

A new reliability measure in ranked set sampling

M. Mahdizadeh¹  · Ehsan Zamanzade² 

Received: 2 March 2015 / Revised: 6 April 2016 / Published online: 5 July 2016
© Springer-Verlag Berlin Heidelberg 2016

Abstract In this article, a dynamic reliability measure based on ranked set sampling is introduced, and its properties are investigated in theory and simulation. The results support the preference of the suggested index over the analogous one in simple random sampling. A data set from an agricultural experiment is analyzed for illustration.

Keywords Judgment ranking · Life testing · Stress-strength model

Mathematics Subject Classification 62G30 · 62N05

1 Introduction

Ranked set sampling (RSS) is a sampling design that often leads to improved statistical inference as compared with simple random sampling (SRS). It is appropriate when the sampling units are expensive and/or difficult to measure but are reasonably simple and cheap to order according to the variable of interest. Ranking can be based on concomitant variables, expert judgment or any mean that does not involve actual measurement.

A basic description of the RSS procedure is as follows: n^2 units are collected as iid draws from the population. These units are randomly partitioned into n sets,

✉ M. Mahdizadeh
mahdizadeh.m@live.com

Ehsan Zamanzade
e.zamanzade@sci.ui.ac.ir; ehsanzamanzadeh@yahoo.com

¹ Department of Statistics, Hakim Sabzevari University, P.O. Box 397, Sabzevar, Iran

² Department of Statistics, University of Isfahan, Isfahan 81746-73441, Iran

each of size n . In the first set, the response judged to be smallest is taken for full quantification; In the second set, the response judged to be second smallest is taken; and so on, until in the last set, the response judged to be largest is taken. The full measurement, along with the associated ranks form a ranked set sample of size n . Let $X_{[i]}$ be the i th judgement order statistic from the i th set. (The i th true order statistics is represented by $X_{(i)}$ for clarity.) Then, the resulting ranked set sample is denoted by $X_{[1]}, \dots, X_{[n]}$. A ranked set sample, consisting of independent order statistics, is more informative than a simple random sample of the same size. The results in the literature have shown that statistical procedures based on RSS tend to be superior to their SRS analogues. For a book-length treatment of RSS and its applications, see [Chen et al. \(2004\)](#).

The RSS was first introduced by [McIntyre \(1952\)](#) in an agricultural experiment for estimating the mean pasture yield. Since then, it has been well adopted to environmental, ecological and health studies. [Kvam and Samaniego \(1994\)](#) provided an example in reliability context. Consider the lifetime of a k -out-of- n system in which components have independent and identically distributed lifetimes is equal to the $(n - k + 1)$ th order statistic from a sample of size n . When independent lifetimes of several k -out-of- n systems with varying k and n are observed, the experimental data constitute a ranked set sample. In addition, when observations consist of lifetimes of coherent systems together with the number of failed components, the data are stochastically equivalent to a collection of independently observed order statistics. Thus RSS arise naturally in life testing experiments in which one observes the system lifetime rather than the lifetimes of the individual components.

The estimation of system reliability has drawn much attention in the statistical literature. The most widely used approach for this purpose is the well-known stress-strength model. It employs $R = P(X > Y)$ as an index to quantify reliability of a component with strength X which is subjected to stress Y . The estimation of R has been extensively investigated in the literature when X and Y are independent variables, and belong to the same family of distributions. A comprehensive account of this topic appear in [Kotz et al. \(2003\)](#). For example, [Díaz-Francis and Montoya \(2013\)](#) discuss the profile likelihood method in statistical inference for R when X and Y are independent exponential random variables.

Given the fact ranked set sample data could represent lifetime, it is relevant to research on reliability measures in RSS scheme. Interpretation of such indices in other disciplines (as exemplified in Sect. 6) would be an additional support. [Sengupta and Mukhuti \(2008\)](#) studied unbiased estimation of R based on RSS in nonparametric setting. They showed that the proposed estimator is more efficient than its SRS rival, even in the presence of ranking errors. We intend to introduce a time-dependent reliability measure, and to investigate its theoretical properties.

Section 2 presents the new measure along with some notions and results which will be used in the sequel. Some theoretical results are established in Sect. 3. Perfect ranking setup is discussed in Sect. 4. Results of Monte Carlo experiment conducted to get insight of the behavior of the proposed index appear in Sect. 5. An application is discussed in Sect. 6. A summary and direction for future research are given in Sect. 7. Figures are postponed to an appendix.

