



**HAL**  
open science

## Confidence intervals for reliability of stress-strength models in the normal case

Alessandro Barbiero

► **To cite this version:**

Alessandro Barbiero. Confidence intervals for reliability of stress-strength models in the normal case. *Communications in Statistics - Simulation and Computation*, 2011, 40 (06), pp.907-925. 10.1080/03610918.2011.560728 . hal-00681617

**HAL Id: hal-00681617**

**<https://hal.science/hal-00681617>**

Submitted on 22 Mar 2012

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



**Confidence intervals for reliability of stress-strength models in the normal case**

Journal:	<i>Communications in Statistics - Simulation and Computation</i>
Manuscript ID:	LSSP-2010-0183.R2
Manuscript Type:	Original Paper
Date Submitted by the Author:	28-Jan-2011
Complete List of Authors:	Barbiero, Alessandro; Università degli Studi di Milano, Department of Economics, Business and Statistics
Keywords:	asymptotically normal, independent normal, parametric bootstrap, Monte Carlo simulations, stress-strength, variance estimation
Abstract:	In this paper we propose some procedures to get confidence intervals for the reliability in stress-strength models. The confidence intervals are obtained either through a parametric bootstrap procedure or using asymptotic results, and are applied to the particular context of two independent normal random variables. The performance of these estimators and other known approximate estimators are empirically checked through a simulation study which considers several scenarios.
<p>Note: The following files were submitted by the author for peer review, but cannot be converted to PDF. You must view these files (e.g. movies) online.</p> <p>reliability_second_revision.tex</p>	

SCHOLARONE™  
Manuscripts

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

For Peer Review Only

1  
2  
3 **Reply to Editor and Reviewers' comments on manuscript "Confidence intervals for reliability**  
4 **of stress-strength models in the normal case" by A. Barbiero**  
5  
6  
7

8  
9 I would like to thank the Editor, Associate Editor and two Referees for their very helpful comments  
10 on my submitted manuscript. I agree with all their comments and believe that they will contribute to  
11 the improvement of the paper.  
12

13 You will find the modifications/additions to the original text coloured in red in the pdf file I have  
14 enclosed (`\textcolor{red}{} in the LaTeX file). I have corrected some minor typos or`  
15 imprecisions (e.g. added bold font for vectors of parameters in some formulas).  
16  
17

18 A point-to-point list of responses is given below (my answers after "R:").  
19  
20  
21

22  
23 **Reviewer #1**

24 Comments to the Author

25  
26 Comments on the paper "Confidence intervals for reliability in the normal case" by Barbiero,  
27 Alessandro  
28

29 This paper considers a bootstrap and some approximate methods for the interval estimation of  
30 reliability for normally distributed  
31 stress and strength. The paper is nicely written and can be easily followed.  
32  
33

34  
35  
36  
37 R: Thanks for your appreciation.  
38  
39

40  
41 **Reviewer #2**

42 Comments to the Author

43  
44 Comments on the paper "Confidence Intervals for Reliability in the Normal Case" by Barbiero

45 The revised version looks better than the first submission, but still I have some concerns.  
46  
47

48  
49 1. The title should include the term "stress-strength", not just reliability.

50  
51 R: Indeed "reliability" alone is a too vague term; I have changed the title, which now is "Confidence  
52 intervals for reliability of stress-strength models in the normal case"  
53  
54

55  
56 2. Page 4, below Eqn (2): Write "Chang's method is overly conservative producing CIs that are  
57 unnecessarily wider."  
58

59  
60 R: Done

1  
2  
3 3. Page 5, line 10 from below: noncentral t, not noncentral Student's T.  
4

5 R: Done  
6  
7

8  
9 4. Page 5: Cumulative distribution function, not cumulative density function.  
10

11 R: Done  
12  
13

14 5. Section 3, first line: We will now describe our new method....  
15

16 R: Done  
17  
18

19 6. Section 3.1: Check the reference to Efron's book.  
20

21 R: checked; now it is referenced to as "Efron and Tibshirani, 1994"  
22  
23

24 7. I also have some concerns with the description of the PB approach. Note that all procedures  
25 depend on only the sample means and variances, and so we do not have to generate sample from  
26  $N(\bar{x}; s^2_x)$  and  $N(\bar{y}; s^2_y)$ . Indeed, we only have to generate  $\bar{x} \sim N(\bar{x}; s^2_x/n)$   
27 and  $s_x^2 \sim s_x^2 \chi^2_{n-1}/(n-1)$ . Similarly,  $\bar{y} \sim N(\bar{y}; s^2_y/n)$  and  $s_y^2 \sim s_y^2 \chi^2_{n-1}/(n-1)$ .  
28  $\bar{x}$  and  $s_x^2$  can be generated.  
29

30 R: due to the properties of (the sample mean of) the normal distribution, actually one can directly  
31 generate the sample means and variances, but in a more general case one has to bootstrap the  
32 original distribution. I have inserted your note in the algorithm: "... since all procedures depend on  
33 only the sample means and variances, for the well-known properties of the normal distribution, for  
34 this case we only have to generate  $\bar{X} \sim N(\bar{x}; s^2_x/n)$  and  $S_x^2 \sim s_x^2 \chi^2_{n-1}/(n-1)$ .  
35 Similarly,  $\bar{Y} \sim N(\bar{y}; s^2_y/n)$  and  $S_y^2 \sim s_y^2 \chi^2_{n-1}/(n-1)$ .  
36  $\bar{Y}$  and  $S_y^2$  can be generated."  
37  
38  
39  
40  
41  
42  
43  
44  
45

46 8. The asymptotic procedures perform well for constructing two-sided CIs, but not for one-sided  
47 CIs. Indeed, Krishnamoorthy and Lin (2010) noted that for the Weibull case asymptotic CIs  
48 overcover on one-tail and undercover on the other, and so overall coverage probabilities of  
49 asymptotic two-sided CIs seem to be OK. I observe similar behavior for the normal case also. The  
50 author should note this point.  
51

52 R: Thanks for the explanation; I put a note when commenting the results for the one-side case: "One  
53 should also note an overall light worsening of the performance of asymptotic procedures passing  
54 from two-sided to one-sided CIs: this may be ascribed to the fact that asymptotic CIs overcover on  
55 one-tail and undercover on the other, and so overall coverage probabilities of asymptotic two-sided  
56  
57  
58  
59  
60

1  
2  
3 CIs turn out to be close to the nominal  $1-\alpha$ . This behavior has been detected by \cite{Kri} for  
4  
5 the Weibull case."  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

For Peer Review Only

# Confidence intervals for reliability of stress-strength models in the normal case

Alessandro Barbiero

Department of Economics, Business and Statistics,

Università degli Studi di Milano, Italy.

alessandro.barbiero@unimi.it

## Abstract

In this paper we propose some procedures to get confidence intervals for the reliability in stress-strength models. The confidence intervals are obtained either through a parametric bootstrap procedure or using asymptotic results, and are applied to the particular context of two independent normal random variables. The performance of these estimators and other known approximate estimators are empirically checked through a simulation study which considers several scenarios.

