \) converge in distribution to the standard normal distribution with rate of convergence \( O(k^{-1/2}) \) where \( k = \min\{m,n\} \).

### 3 The proposed method

In recent years, various methods have been proposed to improve the accuracy of the first-order methods. In particular, Barndorff-Nielsen (1986, 1991) introduced the modified signed log likelihood ratio statistic

\[
r^*(\psi) = r(\psi) + \frac{1}{r(\psi)} \log \left( \frac{Q(\psi)}{Q(r(\psi))} \right),
\]

where \( Q(\psi) \) is the quadratic form of the observed information matrix at the MLE \( \hat{\theta}_\psi \) given by

\[
Q(\psi) = \mathbf{j}_{\theta\theta}(\hat{\theta}_\psi) - \mathbf{j}_{\theta\theta}(\hat{\theta}_0) = \mathbf{j}_{\theta\theta}(\hat{\theta}_\psi) - \mathbf{j}_{\theta\theta}(\hat{\theta}) = \tilde{\mathbf{j}}_{\psi\psi}(\hat{\theta}_\psi).
\]
where $r(\psi)$ is the signed log likelihood ratio statistic defined in (16), and $Q(\psi)$ is a statistic that is based on the log likelihood function and an ancillary statistic. Barndorff-Nielsen (1986, 1991) showed that $r^*(\psi)$ is asymptotically distributed as standard normal with third-order accuracy. Fraser & Reid (1995) showed that for the exponential family model, $Q(\psi)$ is the standardized maximum likelihood estimate calculated in the canonical parameter scale. Reid (1996) and Severini (2000) provide a detailed overview of this development.

For the stress-strength reliability with independent normal random variables problem, as stated in (1), we have an exponential family model with canonical parameter

$$
\varphi(\theta) = \left( \frac{\mu_x}{\sigma_x^2}, \frac{1}{\sigma_x^2}, \frac{\mu_y}{\sigma_y^2}, \frac{1}{\sigma_y^2} \right)'.
$$

(22)

To re-express our parameter of interest on this canonical parameter scale, we require $\varphi_\theta(\theta)$ and $\psi_\theta(\theta)$ which denote the derivatives of $\varphi(\theta)$ and $\psi(\theta)$ with respect to $\theta$. The $(4 \times 4)$ matrix of derivatives for $\varphi(\theta)$ is,

$$
\varphi_\theta(\theta) = \begin{bmatrix}
1/\sigma_x^2 & -\mu_x/\sigma_x^2 & 0 & 0 \\
0 & -1/\sigma_x^2 & 0 & 0 \\
0 & 0 & 1/\sigma_y^2 & -\mu_y/\sigma_y^2 \\
0 & 0 & 0 & -1/\sigma_y^2
\end{bmatrix}.
$$

(23)

The parameter of interest $\psi(\theta)$ is given in (5) with

$$
\psi_\theta(\theta) = \left( \frac{1}{(\sigma_x^2 + \sigma_y^2)^{1/2}}, -\frac{\mu_x - \mu_y}{2(\sigma_x^2 + \sigma_y^2)^{3/2}}, -\frac{1}{(\sigma_x^2 + \sigma_y^2)^{1/2}}, -\frac{\mu_x - \mu_y}{2(\sigma_x^2 + \sigma_y^2)^{3/2}} \right)'.
$$

(24)

By change-of-basis then, $\hat{\psi} - \psi = \psi(\hat{\theta}) - \psi(\theta)$ calculated in the canonical parameter, $\varphi(\theta)$, scale is

$$
\text{sgn}(\hat{\psi} - \psi)|\chi(\hat{\theta}) - \chi(\hat{\theta}_\psi)|,
$$

where

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

(25)

In addition, since $l(\theta) = l(\varphi(\theta))$ and $\tilde{l}(\theta) = \tilde{l}(\varphi(\theta))$, the determinant of the observed information matrix based on the log likelihood function $l(\theta)$ and the titled log likelihood function $\tilde{l}(\theta)$ calculated in $\varphi(\theta)$ scale can be obtained by using the chain-rule in differentiation, and we have

$$
|\tilde{j}_{\varphi'\varphi'}(\tilde{\theta})| = |\tilde{j}_{\varphi'\varphi'}(\varphi(\tilde{\theta}))|\varphi_\theta(\tilde{\theta})^{-2}
$$

and

$$
|\tilde{j}_{\varphi'\varphi}(\tilde{\theta}_\psi)| = |\tilde{j}_{\varphi'\varphi}(\varphi(\tilde{\theta}_\psi))|\varphi_\theta(\tilde{\theta}_\psi)^{-2}.
$$