2 Preliminaries

In reliability theory, there are several approaches to compare lifetimes of two components. Among the most popular methods are to compare the survival functions, the failure rates and the mean residual lifetime functions. Let the random variables X and Y be the lifetimes of two systems. The density, distribution and survival function of $X(Y)$ are denoted by $f(g)$, $F(G)$ and $\bar{F} = 1 - F(\bar{G} = 1 - G)$, respectively. Assume that both systems are operating at time $t > 0$. Then the residual lifetimes of them are $X_t = (X - t|X > t)$ and $Y_t = (Y - t|Y > t)$, respectively. Taking into account the age of systems, Zardasht and Asadi (2010) introduced a time-dependent criterion to compare the two residual lifetimes. They considered function $R(t) = P(X_t > Y_t)$. Note that $R(t)$ can be written as

$$R(t) = \frac{R_1(t)}{R_2(t)},$$

where $R_1(t) = P(X > Y > t)$ and $R_2(t) = P(X > t, Y > t)$. Using simple random samples X_1, \dots, X_m and Y_1, \dots, Y_n from F and G , an estimate of $R(t)$ can be made as

$$\hat{R}(t) = \frac{\hat{R}_1(t)}{\hat{R}_2(t)},$$

where

$$\hat{R}_1(t) = \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n I(X_i > Y_j > t)$$

and

$$\hat{R}_2(t) = \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n I(X_i > t, Y_j > t).$$

Although $\hat{R}_1(t)$ and $\hat{R}_2(t)$ are respectively unbiased estimators of $R_1(t)$ and $R_2(t)$, $\hat{R}(t)$ is only asymptotically unbiased.

Let $X_{[1]}, \dots, X_{[m]}$ and $Y_{[1]}, \dots, Y_{[n]}$ be two ranked set samples from F and G , respectively. The estimator of $R(t)$ in RSS is given by

$$\hat{R}^*(t) = \frac{\hat{R}_1^*(t)}{\hat{R}_2^*(t)},$$

where

$$\hat{R}_1^*(t) = \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n I(X_{[i]} > Y_{[j]} > t)$$

and

$$\hat{R}_2^*(t) = \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n I(X_{[i]} > t, Y_{[j]} > t).$$

When $t = -\infty$, the above estimator reduces to that proposed by [Sengupta and Mukhuti \(2008\)](#). In the next section, we study properties of $\hat{R}^*(t)$. It is to be noted that $\hat{R}(t)$ and $\hat{R}^*(t)$ are defined for $t \in (-\infty, t^*)$, where $t^* = \min\{t_X, t_Y\}$ with t_X and t_Y being the supremum of the set of values in the support of X and Y , respectively. However, in practice, there may be situations that all sample units of X and/or Y are less than the value of t . In such cases, the numerator and denominator of the corresponding estimator are zero, and we define it to be zero at t .

As the ranking process in RSS is done either by expert judgment or using an easily available covariate, it need not match the true ranking according to the variable of interest. This situation is called imperfect ranking. It is unavoidable because true orders are unknown unless all units are actually measured. There are certain models that allow for considering such cases. We build on an imperfect ranking model introduced by [Bohn and Wolfe \(1994\)](#). It is now briefly described for the two-sample problem mentioned above.

The density, distribution and survival function of the i th true (judgement) order statistic of a random sample of size m from F are denoted by $f_{(i)}(f_{[i]})$, $F_{(i)}(F_{[i]})$, and $\bar{F}_{(i)}(\bar{F}_{[i]})$, respectively. Similar notations are used for a random sample of size n from G . We postulate an imperfect ranking model M_X under which $X_{[i]}$'s are assumed to be independently distributed as

$$P(X_{[i]} = X_{(r)}) = p_{ir}, \quad (r = 1, \dots, m),$$

where p_{ir} is the probability that the r th order statistic is judged to have rank i , and thus $\sum_{r=1}^m p_{ir} = 1$. It is further assumed that $\sum_{i=1}^m p_{ir} = 1$. Obviously, this is true in the perfect ranking scenario, i.e. when $p_{ii} = 1$ and $p_{ir} = 0$ ($r \neq i$). Similarly, we postulate an imperfect ranking model M_Y under which $Y_{[j]}$'s are assumed to be independently distributed as

$$P(Y_{[j]} = Y_{(s)}) = q_{js}, \quad (s = 1, \dots, n),$$

where q_{js} is the probability that the s th order statistic is judged to have rank j , and therefore $\sum_{s=1}^n q_{js} = 1$. Moreover, it is assumed that $\sum_{j=1}^n q_{js} = 1$. The model considering M_X and M_Y together is referred to as M .