**Keywords:** asymptotically normal, independent normal, parametric bootstrap, Monte Carlo simulations, stress-strength, variance estimation

## 1 Introduction

The aim of this paper is to propose confidence intervals for the probability  $P(Y < X)$ , where  $X$  and  $Y$  are independent random variables and when sample values  $x$  and  $y$  are observed. The recent years have seen a lot of publications on this subject: the basic impetus to these developments can be attributed to the specific practical problem of applied statistics summarized by

1 the term “stress-strength”. In the simplest terms this can be described as an assessment of “re-  
2 liability” of a “component” in terms of random variables  $Y$  representing “stress” experienced  
3 by the component and  $X$  representing the “strength” of the component available to overcome  
4 the stress. According to this simplified scenario, if the stress exceeds the strength ( $Y > X$ ) the  
5 component would fail; and vice versa. Reliability is defined as the probability of not failing:  
6  $R = P(Y < X)$ . The reliability problem arises in the fields of aeronautical, civil, mechanical and  
7 electronic engineering. For example,  $X$  may be the breakdown voltage of a capacitor, while  $Y$   
8 may represent the voltage output of a transverter (Hall, 1984); or  $X$  may be the chamber burst  
9 strength and  $Y$  the operating pressure for a rocket-motor (Reiser and Guttman, 1986).

10 The beginning point of this idea was introduced by Birnbaum (1956) and developed by  
11 Birnbaum and McCarty (1958). The latter paper does for the first time include  $P(Y < X)$  in  
12 its title. The formal term “stress-strength” appears for the first time in the title of the paper by  
13 Church and Harris (1970). In the course of time, there have been attempts to introduce further  
14 elements of adherence to reality, including various generalizations and applications; many pa-  
15 pers are devoted both to merely probabilistic problems associated with the evaluation of  $R$  and  
16 the construction of efficient and reliable estimators of this parameter, based on sample values  
17 with various assumptions on the distributions of  $X$  and  $Y$ . So the research has not only dealt  
18 with the problems of deriving theoretical expressions for  $P(Y < X)$  and its modifications and  
19 extensions under various distributional assumptions, but also with estimation of these proba-  
20 bilities based on samples of various structure, providing for approximations to variances and  
21 confidence bounds (Kotz *et al.*, 2003). Many of these works presume that both random vari-  
22 ables has distributions belonging to the same family (such as normal, exponential, log-normal,  
23 Weibull, etc.) and more importantly they assume independence between them. Some authors,  
24 however, considered the case in which  $X$  and  $Y$  admit a specified form of dependence (bivariate  
25 normal, bivariate exponential, etc.): see for example Mukherjee and Sharan (1985); Nadarajah  
26 and Kotz (2005); Roy (1993). Harris and Soms (1991) emphasize the fact that in many ap-  
27 plications the “reliability” has to be very close to one for the device to have any possibility of  
28 “useful” life: this implies that very large samples may be needed to obtain sufficiently precise  
29 estimates of reliability, since we are dealing with extreme tails of distributions.

30 In this paper we will consider the estimation of reliability when  $X$  and  $Y$  are independent  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51



random variables and their distribution belongs to a known family (normal) with unknown parameters. Based upon ML estimators, we will consider and refine approximate interval estimators already appeared in literature, we will propose an original bootstrap interval estimator and empirically investigate and compare their performance in terms of coverage and accuracy through an extensive simulation study.

The paper is structured as follows: in Section 2 we describe available methods for interval estimation of reliability for an independent setup; in Section 3 we describe our proposals; in Section 4 we describe the simulation design, show the results and discuss them; in Section 5, we present a real application, in Section 6 we give some final remarks.

## 2 Available methods

We will now briefly introduce the concept of “reliability” for a stress-strength model and summarize the methods available in literature for its interval estimation. We will confine ourselves to the parametric case only.

Let the strength be modeled by r.v.  $X$  and the stress by r.v.  $Y$ . Let us suppose that  $X$  follows a continuous distribution with vector of parameters  $\theta$  and  $Y$  follows a continuous distributions with vector of parameters  $\lambda$ . The reliability of a stress-strength model is formally defined as

$$R = P(Y < X) = P(Y - X < 0)$$

and the following expression can be used to obtain the reliability:

$$R = \int_{-\infty}^{+\infty} f_x(x, \theta) \int_{-\infty}^x f_y(y, \lambda) dy dx$$

where  $f_x$  and  $f_y$  denote the p.d.f. of  $X$  and  $Y$  respectively. For independent normal r.v.  $X \sim N(\mu_x, \sigma_x)$  and  $Y \sim N(\mu_y, \sigma_y)$ , the reliability presents the easy expression

$$R = \Phi \left( \frac{\mu_x - \mu_y}{\sqrt{\sigma_x^2 + \sigma_y^2}} \right) \quad (1)$$

where  $\Phi$  denotes the standard normal c.d.f.

A straightforward procedure for getting conservative confidence intervals for the reliability  $R$  was described by Chang (Chang, 1995). The basic idea was to start from confidence intervals

for the parameters of the two r.v. and then numerically obtain a lower and upper bound for reliability from the lower and upper bounds of these intervals. Chang's method provides for the reliability of a stress-strength model with two independent normal variables the following conservative  $100 \cdot (1 - 2\alpha)^2\%$  interval:

$$[\underline{R}, \bar{R}] = \left[ \inf_{\Omega} R, \sup_{\Omega} R \right] = \left[ \Phi \left[ \frac{\underline{\mu}_x - \bar{\mu}_y}{\sqrt{\underline{\sigma}_x^2 + \bar{\sigma}_y^2}} \right], \Phi \left[ \frac{\bar{\mu}_x - \underline{\mu}_y}{\sqrt{\bar{\sigma}_x^2 + \underline{\sigma}_y^2}} \right] \right] \quad (2)$$

where  $\Omega = \left\{ \underline{\mu}_x \leq \mu_x \leq \bar{\mu}_x, \underline{\mu}_y \leq \mu_y \leq \bar{\mu}_y, \underline{\sigma}_x^2 \leq \sigma_x^2 \leq \bar{\sigma}_x^2, \underline{\sigma}_y^2 \leq \sigma_y^2 \leq \bar{\sigma}_y^2 \right\}$  and  $(\underline{\mu}_x, \bar{\mu}_x)$  and  $(\underline{\sigma}_x^2, \bar{\sigma}_x^2)$  are a  $(1 - \alpha)$  confidence interval for  $\mu_x$  and  $\sigma_x^2$  respectively, and analogously for the  $Y$  parameters. Yet, **Chang's method is overly conservative producing CIs that are unnecessarily wider.**

Reiser and Guttman (1986) examine statistical inference for  $P(Y < X)$ , where  $X$  and  $Y$  are independent normal variates with unknown means and variances. Two approximate methods for obtaining confidence intervals and an approximate Bayesian probability interval are described. The actual coverage probabilities of these intervals are examined by simulation. Specifically, they first consider

$$\delta = \frac{\mu_x - \mu_y}{\sqrt{\sigma_x^2 + \sigma_y^2}}$$

Since  $R$  is a monotonic function of  $\delta$ , finding a CI for  $R$  is equivalent to finding a CI for  $\delta$ . They argue that it seems more reasonable to base inference on a normal approximation to  $\delta$  than on the normal approximation to  $R$  since  $\delta$  is unbounded, while  $R$  is bounded in  $(-1, 1)$ . An approximate  $(1 - \alpha)$  CI they propose is given by

$$(\Phi(\delta_L), \Phi(\delta_U)) \quad (3)$$

where  $\delta_L$  and  $\delta_U$  are the lower and upper bound of the interval

$$\hat{\delta} \pm \left( \frac{1}{\hat{M}} + \frac{\hat{\delta}^2}{2\hat{f}} \right) z_{1-\alpha/2}$$

with

$$\hat{\delta} = \frac{\bar{x} - \bar{y}}{\sqrt{s_x^2 + s_y^2}}$$

$$\hat{f} = (s_x^2 + s_y^2)^2 / \left( \frac{s_x^2}{n-1} + \frac{s_y^2}{m-1} \right)$$

and

$$\hat{M} = \frac{s_x^2 + s_y^2}{s_x^2/n + s_y^2/m}$$

A  $(1 - \alpha)$  confidence lower bound can be obtained as  $\Phi(\delta_l)$  with  $\delta_l = \hat{\delta} - \left(\frac{1}{\hat{M}} + \frac{\hat{\delta}^2}{2\hat{f}}\right) z_{1-\alpha}$ .