Hence the asymptotic variance of $\chi(\hat{\theta}) - \chi(\hat{\theta}_\psi)$ calculated in $\varphi(\theta)$ scale is

$$
\text{Var}(\chi(\hat{\theta}) - \chi(\hat{\theta}_\psi)) = \left\{ \frac{|\tilde{j}_{\varphi'\varphi'}(\tilde{\theta}_\psi)|}{|\tilde{j}_{\varphi'\varphi'}(\varphi(\tilde{\theta}_\psi))|} \right\} \psi_\theta(\hat{\theta}_\psi)\tilde{j}_{\varphi'\varphi}(\varphi(\tilde{\theta}_\psi))^2 \tilde{j}_{\varphi'\varphi}(\varphi(\tilde{\theta}_\psi)).
$$

(26)
Thus the standardized MLE of $\hat{\psi}$ calculated in the $\varphi(\theta)$ scale is

$$Q(\psi) = \text{sgn}(\hat{\psi} - \psi) \frac{|\chi(\hat{\theta}) - \chi(\hat{\theta}_\psi)|}{\sqrt{\text{Var}(\chi(\hat{\theta}) - \chi(\hat{\theta}_\psi))}}.$$  

(27)

Finally, the modified signed log likelihood ratio statistic can be obtained from (21) and is asymptotically distributed as standard normal with accuracy $O(k^{-3/2})$ where $k = \min\{m, n\}$. The $(1 - \alpha)100\%$ confidence interval for $\psi$ is given by

$$I_{\psi}^{1-\alpha} = \{ \psi : |r^*(\psi)| \leq z_{\alpha/2} \}.$$  

(28)

Hence, the $(1 - \alpha)100\%$ confidence interval for $R$ is then given by $I_R^{1-\alpha} = \Phi(I_{\psi}^{1-\alpha})$, where $\Phi(\cdot)$ is the cumulative density function of the standard normal distribution. Note that the added accuracy, from first-order to third-order, does not require many additional calculation as all the ingredients are available from the calculations of the signed log likelihood ratio statistic.

4 Empirical example and Monte Carlo Simulations

In this section, we provide an empirical example and a small sample Monte Carlo simulation study for inference on the stress-strength reliability. We compare our proposed method (Proposed) to existing methods, namely the log likelihood ratio method ($r$), the Reiser & Guttman approximation (RG), and the Guo & Krishnamoorthy (GK) approach.

4.1 Empirical example

This example is drawn from the study of Azuma et al. (2011) regarding human abdominal aortic aneurysm, a pathologic dilatation of the abdominal aortic diameter which is a significant cause of morbidity and mortality in the United States and worldwide. Accurate measurements of this diameter are crucial for detecting the seriousness of the disease. In order to assess the validity and accuracy of a new measurement instrument based on ultrasound (US), measurements of the aortic luminal diameter (ALD) of two groups of patients were performed using the ultrasound. The first group consisted of 20 patients classified with a small aneurysm diameter (the control or low risk group) and the second group consisted of 20 patients with a large aneurysm diameter (the treatment or high risk group). The ultrasound measurements obtained are given in Table 1.

The goal of the study was to evaluate the diagnostic accuracy of this new instrument in discriminating between patients with low and high rupture risk. The ultrasound measurements were assumed to be normally distributed by the authors. This assumption is also supported by our QQ-plot representation given in Figure 1. The graphs in this figure show that the distributions of the control and the treatment groups can be fairly assumed to follow some normal distributions.

---

2In contrast to the Reiser & Guttman (1986) and Guo & Krishnamoorthy (2004) methods, whose rates of convergence are unknown.