We close this section by pointing out some results which are repeatedly used in this work. According to a basic identity in RSS,

$$\frac{1}{m} \sum_{i=1}^m F_{(i)}(x) = F(x), \quad \frac{1}{n} \sum_{j=1}^n G_{(j)}(y) = G(y). \quad (1)$$

It is easy to verify that these equations hold under the model M , i.e.

$$\frac{1}{m} \sum_{i=1}^m F_{[i]}(x) = F(x), \quad \frac{1}{n} \sum_{j=1}^n G_{[j]}(y) = G(y). \tag{2}$$

All identities in (1) and (2) can be expressed in terms of density functions.

3 Main results

First we show that $\hat{R}_1^*(t)$ and $\hat{R}_2^*(t)$ are unbiased estimators. Next, it is proved that they are more precise than their SRS analogs.

Proposition 1 *The estimators $\hat{R}_1^*(t)$ and $\hat{R}_2^*(t)$ are unbiased.*

Proof Using equations in (2) we have

$$\begin{aligned} E \left\{ \sum_{i=1}^m \sum_{j=1}^n I(X_{[i]} > Y_{[j]} > t) \right\} &= \sum_{i=1}^m \sum_{j=1}^n P(X_{[i]} > Y_{[j]} > t) \\ &= \sum_{i=1}^m \sum_{j=1}^n \int P(X_{[i]} > y > t) g_{[j]}(y) dy \\ &= n \sum_{i=1}^m \int P(X_{[i]} > y > t) g(y) dy \\ &= n \sum_{i=1}^m P(X_{[i]} > Y > t) \\ &= n \sum_{i=1}^m \int P(x > Y > t) f_{[i]}(x) dx \\ &= mn \int P(x > Y > t) f(x) dx \\ &= mnP(X > Y > t) = mnR_1(t). \end{aligned}$$

Similarly,

$$\begin{aligned} E \left\{ \sum_{i=1}^m \sum_{j=1}^n I(X_{[i]} > t, Y_{[j]} > t) \right\} &= \sum_{i=1}^m \sum_{j=1}^n P(X_{[i]} > t, Y_{[j]} > t) \\ &= \sum_{i=1}^m P(X_{[i]} > t) \sum_{j=1}^n P(Y_{[j]} > t) \end{aligned}$$

$$\begin{aligned}
 &= \sum_{i=1}^m \bar{F}_{[i]}(t) \sum_{j=1}^n \bar{G}_{[j]}(t) \\
 &= mn\bar{F}(t)\bar{G}(t) = mnR_2(t).
 \end{aligned}$$

□

Proposition 2 *The variances of $\hat{R}_1(t)$ and $\hat{R}_1^*(t)$ (under model M) are given by*

$$\begin{aligned}
 m^2n^2Var\left(\hat{R}_1(t)\right) &= m(m-1)n(n-1)R_1^2(t) + nm(m-1)E\left\{\bar{F}(Y)I(Y > t)\right\}^2 \\
 &+ mn(n-1)E\left\{G(X) - G(t)\right\}^2 + mnR_1(t) - m^2n^2R_1^2(t) \quad (3)
 \end{aligned}$$

and

$$\begin{aligned}
 m^2n^2Var\left(\hat{R}_1^*(t)\right) &= E\left\{m^2\left[\sum_{j=1}^n \bar{F}(Y_{[j]}) I(Y_{[j]} > t)\right]^2\right. \\
 &\quad \left.- \sum_{i=1}^m \left[\sum_{j=1}^n \bar{F}_{[i]}(Y_{[j]}) I(Y_{[j]} > t)\right]^2\right\} \\
 &+ mE\left\{n^2[G(X) - G(t)]^2 - \sum_{j=1}^n [G_{[j]}(X) - G_{[j]}(t)]^2\right\} \\
 &+ mnR_1(t) - m^2n^2R_1^2(t). \quad (4)
 \end{aligned}$$