The simulations show that the coverage is not badly influenced by unequal sample sizes.

Church and Harris (1970) obtain confidence intervals for  $P(Y < X)$  under the assumptions that  $X$  and  $Y$  are independently normally distributed and the distribution of  $Y$  is known. Downton (1973) derives the minimum variance unbiased estimator of  $P(Y < X)$  under the same assumptions. Approximations to this “best” estimator are obtained, and they are suggested as alternatives to the asymptotically equivalent estimator used by Church and Harris to obtain confidence intervals for that probability; they generalize the estimator to the case the distribution of  $Y$  is unknown. Weerahandi and Johnson (1992) derive approximate confidence intervals for the reliability, based upon generalized  $p$ -values, when  $X$  and  $Y$  are independently normally distributed and all the parameters are unknown. Guo and Krishnamoorthy (2004) propose an approximate (one-sided) CI for  $R$  by first computing a lower bound for  $\delta$  and then transforming it to a lower bound for  $R$  as done by Reiser and Guttman (1986). The proposal involves noncentral  $t$  with a non-integer number of degrees of freedom and requires a numerical solution. More in detail, if we denote with  $F(x; df, \xi)$  the noncentral  $t$  cumulative distribution function with  $df$  degrees of freedom and noncentrality parameter  $\xi$ , a first lower bound  $\delta_{1L}$  is the solution of the equation

$$F(\sqrt{\hat{m}_1} \hat{\delta}; \hat{f}_1, \sqrt{\hat{m}_1} \delta_{1L}) = 1 - \alpha$$

where

$$\hat{m}_1 = \frac{n(1 + \hat{q}_1)}{\hat{q}_1 + n/m}; \hat{f}_1 = \frac{(n-1)(\hat{q}_1 + 1)^2}{\hat{q}_1^2 + (n-1)/(m-1)}; \hat{q}_1 = s_1^2/s_2^2$$

whereas a second lower bound  $\delta_{2L}$  is the solution of the equation

$$F(\sqrt{\hat{m}_2} \hat{\delta}; \hat{f}_2, \sqrt{\hat{m}_2} \delta_{2L}) = 1 - \alpha$$

where

$$\hat{m}_2 = \frac{m(1 + \hat{q}_2)}{\hat{q}_2 + m/n}; \hat{f}_2 = \frac{(m-1)(\hat{q}_2 + 1)^2}{\hat{q}_2^2 + (m-1)/(n-1)}; \hat{q}_2 = s_2^2/s_1^2$$

The actual lower bound for  $R$  is obtained as

$$\hat{R}_L = \Phi(\min\{\delta_{1L}, \delta_{2L}\}). \quad (4)$$

Mukherjee and Mahiti (1998) develop an interval estimation procedure on the basis of the asymptotic normality of the maximum likelihood estimator (ML) of  $R$  and also provide interval estimation procedures based on variance stabilizing transformations such as logit and arcsine. Specifically, they consider the case of two independent Weibull distributions. For the same set-up, Krishnamoorthy and Lin (2010) propose an alternative method using a generalized variable approach.

In the next Section, we will provide some new procedures to obtain an approximate CI for  $R$ .

### 3 The proposals

We will now describe our new method for the interval estimation of reliability, based upon two independent samples from  $X$  and  $Y$ , denoted as  $x$  and  $y$ , of size  $n$  and  $m$  respectively. The first is based upon parametric bootstrap, the second ones upon asymptotic results.

#### 3.1 A parametric bootstrap procedure

Our first method is a straightforward extension and application of parametric bootstrap for the i.i.d. case (Efron and Tibshirani, 1994) to a two-sample case. It works as follows:

1. estimate the parameters of the r.v.  $X$  and  $Y$ , through  $\hat{\theta}$  and  $\hat{\lambda}$  (e.g. maximum likelihood)
2. estimate  $R$  as

$$\hat{R} = \int_{-\infty}^{+\infty} f_x(x, \hat{\theta}) \int_{-\infty}^x f_y(y, \hat{\lambda}) dy dx;$$

For normal r.v.  $R$  has an analytical expression depending upon the parameters of  $X$  and  $Y$  (see Equation (1)), and can be easily estimated e.g. through (see Downton, 1973)

$$\hat{R} = \Phi \left( \frac{\bar{x} - \bar{y}}{\sqrt{s_x^2 + s_y^2}} \right)$$

or

$$\hat{R} = \Phi \left( \frac{\bar{x} - \bar{y}}{\sqrt{\hat{\sigma}_x^2 + \hat{\sigma}_y^2}} \right),$$

which is the ML estimator of  $R$

3. draw independently a bootstrap sample  $x^*$  of size  $n$  from a r.v.  $X^*$  with the same distribution family of  $X$  and parameter  $\hat{\theta}$  and a bootstrap sample  $y^*$  of size  $m$  from a r.v.  $Y^*$  with the same distribution family of  $Y$  and parameter  $\hat{\lambda}$ . In the normal case,  $X^* \sim N(\bar{x}, \hat{\sigma}_x^2)$  (or  $X^* \sim N(\bar{x}, s_x^2)$ ) and similarly for  $Y^*$ . Since all procedures depend on only the sample means and variances, for the well-known properties of the normal distribution, for this case we only have to generate  $\bar{X}^* \sim N(\bar{x}; s_x^2/n)$  and  $S_x^{*2} \sim s_x^2 \chi_{n-1}^2/(n-1)$ ; similarly,  $\bar{Y}^*$  and  $S_y^{*2}$  can be generated.
4. estimate  $R^*$ , using the same expression for the estimator used in 2.
5. repeat steps 3. and 4.  $B$  times ( $B$  sufficiently large, e.g. 2,000), thus obtaining the bootstrap distribution  $\{\hat{R}^*\}$
6. estimate a  $(1 - \alpha)$  bootstrap percentile CI for  $R$  from  $\hat{R}^*$  distribution, taking the  $\alpha/2$  and  $1 - \alpha/2$  quantiles:

$$\left( \hat{R}_{\alpha/2}^*, \hat{R}_{1-\alpha/2}^* \right) \quad (5)$$

or a  $(1 - \alpha)$  bootstrap percentile lower bound for  $R$  as  $\hat{R}_\alpha^*$

The procedure is quite easy, even if time-consuming, due to the bootstrap step which requires a huge number of runs (2,000 is a recommended value for percentile bootstrap). Here it is applied to the normal case; indeed, it only preferably requires that the reliability has got a known expression for the two distributions; otherwise the reliability estimate  $\hat{R}$  has to be numerically computed, and the bootstrap estimates  $\hat{R}^*$  too. It is a general purpose algorithm, since it can be applied to all possible distribution families; anyway it requires an a priori parametric hypothesis about the distribution of  $X$  and  $Y$ , which is needed in order to estimate the parameters and the model reliability. The interval estimation is indeed performed only recalling the (parametric) bootstrap principle, which is independently applied to the two samples, with no concern about the different sample size, and then used to get a bootstrap distribution for  $\hat{R}^*$ , which is an estimate for the unknown distribution of  $\hat{R}$ .