3Practically, the confidence interval $I_{\psi}^{1-\alpha}$ is obtained by numerically solving for $\psi$ in the inequality $|r^*(\psi)| \leq z_{\alpha/2}$ using a sufficient number of grid points of $\psi$ chosen in an appropriate range.
Table 1: Ultrasound measurements of the abdominal aortic diameter

<table>
<thead>
<tr>
<th>Patient #</th>
<th>Control Group</th>
<th>Treatment Group</th>
<th>Patient #</th>
<th>Control Group</th>
<th>Treatment Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.586</td>
<td>0.880</td>
<td>11</td>
<td>0.684</td>
<td>1.132</td>
</tr>
<tr>
<td>2</td>
<td>0.623</td>
<td>1.075</td>
<td>12</td>
<td>0.743</td>
<td>1.210</td>
</tr>
<tr>
<td>3</td>
<td>0.552</td>
<td>0.997</td>
<td>13</td>
<td>0.735</td>
<td>1.273</td>
</tr>
<tr>
<td>4</td>
<td>0.606</td>
<td>0.921</td>
<td>14</td>
<td>0.712</td>
<td>1.412</td>
</tr>
<tr>
<td>5</td>
<td>0.626</td>
<td>1.093</td>
<td>15</td>
<td>0.765</td>
<td>1.367</td>
</tr>
<tr>
<td>6</td>
<td>0.586</td>
<td>1.154</td>
<td>16</td>
<td>0.756</td>
<td>1.486</td>
</tr>
<tr>
<td>7</td>
<td>0.623</td>
<td>1.074</td>
<td>17</td>
<td>0.820</td>
<td>1.427</td>
</tr>
<tr>
<td>8</td>
<td>0.606</td>
<td>1.348</td>
<td>18</td>
<td>0.781</td>
<td>1.355</td>
</tr>
<tr>
<td>9</td>
<td>0.634</td>
<td>1.153</td>
<td>19</td>
<td>0.812</td>
<td>1.562</td>
</tr>
<tr>
<td>10</td>
<td>0.684</td>
<td>1.331</td>
<td>20</td>
<td>0.879</td>
<td>1.761</td>
</tr>
</tbody>
</table>

Figure 1: QQ plots of controls and treatments ultrasound measurements

We let $Y$ be the ultrasound measurements of the control group and $X$ be the measurements of the treatment group. Then $R = P(X > Y)$ is the probability that abdominal aortic diameter measurement of a patient with high rupture risk is higher than the abdominal aortic diameter measurement of a patient with low rupture risk. It thus evaluates the diagnostic accuracy of the ultrasound instrument in discriminating between these two groups of patients. We computed the $p$-value function of $R$

$$p(R^0) = P(R > R^0)$$

where $R^0$ is a specific value of $R$, for the four methods discussed in this paper. The higher the value of $R^0$ the more accurate the instrument is in discriminating between the two types of patients. In particular, a value of $R^0$ bigger than 0.5 would mean that the ultrasound instrument is reliable more than 50% of the time. The plot of $R^0$ vs $p(R^0)$ is presented in Figure 2 and the lower confidence bounds of the various methods are given in Table 2. It is clear that the four discussed methods give very different results for this data set. Hence, it is important to compare the accuracy of the four discussed methods.
4.2 Monte Carlo simulation results

In order to study the accuracy of the four methods discussed in this paper, we perform Monte Carlo simulation studies. The accuracy is evaluated by computing confidence intervals for $R$ with each method using the following criteria:

- **Central Coverage Probability**: The proportion of the true parameter value falling within the confidence interval.

- **Upper Probability Error**: The proportion of the true parameter value that falls above the confidence interval.

- **Lower Probability Error**: The proportion of the true parameter value that falls below the confidence interval.

<table>
<thead>
<tr>
<th>Method</th>
<th>90%</th>
<th>95%</th>
<th>97.5%</th>
<th>99%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed</td>
<td>0.9848</td>
<td>0.9816</td>
<td>0.9787</td>
<td>0.9743</td>
</tr>
<tr>
<td>$r$</td>
<td>0.9829</td>
<td>0.9794</td>
<td>0.9760</td>
<td>0.9712</td>
</tr>
<tr>
<td>RG</td>
<td>0.9652</td>
<td>0.9526</td>
<td>0.9390</td>
<td>0.9195</td>
</tr>
<tr>
<td>GK</td>
<td>0.9634</td>
<td>0.9509</td>
<td>0.9376</td>
<td>0.9191</td>
</tr>
</tbody>
</table>
We performed 10,000 simulations of pairs of data generated from the distributions of $X \sim N(1.2506, 0.0494)$, $Y \sim N(0.6906, 0.0085)$ and $R = 0.99$. These distributions were chosen so as to mimic the clinical data in our empirical example above. Various combinations of sample sizes were considered, $n = 10, 15, 20$ and $m = 10, 15, 20$, and 95% confidence intervals for the reliability parameter were obtained for each combination of samples. Simulations results are presented in Table 3. The nominal values for the central coverage, the lower probability error and the upper probability error are 0.95, 0.025 and 0.025, with standard errors 0.0022, 0.0016 and 0.0016 respectively.