Proof It is easy to show that

$$m^2n^2E\left(\hat{R}_1^2(t)\right) = E(A_1 + A_2 + A_3 + A_4), \quad (5)$$

where

$$\begin{aligned}
 E(A_1) &= E\left\{\sum_{i \neq i'=1}^m \sum_{j \neq j'=1}^n I(X_i > Y_j > t) I(X_{i'} > Y_{j'} > t)\right\} \\
 &= m(m-1)n(n-1)R_1^2(t), \quad (6)
 \end{aligned}$$

$$\begin{aligned}
 E(A_2) &= E\left\{\sum_{j=1}^n \sum_{i \neq i'=1}^m I(X_i > Y_j > t) I(X_{i'} > Y_j > t)\right\} \\
 &= \sum_{j=1}^n \sum_{i \neq i'=1}^m EE\left\{I(X_i > Y_j > t) I(X_{i'} > Y_j > t) \mid Y_j\right\}
 \end{aligned}$$

$$\begin{aligned}
 &= \sum_{j=1}^n \sum_{i \neq i'=1}^m E \left\{ \bar{F}(Y) I(Y > t) \right\}^2 \\
 &= nm(m-1) E \left\{ \bar{F}(Y) I(Y > t) \right\}^2, \tag{7}
 \end{aligned}$$

$$\begin{aligned}
 E(A_3) &= E \left\{ \sum_{i=1}^m \sum_{j \neq j'=1}^n I(X_i > Y_j > t) I(X_i > Y_{j'} > t) \right\} \\
 &= \sum_{i=1}^m \sum_{j \neq j'=1}^n EE \left\{ I(X_i > Y_j > t) I(X_i > Y_{j'} > t) \mid X_i \right\} \\
 &= \sum_{i=1}^m \sum_{j \neq j'=1}^n E \left\{ G(X) - G(t) \right\}^2 = mn(n-1) E \left\{ G(X) - G(t) \right\}^2 \tag{8}
 \end{aligned}$$

and

$$E(A_4) = E \left\{ \sum_{i=1}^m \sum_{j=1}^n I(X_i > Y_j > t) \right\} = mnR_1(t). \tag{9}$$

From (5)–(9) and unbiasedness of $\hat{R}_1(t)$, the proof of the first part is complete. Similarly,

$$m^2n^2 E \left(\hat{R}_1^{*2}(t) \right) = E(B_1 + B_2 + B_3), \tag{10}$$

where

$$\begin{aligned}
 E(B_1) &= E \left\{ \sum_{i \neq i'=1}^m \sum_{j \neq j'=1}^n I(X_{[i]} > Y_{[j]} > t) I(X_{[i']} > Y_{[j']} > t) \right. \\
 &\quad \left. + \sum_{j=1}^n \sum_{i \neq i'=1}^m I(X_{[i]} > Y_{[j]} > t) I(X_{[i']} > Y_{[j]} > t) \right\} \\
 &= \sum_{i \neq i'=1}^m \sum_{j \neq j'=1}^n EE \left\{ I(X_{[i]} > Y_{[j]} > t) \mid Y_{[j]} \right\} EE \left\{ I(X_{[i']} > Y_{[j']} > t) \mid Y_{[j']} \right\} \\
 &\quad + \sum_{j=1}^n \sum_{i \neq i'=1}^m EE \left\{ I(X_{[i]} > Y_{[j]} > t) I(X_{[i']} > Y_{[j]} > t) \mid Y_{[j]} \right\} \\
 &= E \left\{ \sum_{i \neq i'=1}^m \sum_{j \neq j'=1}^n \left[\bar{F}_{[i]}(Y_{[j]}) I(Y_{[j]} > t) \right] \left[\bar{F}_{[i']}(Y_{[j']}) I(Y_{[j']} > t) \right] \right. \\
 &\quad \left. + \sum_{j=1}^n \sum_{i \neq i'=1}^m \left[\bar{F}_{[i]}(Y_{[j]}) I(Y_{[j]} > t) \right] \left[\bar{F}_{[i']}(Y_{[j]}) I(Y_{[j]} > t) \right] \right\}
 \end{aligned}$$

$$\begin{aligned}
 &= E \left\{ \left[\sum_{i=1}^m \sum_{j=1}^n \bar{F}_{[i]}(Y_{[j]}) I(Y_{[j]} > t) \right]^2 - \sum_{i=1}^m \sum_{j=1}^n \left[\bar{F}_{[i]}(Y_{[j]}) I(Y_{[j]} > t) \right]^2 \right. \\
 &\quad \left. - \sum_{i=1}^m \sum_{j \neq j'=1}^n \left[\bar{F}_{[i]}(Y_{[j]}) I(Y_{[j]} > t) \right] \left[\bar{F}_{[i]}(Y_{[j']}) I(Y_{[j']} > t) \right] \right\} \\
 &= E \left\{ m^2 \left[\sum_{j=1}^n \bar{F}(Y_{[j]}) I(Y_{[j]} > t) \right]^2 - \sum_{i=1}^m \left[\sum_{j=1}^n \bar{F}_{[i]}(Y_{[j]}) I(Y_{[j]} > t) \right]^2 \right\},
 \end{aligned}
 \tag{11}$$