Actually, only Guo and Krishnamoorthy (2004) have considered the bootstrap method for reliability estimation, specifically as a way to derive approximate test for reliability, at least for the normal case; yet, they have observed that the parametric bootstrap does not give satisfactory

results: their extensive simulation studies showed that the test based on parametric bootstrap method is liberal, i.e. its size is far greater than the nominal level. We will see that the bootstrap method can provide acceptable results, even if not as good as other approximate methods; yet it turns out as an alternative approach when all the parameters are unknown.

### 3.2 Confidence intervals based upon asymptotic results

Another way to obtain approximate confidence intervals for reliability is to recall some asymptotic results and/or to use some variance stabilizing functions. Asymptotic confidence intervals can be easily obtained. The asymptotic variance of the ML estimator  $\hat{R}$  can be derived following the same steps used in Mukherjee and Sharan (1985), and it can be estimated by

$$v(\hat{R}) = \frac{1}{2\pi s^2} e^{-(\bar{x}-\bar{y})^2/s^2} \left[ s_x^2/n + s_y^2/m + 1/2 \frac{(\bar{x}-\bar{y})^2}{s^4} (s_x^4/n + s_y^4/m) \right]$$

where  $s^2 = s_x^2 + s_y^2$ . A confidence interval for  $R$  can be then built using the variance estimate above; however, for the same reason recalled by Reiser and Guttman, it may be better to adopt some specific modifications in order to produce better CI. For example, it is possible to find first the estimate of the asymptotic variance of the ML estimator of  $\delta$ ,

$$\hat{d} = \frac{\bar{x} - \bar{y}}{\hat{\sigma}_x^2 + \hat{\sigma}_y^2},$$

that can be obtained as:

$$v(\hat{d}) = \frac{1}{s^2} \left( s_x^2/n + s_y^2/m + \frac{1}{2} \frac{(\bar{x} - \bar{y})^2}{s^4} (s_x^2/n + s_y^2/m) \right)$$

and can be employed to construct an approximate  $(1 - \alpha)$  CI for  $d$ :

$$(d_L, d_U) = (\hat{d} \pm z_{1-\alpha/2} \sqrt{v(\hat{d})})$$

and then an approximate  $(1 - \alpha)$  CI for  $R$ :

$$(\Phi(d_L), \Phi(d_U)) \tag{6}$$

or an approximate  $(1 - \alpha)$  lower bound as  $\Phi(\hat{d} - z_{1-\alpha} \sqrt{v(\hat{d})})$ .

The variance estimate of  $\hat{R}$  can be employed to build a CI for  $R$  by previously “stabilizing” it through a proper transformation. For a normalizing transformation  $g(\hat{R})$ , the approximate

variance can be obtained, according to the delta method, as

$$V(g(\hat{R})) \simeq [g'(R)]^2 V(\hat{R})$$

Then, see Krishnamoorthy and Lin (2010), for the logit transformation  $g(R) = \ln\left(\frac{R}{1-R}\right)$  we get the approximate  $(1 - \alpha)$  CI

$$(e^L(1+e^L)^{-1}, e^U(1+e^U)^{-1}) \quad (7)$$

where  $L$  and  $U$  are the lower and upper bound of the CI:

$$\left( \ln\left(\frac{\hat{R}}{1-\hat{R}}\right) \pm z_{1-\alpha/2} \frac{\sqrt{v(\hat{R})}}{\hat{R}(1-\hat{R})} \right)$$

For the arcsin transformation, the approximate  $(1 - \alpha)$  CI is given by

$$(\sin^2(L), \sin^2(U)) \quad (8)$$

where  $L$  and  $U$  are the lower and upper bound of the CI:

$$\left( \sin^{-1}\left(\sqrt{\hat{R}}\right) \pm z_{1-\alpha/2} \sqrt{\frac{v(\hat{R})}{4\hat{R}(1-\hat{R})}} \right)$$

Analogously, an approximate  $(1 - \alpha)$  lower bound can be obtained for  $R$  recalling the logit and arcsin transformations. The three latest confidence intervals are based upon the asymptotic properties of the ML estimator, so we expect they may fail or show poor results when the sample sizes  $n$  and  $m$  are small. The simulation study we present in the next Section will point out if and when this expectation is well-founded.

## 4 Simulation study

The simulation study we performed **aims** at empirically checking the statistical properties of the proposed estimators and its competitors, specifically coverage and average length at a nominal level (95%). The simulation study works as follows:

1. set the parameters for the r.v.  $X$  and  $Y$
2. compute  $R$

3. draw independently a sample  $x$  of size  $n$  from  $X$  and a sample  $y$  of size  $m$  from  $Y$
4. estimate  $R$  and a CI (either two-sided or one-sided) for  $R$
5. check out if the CI computed at the previous step contains  $R$ ; compute its length
6. repeat steps 3-5  $nSim$  (10,000) times and compute the overall CI coverage and average length

The confidence intervals empirically investigated by simulation are:

- Reiser and Guttman approximate CI (RG, Eq.3)
- Guo and Krishnamoorthy approximate CI (GK, Eq.4)
- bootstrap percentile CI (B, Eq.5)
- approximate CI based on asymptotic variance of  $d$  (AS, Eq.6)
- approximate CI based on logit transformation (LO, Eq.7)
- approximate CI based on sin transformation (SIN, Eq.8)

We considered different scenarios, each corresponding to a different combination of distributional parameters (and thus different reliabilities), reported in Table 1, and sample sizes. Without any loss in generality, we fixed  $Y \sim N(\mu_y = 0, \sigma_y = 1)$  and varied the parameters of  $X \sim N(\mu_x, \sigma_x)$ . We set the parameters in order to get a high value for the reliability, since, as we have stressed before, in real practice we usually look for high reliability for the study component/system.

parameter	values					
$\mu_x$	2	3	4	2	3	4
$\mu_y$	0	0	0	0	0	0
$\sigma_x$	1	1	1	3	3	3
$\sigma_y$	1	1	1	1	1	1
$R$	0.9214	0.9831	0.9977	0.7365	0.8286	0.8970

Table 1: Parameters for the simulation study

The Monte Carlo (MC) relative bias of  $R$  has been taken under control by checking out the condition

$$|(E_{MC}(\hat{R}) - R) / R| \leq 2\%$$



$\mu_x = 2$		$\mu_x = 3$		$\mu_x = 4$		$\mu_x = 2$		$\mu_x = 3$		$\mu_x = 4$	
$\sigma_x = 1$	$\sigma_x = 3$	$\sigma_x = 1$	$\sigma_x = 3$	$\sigma_x = 1$	$\sigma_x = 3$	$\sigma_x = 1$	$\sigma_x = 3$	$\sigma_x = 1$	$\sigma_x = 3$	$\sigma_x = 1$	$\sigma_x = 3$
$n = 10, m = 10$						$n = 20, m = 50$					
+0.257	+1.298	-0.167	+1.054	-0.100	+0.338	+0.083	+0.804	-0.057	+0.544	-0.048	+0.118
$n = 10, m = 20$						$n = 50, m = 10$					
+0.152	+1.488	-0.115	+1.146	+0.588	+0.588	+0.087	+0.477	-0.091	+0.175	-0.064	+0.343
$n = 10, m = 50$						$n = 50, m = 20$					
+0.130	+1.236	-0.107	+0.995	-0.067	+0.733	+0.096	+0.284	-0.042	+0.256	-0.051	+0.123
$n = 20, m = 10$						$n = 50, m = 50$					
+0.246	+0.675	-0.084	+0.670	-0.080	+0.420	+0.065	+0.243	-0.038	+0.253	-0.034	+0.000
$n = 20, m = 20$											
+0.131	+0.431	-0.040	+0.279	-0.040	+0.683						

Table 2: Simulation results:  $\hat{R}$  percentage relative bias

Actually, the reliability estimator employed is not unbiased, and the MC relative error is an estimate of its relative bias  $(E(\hat{R}) - R)/R$ .