<table>
<thead>
<tr>
<th>(n, m)</th>
<th>Method</th>
<th>Upper Error</th>
<th>Lower Error</th>
<th>Central Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>n =10, m =10</td>
<td>Proposed</td>
<td>0.0271</td>
<td>0.0252</td>
<td>0.9477</td>
</tr>
<tr>
<td></td>
<td>r</td>
<td>0.0493</td>
<td>0.0176</td>
<td>0.9331</td>
</tr>
<tr>
<td></td>
<td>RG</td>
<td>0.0588</td>
<td>0.0154</td>
<td>0.9258</td>
</tr>
<tr>
<td></td>
<td>GK</td>
<td>0.0274</td>
<td>0.0257</td>
<td>0.9469</td>
</tr>
<tr>
<td>n =10, m =15</td>
<td>Proposed</td>
<td>0.0265</td>
<td>0.0256</td>
<td>0.9479</td>
</tr>
<tr>
<td></td>
<td>r</td>
<td>0.0457</td>
<td>0.0217</td>
<td>0.9326</td>
</tr>
<tr>
<td></td>
<td>RG</td>
<td>0.0634</td>
<td>0.0158</td>
<td>0.9208</td>
</tr>
<tr>
<td></td>
<td>GK</td>
<td>0.0432</td>
<td>0.0243</td>
<td>0.9325</td>
</tr>
<tr>
<td>n =10, m =20</td>
<td>Proposed</td>
<td>0.0292</td>
<td>0.0271</td>
<td>0.9437</td>
</tr>
<tr>
<td></td>
<td>r</td>
<td>0.0553</td>
<td>0.0189</td>
<td>0.9258</td>
</tr>
<tr>
<td></td>
<td>RG</td>
<td>0.0685</td>
<td>0.0149</td>
<td>0.9166</td>
</tr>
<tr>
<td></td>
<td>GK</td>
<td>0.0481</td>
<td>0.0225</td>
<td>0.9294</td>
</tr>
<tr>
<td>n =15, m =10</td>
<td>Proposed</td>
<td>0.0249</td>
<td>0.0242</td>
<td>0.9509</td>
</tr>
<tr>
<td></td>
<td>r</td>
<td>0.0411</td>
<td>0.0163</td>
<td>0.9426</td>
</tr>
<tr>
<td></td>
<td>RG</td>
<td>0.0452</td>
<td>0.0149</td>
<td>0.9399</td>
</tr>
<tr>
<td></td>
<td>GK</td>
<td>0.0383</td>
<td>0.0216</td>
<td>0.9401</td>
</tr>
<tr>
<td>n =15, m =15</td>
<td>Proposed</td>
<td>0.0251</td>
<td>0.0248</td>
<td>0.9501</td>
</tr>
<tr>
<td></td>
<td>r</td>
<td>0.0407</td>
<td>0.0155</td>
<td>0.9438</td>
</tr>
<tr>
<td></td>
<td>RG</td>
<td>0.0454</td>
<td>0.0154</td>
<td>0.9392</td>
</tr>
<tr>
<td></td>
<td>GK</td>
<td>0.0266</td>
<td>0.0245</td>
<td>0.9489</td>
</tr>
<tr>
<td>n =15, m =20</td>
<td>Proposed</td>
<td>0.0266</td>
<td>0.0261</td>
<td>0.9473</td>
</tr>
<tr>
<td></td>
<td>r</td>
<td>0.0517</td>
<td>0.0168</td>
<td>0.9315</td>
</tr>
<tr>
<td></td>
<td>RG</td>
<td>0.0494</td>
<td>0.0162</td>
<td>0.9344</td>
</tr>
<tr>
<td></td>
<td>GK</td>
<td>0.0343</td>
<td>0.0261</td>
<td>0.9396</td>
</tr>
<tr>
<td>n =20, m =10</td>
<td>Proposed</td>
<td>0.0249</td>
<td>0.0247</td>
<td>0.9504</td>
</tr>
<tr>
<td></td>
<td>r</td>
<td>0.0351</td>
<td>0.0193</td>
<td>0.9456</td>
</tr>
<tr>
<td></td>
<td>RG</td>
<td>0.0381</td>
<td>0.0168</td>
<td>0.9451</td>
</tr>
<tr>
<td></td>
<td>GK</td>
<td>0.0391</td>
<td>0.0207</td>
<td>0.9402</td>
</tr>
<tr>
<td>n =20, m =15</td>
<td>Proposed</td>
<td>0.0253</td>
<td>0.0244</td>
<td>0.9503</td>
</tr>
<tr>
<td></td>
<td>r</td>
<td>0.0417</td>
<td>0.0161</td>
<td>0.9422</td>
</tr>
<tr>
<td></td>
<td>RG</td>
<td>0.0425</td>
<td>0.0160</td>
<td>0.9415</td>
</tr>
<tr>
<td></td>
<td>GK</td>
<td>0.0328</td>
<td>0.0236</td>
<td>0.9436</td>
</tr>
<tr>
<td>n =20, m =20</td>
<td>Proposed</td>
<td>0.0253</td>
<td>0.0249</td>
<td>0.9498</td>
</tr>
<tr>
<td></td>
<td>r</td>
<td>0.0448</td>
<td>0.0159</td>
<td>0.9393</td>
</tr>
<tr>
<td></td>
<td>RG</td>
<td>0.0469</td>
<td>0.0162</td>
<td>0.9389</td>
</tr>
<tr>
<td></td>
<td>GK</td>
<td>0.0301</td>
<td>0.0209</td>
<td>0.9490</td>
</tr>
</tbody>
</table>
From Table 3, it is evident that the proposed method yields better results than the three existing methods described in this paper. The central coverage, lower errors and upper errors of the proposed method are all within three standard errors of the nominal values even for small and unequal sample sizes. In fact, the central coverage, lower errors and upper errors of the proposed method are actually within two standard errors of the nominal values for all cases but one (which is within three standard errors of the nominal values). On the other hand, the other methods give much less accurate central coverages. Further, the proposed method delivers symmetric errors whereas the other three methods do not. The GK method yields better results than the RG and the signed log likelihood methods, especially for the central coverage, when the sample sizes of $X$ and $Y$ are equal. However the quality of the GK results tend to depreciate when sample sizes are different. Simulation studies with many other combinations of the parameter values and sample sizes have also been calculated and the results are very similar to what was presented here. Thus, they are not reported in this paper but are available from the authors upon request.