$$\begin{aligned}
 E(B_2) &= E \left\{ \sum_{i=1}^m \sum_{j \neq j'=1}^n I(X_{[i]} > Y_{[j]} > t) I(X_{[i]} > Y_{[j']} > t) \right\} \\
 &= m \sum_{j \neq j'=1}^n E \left\{ I(X > Y_{[j]} > t) I(X > Y_{[j']} > t) \right\} \\
 &= m \sum_{j \neq j'=1}^n EE \left\{ I(X > Y_{[j]} > t) I(X > Y_{[j']} > t) \mid X \right\} \\
 &= m \sum_{j \neq j'=1}^n E \left\{ \left[G_{[j]}(X) - G_{[j]}(t) \right] \left[G_{[j']}(X) - G_{[j']}(t) \right] \right\} \\
 &= mE \left\{ n^2 \left[G(X) - G(t) \right]^2 - \sum_{j=1}^n \left[G_{[j]}(X) - G_{[j]}(t) \right]^2 \right\}
 \end{aligned}
 \tag{12}$$

and

$$E(B_3) = E \left\{ \sum_{i=1}^m \sum_{j=1}^n I(X_{[i]} > Y_{[j]} > t) \right\} = mnR_1(t).
 \tag{13}$$

Now the second part follows from (10)–(13) and unbiasedness of $\hat{R}_1^*(t)$. □

The variances of $\hat{R}_1(t)$ and $\hat{R}_1^*(t)$ are compared in the next proposition.

Proposition 3 *Under model M, $Var(\hat{R}_1^*(t)) \leq Var(\hat{R}_1(t))$, and the equality holds if and only if $F_{[i]} = F$ ($i = 1, \dots, m$) and $G_{[j]} = G$ ($j = 1, \dots, n$). The latter happens when either $m = n = 1$, or $p_{ir} = 1/m$ ($i, r = 1, \dots, m$) and $q_{js} = 1/n$ ($j, s = 1, \dots, n$).*

Proof Using equations (3) and (4), it can be shown

$$m^2n^2 \left[Var(\hat{R}_1(t)) - Var(\hat{R}_1^*(t)) \right] = \Gamma_1 + \Gamma_2 + \Gamma_3,
 \tag{14}$$

where

$$\begin{aligned} \Gamma_1 &= E \left\{ \sum_{i=1}^m \left[\sum_{j=1}^n \bar{F}_{[i]}(Y_{[j]}) I(Y_{[j]} > t) \right]^2 - m \left[\sum_{j=1}^n \bar{F}(Y_{[j]}) I(Y_{[j]} > t) \right]^2 \right\} \\ &= E \left\{ \sum_{i=1}^m \left(\sum_{j=1}^n \left[\bar{F}_{[i]}(Y_{[j]}) I(Y_{[j]} > t) - \bar{F}(Y_{[j]}) I(Y_{[j]} > t) \right] \right)^2 \right\}, \end{aligned} \tag{15}$$

$$\begin{aligned} \Gamma_2 &= mn(n-1)E \left\{ G(X) - G(t) \right\}^2 \\ &\quad - mE \left\{ n^2 \left[G(X) - G(t) \right]^2 - \sum_{j=1}^n \left[G_{[j]}(X) - G_{[j]}(t) \right]^2 \right\} \\ &= mE \left\{ \sum_{j=1}^n \left[G_{[j]}(X) - G_{[j]}(t) \right]^2 - n \left[G(X) - G(t) \right]^2 \right\} \\ &= mE \left\{ \sum_{j=1}^n \left[\left(G_{[j]}(X) - G_{[j]}(t) \right) - \left(G(X) - G(t) \right) \right]^2 \right\} \end{aligned} \tag{16}$$