The results of the simulation study are reported in Tables 3 and 4 (coverage and average length of the two-sided CIs) and Tables 5 and 6 (coverage and average lower bound of one-sided CIs). The MC percentage relative bias of  $\hat{R}$  is reported in Table 2.

With regard to the behavior of  $\hat{R}$ , even if the ML estimator is only asymptotically unbiased, its MC bias is here shown to be negligible, since its absolute value is always smaller than 1.5%. As we would have expected, sample sizes affect it significantly: for  $n = m = 50$  it decreases to 0.25%. The MC bias is also strongly affected by the experimental conditions: in particular, it takes negative values when the true  $R$  is closer to 1, while takes the greatest positive values when  $R$  is nearer to 0.5.

Looking at the results for two-sided CIs, it is clearly visible that Reiser and Guttman inferential procedure gives very good results, which seem not to be affected by the samples' sizes and by their discrepancy, even if one can note a slight departure of coverage rate from the nominal level when  $n = 10$  and  $m = 50$ . The asymptotic normal CI provide very good results too, in terms of coverage, which is always very close to the nominal level, yet its performance is a bit worse than Reiser and Guttman's. The bootstrap proposal show on average a good performance, assuring a coverage always greater than 90% even for small samples; moreover, it is the method that provides, on average, the shortest confidence interval, under each experimental

method	$\mu_x = 2$		$\mu_x = 3$		$\mu_x = 4$		method	$\mu_x = 2$		$\mu_x = 3$		$\mu_x = 4$	
	$\sigma_x = 1$	$\sigma_x = 3$	$\sigma_x = 1$	$\sigma_x = 3$	$\sigma_x = 1$	$\sigma_x = 3$		$\sigma_x = 1$	$\sigma_x = 3$	$\sigma_x = 1$	$\sigma_x = 3$	$\sigma_x = 1$	$\sigma_x = 3$
$n = 10, m = 10$							$n = 20, m = 50$						
RG	0.954	0.943	0.957	0.950	0.959	0.951	RG	0.952	0.949	0.951	0.945	0.957	0.949
B	0.910	0.896	0.899	0.907	0.894	0.901	B	0.935	0.927	0.926	0.922	0.939	0.926
AS	0.950	0.939	0.952	0.947	0.949	0.950	AS	0.951	0.948	0.948	0.942	0.956	0.948
LO	0.978	0.973	0.983	0.970	0.994	0.983	LO	0.960	0.953	0.970	0.954	0.974	0.967
SIN	0.922	0.911	0.906	0.923	0.887	0.922	SIN	0.936	0.935	0.920	0.927	0.926	0.938
$n = 10, m = 20$							$n = 50, m = 10$						
RG	0.950	0.944	0.949	0.945	0.952	0.944	RG	0.938	0.951	0.940	0.950	0.942	0.949
B	0.915	0.903	0.897	0.899	0.891	0.899	B	0.905	0.936	0.899	0.935	0.889	0.930
AS	0.947	0.941	0.945	0.942	0.945	0.939	AS	0.935	0.950	0.935	0.949	0.932	0.948
LO	0.977	0.963	0.973	0.960	0.988	0.978	LO	0.957	0.953	0.974	0.954	0.987	0.957
SIN	0.923	0.919	0.918	0.915	0.874	0.909	SIN	0.911	0.941	0.894	0.938	0.869	0.933
$n = 10, m = 50$							$n = 50, m = 20$						
RG	0.936	0.945	0.937	0.943	0.951	0.942	RG	0.952	0.945	0.949	0.948	0.959	0.953
B	0.909	0.902	0.900	0.897	0.912	0.895	B	0.931	0.936	0.927	0.937	0.933	0.954
AS	0.933	0.942	0.932	0.938	0.944	0.939	AS	0.950	0.946	0.947	0.948	0.956	0.951
LO	0.953	0.958	0.972	0.968	0.988	0.975	LO	0.962	0.946	0.969	0.953	0.984	0.959
SIN	0.915	0.919	0.896	0.913	0.891	0.912	SIN	0.932	0.938	0.919	0.940	0.916	0.951
$n = 20, m = 10$							$n = 50, m = 50$						
RG	0.951	0.947	0.947	0.954	0.947	0.926	RG	0.953	0.951	0.948	0.950	0.951	0.936
B	0.911	0.927	0.901	0.928	0.901	0.905	B	0.937	0.943	0.939	0.941	0.945	0.936
AS	0.947	0.947	0.943	0.953	0.943	0.925	AS	0.952	0.951	0.947	0.949	0.951	0.936
LO	0.972	0.952	0.983	0.965	0.990	0.961	LO	0.958	0.953	0.961	0.953	0.965	0.942
SIN	0.919	0.932	0.900	0.936	0.881	0.907	SIN	0.939	0.946	0.935	0.944	0.937	0.939
$n = 20, m = 20$													
RG	0.948	0.943	0.954	0.949	0.966	0.953							
B	0.926	0.927	0.920	0.929	0.924	0.906							
AS	0.945	0.942	0.950	0.948	0.961	0.949							
LO	0.965	0.947	0.982	0.948	0.992	0.972							
SIN	0.929	0.932	0.916	0.939	0.905	0.915							