5 Discussion

We have proposed an $O(n^{-3/2})$ likelihood-based method to obtain interval estimation for the stress-strength reliability model $R = P(X > Y)$ when $X$ and $Y$ are two independent normal variables. Our results also hold for the case where $X$ and $Y$ are log-normally distributed, since $P(X > Y) = P(\log X > \log Y)$. Simulation results indicate that the proposed method is extremely accurate even when only small sample sizes of observations are available.

The main contribution of this study is that it fills the gap in the literature by providing a highly accurate estimation procedure not only for the case where both the mean and variance of $X$ and $Y$ are unknown but also when sample sizes of observations of $X$ and $Y$ are different and possibly very small. In fact, when estimating the stress-strength reliability in the presence of small samples, first-order methods will be highly sensitive to parametric assumptions, even to those commonly used like normality (Neal et al. 1991). Our results show that these methods can be seriously misleading in practice. The nonparametric approach on the other hand will not be helpful in this situation because of its extremely conservative nature (Basu 1981). Our procedure is therefore a useful alternative.

One difficulty that we point out with our proposed method is that the numerical calculations of p-values can be tricky when the parameter estimates $\hat{\psi}$ are too close to the hypothesized values $\psi$, which may then be singular for Equation (21). However, this difficulty is easily overcome by applying the bridging procedure proposed by Fraser et al. (2003). We note that the $R$ programs for all calculations appearing in this paper are available from the authors upon request. A possible problem to examine in future research could be to extend the method to the case where observations of $X$ and $Y$ are time series data exhibiting statistical dependence or correlation across time.
References