and

$$\begin{aligned} \Gamma_3 &= m(m-1)n(n-1)R_1^2(t) + nm(m-1)E \left\{ \bar{F}(Y)I(Y > t) \right\}^2 \\ &\quad - m(m-1)E \left\{ \left[\sum_{j=1}^n \bar{F}(Y_{[j]}) I(Y_{[j]} > t) \right]^2 \right\} \\ &= m(m-1) \left[\left(1 - \frac{1}{n} \right) \left(\sum_{j=1}^n E \left\{ \bar{F}(Y_{[j]}) I(Y_{[j]} > t) \right\} \right)^2 \right. \\ &\quad \left. - \sum_{j \neq j'=1}^n E \left\{ \bar{F}(Y_{[j]}) I(Y_{[j]} > t) \right\} E \left\{ \bar{F}(Y_{[j']}) I(Y_{[j']} > t) \right\} \right] \\ &= m(m-1) \left[\sum_{j=1}^n E^2 \left\{ \bar{F}(Y_{[j]}) I(Y_{[j]} > t) \right\} \right. \\ &\quad \left. - \frac{1}{n} \left(\sum_{j=1}^n E \left\{ \bar{F}(Y_{[j]}) I(Y_{[j]} > t) \right\} \right)^2 \right] \\ &= m(m-1) \sum_{j=1}^n E^2 \left\{ \bar{F}(Y_{[j]}) I(Y_{[j]} > t) - \bar{F}(Y)I(Y > t) \right\}. \end{aligned} \tag{17}$$

Clearly, $\Gamma_i \geq 0$ ($i = 1, 2, 3$), and hence $Var(\hat{R}_1^*(t)) \leq Var(\hat{R}_1(t))$.

If $F_{[i]} = F$ ($i = 1, \dots, m$) and $G_{[j]} = G$ ($j = 1, \dots, n$), then $\Gamma_1 = \Gamma_2 = \Gamma_3 = 0$. Conversely, if $\Gamma_1 = \Gamma_2 = \Gamma_3 = 0$, we have the following conclusions. Having $\Gamma_3 = 0$ implies that for $j = 1, \dots, n$,

$$\bar{F}(Y_{[j]}) I(Y_{[j]} > t) = \bar{F}(Y) I(Y > t), \quad \text{a.s.,}$$

and thus

$$P(\bar{F}(Y_{[j]}) I(Y_{[j]} > t) = 0) = P(\bar{F}(Y) I(Y > t) = 0).$$

By continuity of F and G , the above result is equivalent to

$$P(I(Y_{[j]} > t) = 0) = P(I(Y > t) = 0),$$

or $G_{[j]} = G$. Putting this and $\Gamma_1 = 0$ together, it follows that for $i = 1, \dots, m$,

$$\bar{F}_{[i]}(Y) I(Y > t) = \bar{F}(Y) I(Y > t), \quad \text{a.s.,}$$

or $F_{[i]} = F$. □

We now present counterparts of propositions 2 and 3 in estimating $R_2(t)$.

Proposition 4 *The variances of $\hat{R}_2(t)$ and $\hat{R}_2^*(t)$ (under model M) are given by*

$$m^2 n^2 \text{Var}(\hat{R}_2(t)) = [m(m-1)\bar{F}^2(t)][n(n-1)\bar{G}^2(t)] + m\bar{F}(t)[n(n-1)\bar{G}^2(t)] + n\bar{G}(t)[m(m-1)\bar{F}^2(t)] + mnR_2(t) - m^2 n^2 R_2^2(t) \quad (18)$$

and

$$\begin{aligned} m^2 n^2 \text{Var}(\hat{R}_2^*(t)) &= \left[m^2 \bar{F}^2(t) - \sum_{i=1}^m \bar{F}_{[i]}^2(t) \right] \left[n^2 \bar{G}^2(t) - \sum_{j=1}^n \bar{G}_{[j]}^2(t) \right] \\ &\quad + m\bar{F}(t) \left[n^2 \bar{G}^2(t) - \sum_{j=1}^n \bar{G}_{[j]}^2(t) \right] \\ &\quad + n\bar{G}(t) \left[m^2 \bar{F}^2(t) - \sum_{i=1}^m \bar{F}_{[i]}^2(t) \right] + mnR_2(t) - m^2 n^2 R_2^2(t). \end{aligned} \quad (19)$$

Proof It is readily seen that

$$m^2 n^2 E(\hat{R}_2^2(t)) = E(C_1 + C_2 + C_3 + C_4), \quad (20)$$

where

$$\begin{aligned}
 E(C_1) &= E \left\{ \sum_{i \neq i'=1}^m \sum_{j \neq j'=1}^n I(X_i > t, Y_j > t) I(X_{i'} > t, Y_{j'} > t) \right\} \\
 &= [m(m-1)\bar{F}^2(t)] [n(n-1)\bar{G}^2(t)], \tag{21}
 \end{aligned}$$

$$\begin{aligned}
 E(C_2) &= E \left\{ \sum_{i=1}^m \sum_{j \neq j'=1}^n I(X_i > t, Y_j > t) I(X_i > t, Y_{j'} > t) \right\} \\
 &= m\bar{F}(t) [n(n-1)\bar{G}^2(t)], \tag{22}
 \end{aligned}$$

$$\begin{aligned}
 E(C_3) &= E \left\{ \sum_{j=1}^n \sum_{i \neq i'=1}^m I(X_i > t, Y_j > t) I(X_{i'} > t, Y_j > t) \right\} \\
 &= n\bar{G}(t) [m(m-1)\bar{F}^2(t)] \tag{23}
 \end{aligned}$$

and

$$E(C_4) = E \left\{ \sum_{i=1}^m \sum_{j=1}^n I(X_i > t, Y_j > t) \right\} = mnR_2(t). \tag{24}$$