Table 3: Simulation results: CI coverage rate

method	$\mu_x = 2$		$\mu_x = 3$		$\mu_x = 4$		method	$\mu_x = 2$		$\mu_x = 3$		$\mu_x = 4$	
	$\sigma_x = 1$	$\sigma_x = 3$	$\sigma_x = 1$	$\sigma_x = 3$	$\sigma_x = 1$	$\sigma_x = 3$		$\sigma_x = 1$	$\sigma_x = 3$	$\sigma_x = 1$	$\sigma_x = 3$	$\sigma_x = 1$	$\sigma_x = 3$
$n = 10, m = 10$							$n = 20, m = 50$						
RG	0.243	0.402	0.123	0.349	0.052	0.292	RG	0.136	0.288	0.055	0.247	0.017	0.200
B	0.198	0.398	0.080	0.317	0.026	0.240	B	0.125	0.287	0.046	0.236	0.012	0.181
AS	0.238	0.399	0.118	0.344	0.048	0.285	AS	0.135	0.287	0.054	0.245	0.016	0.198
LO	0.294	0.409	0.271	0.374	0.422	0.345	LO	0.142	0.289	0.063	0.252	0.023	0.211
SIN	0.213	0.395	0.088	0.329	0.028	0.259	SIN	0.130	0.286	0.047	0.241	0.012	0.189
$n = 10, m = 20$							$n = 50, m = 10$						
RG	0.206	0.393	0.096	0.341	0.036	0.283	RG	0.181	0.222	0.081	0.186	0.029	0.144
B	0.175	0.389	0.068	0.310	0.020	0.233	B	0.157	0.220	0.060	0.181	0.017	0.135
AS	0.203	0.389	0.093	0.336	0.034	0.277	AS	0.178	0.221	0.079	0.186	0.027	0.143
LO	0.230	0.398	0.144	0.361	0.129	0.328	LO	0.194	0.222	0.105	0.189	0.060	0.148
SIN	0.186	0.387	0.072	0.323	0.020	0.252	SIN	0.167	0.221	0.063	0.184	0.017	0.139
$n = 10, m = 50$							$n = 50, m = 20$						
RG	0.180	0.388	0.082	0.337	0.030	0.278	RG	0.136	0.204	0.055	0.172	0.017	0.135
B	0.156	0.383	0.060	0.307	0.018	0.228	B	0.125	0.204	0.045	0.168	0.012	0.128
AS	0.177	0.384	0.079	0.332	0.028	0.272	AS	0.135	0.204	0.054	0.172	0.016	0.134
LO	0.193	0.392	0.105	0.356	0.061	0.318	LO	0.142	0.204	0.062	0.174	0.023	0.138
SIN	0.166	0.382	0.064	0.320	0.018	0.248	SIN	0.129	0.204	0.047	0.170	0.012	0.131
$n = 20, m = 10$							$n = 50, m = 50$						
RG	0.205	0.308	0.095	0.263	0.037	0.209	RG	0.102	0.192	0.038	0.163	0.010	0.129
B	0.174	0.308	0.067	0.250	0.021	0.186	B	0.097	0.192	0.034	0.160	0.008	0.123
AS	0.201	0.307	0.092	0.261	0.035	0.206	AS	0.101	0.192	0.038	0.162	0.010	0.128
LO	0.229	0.310	0.143	0.270	0.131	0.223	LO	0.104	0.192	0.041	0.164	0.012	0.131
SIN	0.185	0.307	0.071	0.256	0.021	0.195	SIN	0.099	0.192	0.035	0.161	0.008	0.126
$n = 20, m = 20$													
RG	0.165	0.296	0.070	0.255	0.022	0.200							
B	0.148	0.296	0.054	0.243	0.015	0.179							
AS	0.163	0.295	0.068	0.253	0.022	0.198							
LO	0.176	0.298	0.087	0.261	0.039	0.212							
SIN	0.154	0.295	0.056	0.249	0.014	0.187							

Table 4: Simulation results: CI average width

method	$\mu_x = 2$		$\mu_x = 3$		$\mu_x = 4$		method	$\mu_x = 2$		$\mu_x = 3$		$\mu_x = 4$	
	$\sigma_x = 1$	$\sigma_x = 3$	$\sigma_x = 1$	$\sigma_x = 3$	$\sigma_x = 1$	$\sigma_x = 3$		$\sigma_x = 1$	$\sigma_x = 3$	$\sigma_x = 1$	$\sigma_x = 3$	$\sigma_x = 1$	$\sigma_x = 3$
$n = 10, m = 10$							$n = 20, m = 50$						
RG	0.958	0.949	0.958	0.943	0.951	0.951	RG	0.956	0.943	0.952	0.946	0.956	0.949
GK	0.964	0.952	0.964	0.950	0.958	0.957	GK	0.959	0.947	0.956	0.952	0.959	0.950
B	0.887	0.891	0.863	0.874	0.851	0.850	B	0.919	0.914	0.924	0.898	0.920	0.889
AS	0.952	0.945	0.953	0.938	0.941	0.947	AS	0.953	0.942	0.950	0.943	0.950	0.948
LO	1.000	0.960	1.000	0.984	1.000	1.000	LO	0.962	0.945	0.970	0.955	0.981	0.958
SIN	0.906	0.908	0.883	0.895	0.860	0.889	SIN	0.922	0.926	0.919	0.917	0.904	0.901
$n = 10, m = 20$							$n = 50, m = 10$						
RG	0.947	0.940	0.945	0.936	0.940	0.939	RG	0.926	0.950	0.926	0.950	0.917	0.950
GK	0.952	0.944	0.953	0.941	0.948	0.949	GK	0.946	0.960	0.939	0.955	0.929	0.957
B	0.883	0.879	0.858	0.872	0.850	0.865	B	0.876	0.926	0.855	0.917	0.848	0.917
AS	0.942	0.934	0.942	0.930	0.936	0.932	AS	0.923	0.950	0.922	0.948	0.914	0.950
LO	0.977	0.954	1.000	0.975	1.000	0.992	LO	0.946	0.951	0.965	0.954	0.989	0.960
SIN	0.894	0.906	0.869	0.896	0.848	0.887	SIN	0.885	0.935	0.856	0.923	0.83	0.923
$n = 10, m = 50$							$n = 50, m = 20$						
RG	0.944	0.947	0.941	0.941	0.938	0.939	RG	0.950	0.954	0.948	0.955	0.949	0.954
GK	0.953	0.951	0.952	0.950	0.955	0.952	GK	0.953	0.959	0.952	0.958	0.953	0.959
B	0.890	0.889	0.879	0.874	0.864	0.859	B	0.912	0.923	0.900	0.916	0.895	0.913
AS	0.939	0.945	0.939	0.939	0.930	0.932	AS	0.949	0.954	0.946	0.955	0.948	0.953
LO	0.956	0.953	0.970	0.966	0.987	0.984	LO	0.956	0.954	0.974	0.958	0.983	0.961
SIN	0.901	0.913	0.879	0.897	0.850	0.889	SIN	0.917	0.934	0.891	0.926	0.872	0.921
$n = 20, m = 10$							$n = 50, m = 50$						
RG	0.942	0.949	0.943	0.948	0.946	0.945	RG	0.962	0.956	0.957	0.957	0.954	0.953
GK	0.952	0.951	0.952	0.952	0.959	0.948	GK	0.962	0.959	0.960	0.959	0.957	0.956
B	0.875	0.908	0.857	0.894	0.856	0.875	B	0.940	0.939	0.937	0.927	0.932	0.923
AS	0.934	0.947	0.937	0.946	0.941	0.943	AS	0.961	0.956	0.956	0.956	0.954	0.950
LO	0.980	0.951	1.000	0.958	1.000	0.968	LO	0.965	0.956	0.966	0.957	0.976	0.957
SIN	0.891	0.923	0.864	0.912	0.848	0.899	SIN	0.941	0.944	0.924	0.935	0.913	0.929
$n = 20, m = 20$													
RG	0.948	0.932	0.950	0.933	0.953	0.934							
GK	0.950	0.935	0.955	0.939	0.954	0.936							
B	0.888	0.896	0.889	0.877	0.886	0.869							
AS	0.947	0.931	0.948	0.926	0.949	0.930							
LO	0.962	0.935	0.984	0.948	1.000	0.953							
SIN	0.897	0.908	0.893	0.893	0.871	0.883							

Table 5: Simulation results: one-sided CI coverage rate

method	$\mu_x = 2$		$\mu_x = 3$		$\mu_x = 4$		method	$\mu_x = 2$		$\mu_x = 3$		$\mu_x = 4$	
	$\sigma_x = 1$	$\sigma_x = 3$	$\sigma_x = 1$	$\sigma_x = 3$	$\sigma_x = 1$	$\sigma_x = 3$		$\sigma_x = 1$	$\sigma_x = 3$	$\sigma_x = 1$	$\sigma_x = 3$	$\sigma_x = 1$	$\sigma_x = 3$
$n = 10, m = 10$							$n = 20, m = 50$						
RG	0.775	0.535	0.902	0.637	0.963	0.725	RG	0.845	0.597	0.948	0.701	0.987	0.787
GK	0.769	0.530	0.897	0.631	0.960	0.717	GK	0.842	0.594	0.947	0.697	0.986	0.784
B	0.817	0.565	0.934	0.681	0.980	0.776	B	0.858	0.612	0.956	0.723	0.990	0.811
AS	0.778	0.537	0.905	0.641	0.965	0.729	AS	0.845	0.597	0.949	0.702	0.987	0.788
LO	0.749	0.533	0.827	0.625	0.746	0.696	LO	0.843	0.597	0.945	0.699	0.984	0.782
SIN	0.806	0.556	0.927	0.667	0.977	0.759	SIN	0.856	0.607	0.956	0.715	0.990	0.804
$n = 10, m = 20$							$n = 50, m = 10$						
RG	0.802	0.542	0.921	0.645	0.973	0.732	RG	0.820	0.634	0.932	0.739	0.978	0.823
GK	0.795	0.536	0.916	0.637	0.971	0.724	GK	0.811	0.630	0.926	0.735	0.975	0.820
B	0.833	0.571	0.943	0.689	0.984	0.783	B	0.844	0.644	0.949	0.750	0.986	0.835
AS	0.804	0.544	0.923	0.648	0.975	0.737	AS	0.822	0.635	0.933	0.739	0.979	0.823
LO	0.790	0.540	0.896	0.635	0.933	0.709	LO	0.814	0.635	0.920	0.738	0.964	0.822
SIN	0.826	0.562	0.939	0.673	0.983	0.766	SIN	0.839	0.640	0.946	0.747	0.986	0.832
$n = 10, m = 50$							$n = 50, m = 20$						
RG	0.820	0.547	0.932	0.650	0.978	0.737	RG	0.847	0.643	0.949	0.746	0.987	0.828
GK	0.811	0.540	0.926	0.641	0.975	0.728	GK	0.844	0.642	0.947	0.745	0.986	0.827
B	0.844	0.576	0.949	0.693	0.986	0.786	B	0.860	0.651	0.957	0.756	0.990	0.839
AS	0.822	0.549	0.934	0.653	0.979	0.741	AS	0.847	0.644	0.949	0.746	0.987	0.829
LO	0.814	0.546	0.920	0.641	0.964	0.716	LO	0.845	0.644	0.945	0.746	0.984	0.827
SIN	0.839	0.567	0.946	0.678	0.986	0.770	SIN	0.857	0.649	0.956	0.753	0.990	0.836
$n = 20, m = 10$							$n = 50, m = 50$						
RG	0.804	0.592	0.922	0.697	0.974	0.784	RG	0.867	0.647	0.960	0.749	0.991	0.831
GK	0.796	0.589	0.917	0.693	0.971	0.781	GK	0.867	0.647	0.960	0.748	0.991	0.830
B	0.834	0.610	0.944	0.721	0.985	0.810	B	0.874	0.654	0.964	0.758	0.993	0.841
AS	0.806	0.593	0.924	0.698	0.975	0.785	AS	0.867	0.648	0.960	0.750	0.991	0.831
LO	0.792	0.592	0.924	0.694	0.933	0.777	LO	0.867	0.648	0.959	0.749	0.990	0.830
SIN	0.827	0.604	0.940	0.713	0.984	0.803	SIN	0.873	0.652	0.964	0.755	0.993	0.838
$n = 20, m = 20$													
RG	0.831	0.602	0.940	0.705	0.983	0.791							
GK	0.829	0.599	0.939	0.702	0.982	0.788							
B	0.850	0.618	0.952	0.728	0.988	0.816							
AS	0.832	0.602	0.941	0.706	0.983	0.792							
LO	0.826	0.602	0.931	0.703	0.975	0.786							
SIN	0.847	0.613	0.951	0.721	0.988	0.809							

Table 6: Simulation results: one-sided CI average lower bound

1 condition. The little departure of coverage from the nominal level is a known issue of bootstrap  
2 interval estimation, see Schreuder and Williams (2000). The CI based on logit transformation  
3 gives always coverages greater than the nominal level, especially for high values of  $R$ ; on the  
4 other hand, its average width is the largest. The CI based on sin transformation almost always  
5 give coverages never greater than 95%, even for the largest samples' sizes; its performance in  
6 terms of both coverage and average width looks surprisingly similar to the bootstrap interval  
7 estimator. It is worthwhile noting that the two approximate methods proposed (LO and SIN),  
8 based on asymptotic results, perform well even for small sample sizes. It is also interesting to  
9 observe that for sample sizes moderately large ( $n = m = 50$ , depending upon the other param-  
10 eters) the performances of the different confidence intervals tend to get closer, in terms of both  
11 coverage ( $\simeq 95\%$ ) and average length.  
12

13  
14  
15  
16  
17  
18  
19  
20  
21 From the simulation study, it is interesting to note if and how the value of  $R$  and the sample  
22 sizes  $n$  and  $m$  affect the performance of the interval estimators. With regard to the coverage,  
23 as already said, the samples' sizes and their discrepancy seem not to heavily affect the results  
24 (but, as expected, the overall best results are for  $n = m = 50$ ); the same can be told about  
25 the experimental conditions: the values of  $\mu_x$  and  $\sigma_x$  seem not to systematically affect the  
26 performances of the compared procedures. Even in the "worst" scenario ( $\mu_x = 4$ ,  $\sigma_x = 1$ , with  
27  $R$  very close to 1) the methods keep working satisfactorily. Yet LO, for some scenarios, namely  
28  $\mu_x = 4$  and  $\sigma_x = 1$ , provides unreliable results. With regard to the average width, this is clearly  
29 influenced by the sample sizes and experimental conditions. The results get better (i.e. the CI  
30 are more precise) as the sample sizes increase and the reliability  $R$  gets closer to 1.  
31  
32  
33  
34  
35  
36  
37

38 From the simulation results for one-sided confidence intervals, the overall superiority of  
39 GK procedure emerges, followed closed behind by RG and LO. The LO intervals here, in  
40 fact, allows almost always coverages greater than the nominal level, even if in some cases  
41 they are too conservative, providing also smaller lower bounds. The lower bounds based on  
42 asymptotic results provide satisfactory results, which only partially suffer from the finiteness  
43 of sample sizes. The bootstrap procedure gives worse results than in the two-sided cases: it  
44 attains coverage rates close to the nominal level only for large sample sizes, whereas they are  
45 often smaller than 90%; it tends to give higher lower bounds than its competitors. Its "crude"  
46 use is not suggested; some proper modifications should be made, but the practitioner should  
47  
48  
49  
50  
51  
52  
53  
54

1 also take into account the additional computation time this already time consuming method  
2 could bring. With regard to the SIN procedure, its performance is not totally satisfactory and  
3 reliable. One should also note an overall light worsening of the performance of asymptotic  
4 procedures passing from two-sided to one-sided CIs: this may be ascribed to the fact that  
5 asymptotic CIs overcover on one-tail and undercover on the other, and so overall coverage  
6 probabilities of asymptotic two-sided CIs turn out to be close to the nominal level  $1 - \alpha$ . This  
7 behavior has been detected by Krishnamoorthy and Lin (2010) for the Weibull case.  
8  
9  
10  
11  
12  
13