The first part follows from (20)–(24) and unbiasedness of $\hat{R}_2(t)$. By a similar argument,

$$m^2n^2 E \left(\hat{R}_2^{*2}(t) \right) = E(D_1 + D_2 + D_3 + D_4), \tag{25}$$

where

$$\begin{aligned}
 E(D_1) &= E \left\{ \sum_{i \neq i'=1}^m \sum_{j \neq j'=1}^n I(X_{[i]} > t, Y_{[j]} > t) I(X_{[i']} > t, Y_{[j']} > t) \right\} \\
 &= \left[\sum_{i \neq i'=1}^m \bar{F}_{[i]}(t)\bar{F}_{[i']}(t) \right] \left[\sum_{j \neq j'=1}^n \bar{G}_{[j]}(t)\bar{G}_{[j']}(t) \right] \\
 &= \left[\left(\sum_{i=1}^m \bar{F}_{[i]}(t) \right)^2 - \sum_{i=1}^m \bar{F}_{[i]}^2(t) \right] \left[\left(\sum_{j=1}^n \bar{G}_{[j]}(t) \right)^2 - \sum_{j=1}^n \bar{G}_{[j]}^2(t) \right] \\
 &= \left[m^2\bar{F}^2(t) - \sum_{i=1}^m \bar{F}_{[i]}^2(t) \right] \left[n^2\bar{G}^2(t) - \sum_{j=1}^n \bar{G}_{[j]}^2(t) \right], \tag{26}
 \end{aligned}$$

$$\begin{aligned}
 E(D_2) &= E \left\{ \sum_{i=1}^m \sum_{j \neq j'=1}^n I(X_{[i]} > t, Y_{[j]} > t) I(X_{[i]} > t, Y_{[j']} > t) \right\} \\
 &= \sum_{i=1}^m \bar{F}_{[i]}(t) \sum_{j \neq j'=1}^n \bar{G}_{[j]}(t) \bar{G}_{[j']}(t) \\
 &= m \bar{F}(t) \left[\left(\sum_{j=1}^n \bar{G}_{[j]}(t) \right)^2 - \sum_{j=1}^n \bar{G}_{[j]}^2(t) \right] \\
 &= m \bar{F}(t) \left[n^2 \bar{G}^2(t) - \sum_{j=1}^n \bar{G}_{[j]}^2(t) \right], \tag{27}
 \end{aligned}$$

$$\begin{aligned}
 E(D_3) &= E \left\{ \sum_{j=1}^n \sum_{i \neq i'=1}^m I(X_{[i]} > t, Y_{[j]} > t) I(X_{[i']} > t, Y_{[j]} > t) \right\} \\
 &= \sum_{j=1}^n \bar{G}_{[j]}(t) \sum_{i \neq i'=1}^m \bar{F}_{[i]}(t) \bar{F}_{[i']}(t) \\
 &= n \bar{G}(t) \left[\left(\sum_{i=1}^m \bar{F}_{[i]}(t) \right)^2 - \sum_{i=1}^m \bar{F}_{[i]}^2(t) \right] \\
 &= n \bar{G}(t) \left[m^2 \bar{F}^2(t) - \sum_{i=1}^m \bar{F}_{[i]}^2(t) \right] \tag{28}
 \end{aligned}$$

and

$$\begin{aligned}
 E(D_4) &= E \left\{ \sum_{i=1}^m \sum_{j=1}^n I(X_{[i]} > t, Y_{[j]} > t) \right\} = \sum_{i=1}^m \sum_{j=1}^n P(X_{[i]} > t, Y_{[j]} > t) \\
 &= mn R_2(t). \tag{29}
 \end{aligned}$$

From (25)–(29) and unbiasedness of $\hat{R}_2^*(t)$, the second part is concluded. □

The variances of $\hat{R}_2(t)$ and $\hat{R}_2^*(t)$ are compared in the next proposition.