## 14 **5 An example of application**

15  
16 Table 7, taken from Duncan (1986), gives the results of measuring shear strength for spot  
17 welds for two different gauges of steel. Although not a stress-strength situation in the sense  
18 discussed in the introduction, this problem does fit our formulation, and we use it to show how  
19 the inferential procedures work on a practical case. Denoting with  $X$  and  $Y$  the r.v. modeling  
20 the strength of the two gauges of steel, the estimates computed on the data are  $\bar{x} = 975$ ,  $\bar{y} =$   
21  $480$ ,  $\hat{\sigma}_x^2 = 33055$  and  $\hat{\sigma}_y^2 = 7345$ . Assuming  $X$  and  $Y$  are independently normally distributed,  
22 the maximum likelihood estimator  $\hat{R}$  has value 0.9931. The confidence intervals calculated  
23 according each of the methods described in the previous sections are reported in Table 8, in  
24 Table 9 the one-sided confidence intervals; they show that the proposed bootstrap methods  
25 provides the narrowest interval, confirming what we have already seen in the simulation study,  
26 while the LO method provides the largest one. RG, GK and AS bounds are very close to each  
27 other.  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41

## 42 **6 Conclusions**

43  
44 In this paper we propose a bootstrap approach and some approximate methods for the interval  
45 estimation of reliability in a stress-strength model, where both stress and strength are normally  
46 distributed with unknown parameters. We empirically investigate the performance of these  
47 methods and other existing ones, through a simulation study comprising different artificial  
48 scenarios. The results show the superiority (in particular when building one-sided tolerance  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

$Y (m = 10)$	$X (n = 10)$
350	680
380	800
385	780
450	885
465	975
485	1,025
535	1,100
555	1,030
590	1,175
605	1,300

Table 7: Data on shear strengths of two gauges of steel

method	90%		95%		99%	
	lower bound	upper bound	lower bound	upper bound	lower bound	upper bound
RG	0.9216	0.9994	0.8925	0.9997	0.8147	0.9999
AS	0.9262	0.9994	0.8993	0.9997	0.8277	0.9999
SIN	0.9511	1.0000	0.9389	1.0000	0.9115	1.0000
LO	0.8308	0.9998	0.7200	0.9999	0.4205	1.0000
B	0.9587	0.9999	0.9443	1.0000	0.9266	1.0000

Table 8: Results for reliability on Duncan data: 90, 95, and 99% confidence intervals.



method	90%	95%	99%
RG	0.9473	0.9216	0.8496
GK	0.9416	0.9154	0.8467
AS	0.9499	0.9262	0.8599
SIN	0.9636	0.9511	0.9232
LO	0.9120	0.8308	0.5478
B	0.9685	0.9569	0.9303

Table 9: 90, 95, and 99% lower bounds for reliability on Duncan data.

bounds) of two existing methods specifically conceived for the normal set-up. Our bootstrap proposal performs slightly worse than the other approximate methods in terms of coverage, but provides narrower intervals. Yet, the advantage of the bootstrap method versus its competitors lies in the actual computation of CI, which does not require further formulas or approximations, but simply reuses the formula employed for point estimation. With regard to the estimators based on asymptotic results, their performance has been here satisfactory even for moderate sample sizes, however, a particular attention should be devoted if extending these results to other parametric families. The new bootstrap proposal can be easily extended, especially to the other family distributions for which an expression for the reliability  $R$  is analytically derivable; it can be used in the independent set-up, as described here, but also generalized - with proper modifications - to the dependent set-up: future research shall then investigate the performance of the bootstrap interval estimator towards the existing ones for other parametric univariate and bivariate families.

## Acknowledgments

I would like to thank the referees for their helpful comments and suggestions.

I wish to dedicate this paper to the cherished memory of my mother and to my father.

## References

- 1  
2  
3  
4 Birnbaum, Z. W. (1956). On a use of the Mann-Whitney statistic. Proceedings of the Third  
5  
6 Berkeley Symposium on Mathematical Statistics and Probability, Vol. 1, Univ. of Calif.  
7  
8 Press, 13-17.  
9
- 10 Birnbaum, Z. W., McCarty, R. C. (1958). A distribution-free upper confidence bound for  $P(Y <$   
11  
12  $X)$  based on independent samples of  $X$  and  $Y$ . *Annals of Mathematical Statistics* 29: 558-  
13  
14 562.  
15
- 16 Chang D. S. (1995) Reliability bounds for the stress-strength model. *Computers & Industrial*  
17  
18 *Engineering* 29(1): 15-19  
19
- 20 Church, J.D., Harris, B. (1970). The Estimation of Reliability from Stress-Strength Relation-  
21  
22 ships. *Technometrics* 12(1): 551-558  
23  
24
- 25 Downton, F. (1973). The Estimation of  $Pr(Y < X)$  in the Normal Case. *Technometrics* 15(3):  
26  
27 551-558  
28
- 29 Efron, B., Tibshirani, R. (1994) *An Introduction to the Bootstrap*. Chapman & Hall/CRC  
30  
31
- 32 Duncan, A. G. (1986). *Quality and Industrial Statistics* (5th ed.). Irwin, Homewood, IL  
33  
34
- 35 Guo, H., Krishnamoorthy, K. (2004). New Approximate Inferential Methods for the Reliability  
36  
37 Parameter in a Stress-Strength Model: The Normal Case. *Communications in Statistics -*  
38  
39 *Theory and Methods* 33(7): 1715-1731  
40
- 41 Hall, I.J. (1984). Approximate one-sided tolerance limits for the difference or sum of two  
42  
43 independent normal variates. *Journal of Quality Technology* 16: 15-19  
44
- 45 Harris, B., Soms, A. P. (1991) The Theory of Optimal Confidence Limits for Systems Relia-  
46  
47 bility with Counterexamples for Results on Optimal Confidence Limits for Series Systems.  
48  
49 *Statistics & Probability Letters* 11(5): 411-417  
50
- 51 Kotz, S., Lumelskii, Y., Pensky, M. (2003) *The stress-strength model and its generalizations -*  
52  
53 *Theory and Applications*. World Scientific Publishing  
54

- 1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54
- Krishnamoorthy, K., Lin, Y. (2010). Confidence Limits for Stress-Strength Reliability Involving Weibull Models. *Journal of Statistical Planning and Inference* 140(7): 1754-1764
- Mukherjee, S. P., Maiti, S. S. (1998). Stress-strength reliability in the Weibull case. In: *Frontiers In Reliability* 4: 231-248. World Scientific
- Mukherjee, S.P., Sharan, L.K. (1985). Estimation of failure probability from a bivariate normal stress-strength distribution. *Microelectronics Reliability* 25: 699-702
- Nadarajah, S., Kotz, S. (2005). Reliability For A Bivariate Gamma Distribution. *Economic Quality Control* 20(1): 111-119
- Reiser, B., Guttman, I. (1986). Statistical Inference for  $Pr(Y < X)$ : The Normal Case. *Technometrics*, 28(3): 253-257
- Roy, D. (1993) Estimation of failure probability under a bivariate normal stress-strength distribution. *Microelectronics Reliability* 33(15): 2285-2287
- Schreuder, H. T., Williams, M.S. (2000). Reliability of Confidence Intervals Calculated by Bootstrap and Classical Methods Using the FIA 1-ha Plot Design. General Technical Report RMRS-GTR-57. Fort Collins, U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station.
- Weerahandi, S., Johnson, R. A. (1992). Testing Reliability in a Stress-Strength Model When  $X$  and  $Y$  Are Normally Distributed. *Technometrics* 34(1): 83-91