Proposition 5 *Under model M, $Var(\hat{R}_2^*(t)) \leq Var(\hat{R}_2(t))$, and the equality holds if and only if $F_{[i]} = F$ ($i = 1, \dots, m$) and $G_{[j]} = G$ ($j = 1, \dots, n$). The latter happens when either $m = n = 1$, or $p_{ir} = 1/m$ ($i, r = 1, \dots, m$) and $q_{js} = 1/n$ ($j, s = 1, \dots, n$).*

Proof Using equations (18) and (19), it is straightforward to see that

$$m^2 n^2 \left[Var(\hat{R}_2(t)) - Var(\hat{R}_2^*(t)) \right] = \Delta_1 + \Delta_2 + \Delta_3,$$

where

$$\Delta_1 = m\bar{F}(t) \left[\sum_{j=1}^n \bar{G}_{[j]}^2(t) - n\bar{G}^2(t) \right] + n\bar{G}(t) \left[\sum_{i=1}^m \bar{F}_{[i]}^2(t) - m\bar{F}^2(t) \right],$$

$$\Delta_2 = m^2\bar{F}^2(t) \left[\sum_{j=1}^n \bar{G}_{[j]}^2(t) - n\bar{G}^2(t) \right] + n^2\bar{G}^2(t) \left[\sum_{i=1}^m \bar{F}_{[i]}^2(t) - m\bar{F}^2(t) \right]$$

and

$$\Delta_3 = mn\bar{F}^2(t)\bar{G}^2(t) - \left[\sum_{i=1}^m \bar{F}_{[i]}^2(t) \right] \left[\sum_{j=1}^n \bar{G}_{[j]}^2(t) \right].$$

Let $u_i = \bar{F}_{[i]} (i = 1, \dots, m)$ and $v_j = \bar{G}_{[j]} (j = 1, \dots, n)$ with the corresponding means $\bar{u} = \sum_{i=1}^m u_i/m$ and $\bar{v} = \sum_{j=1}^n v_j/n$, and variances $S_u^2 = \sum_{i=1}^m (u_i - \bar{u})^2/m$ and $S_v^2 = \sum_{j=1}^n (v_j - \bar{v})^2/n$. Also, assume $\bar{u}^2 = \sum_{i=1}^m u_i^2/m$ and $\bar{v}^2 = \sum_{j=1}^n v_j^2/n$. Then, we have

$$\Delta_1 = mn \left[\bar{u}S_v^2 + \bar{v}S_u^2 \right],$$

$$\Delta_2 = mn \left[m(\bar{u})^2S_v^2 + n(\bar{v})^2S_u^2 \right]$$

and

$$\Delta_3 = -mn \left[(\bar{u})^2S_v^2 + (\bar{v})^2S_u^2 + S_u^2S_v^2 \right].$$

The first part of the proposition holds owing to the following inequality

$$\begin{aligned} \frac{\Delta_1 + \Delta_2 + \Delta_3}{mn} &= \bar{u}S_v^2 + \bar{v}S_u^2 + m(\bar{u})^2S_v^2 + n(\bar{v})^2S_u^2 - (\bar{u})^2S_v^2 - (\bar{v})^2S_u^2 - S_u^2S_v^2 \\ &\geq \bar{u}^2S_v^2 + \bar{v}^2S_u^2 + m(\bar{u})^2S_v^2 + n(\bar{v})^2S_u^2 - (\bar{u})^2S_v^2 - (\bar{v})^2S_u^2 - S_u^2S_v^2 \\ &= (\bar{u}^2 - \bar{u}^2)S_v^2 + (\bar{v}^2 - \bar{v}^2)S_u^2 + m(\bar{u})^2S_v^2 + n(\bar{v})^2S_u^2 - S_u^2S_v^2 \\ &= m(\bar{u})^2S_v^2 + n(\bar{v})^2S_u^2 + S_u^2S_v^2 \geq 0, \end{aligned} \tag{30}$$

where the first inequality follows from the fact that $\bar{u} \geq \bar{u}^2$ and $\bar{v} \geq \bar{v}^2$.

If $F_{[i]} = F (i = 1, \dots, m)$ and $G_{[j]} = G (j = 1, \dots, n)$, then $\Delta_1 = \Delta_2 = \Delta_3 = 0$. Conversely, if $\Delta_1 = \Delta_2 = \Delta_3 = 0$, then from (30) we get $S_u^2 = S_v^2 = 0$. That is to say $F_{[i]} = F (i = 1, \dots, m)$ and $G_{[j]} = G (j = 1, \dots, n)$. \square

The following proposition provides components needed to approximate $Var(\hat{R}(t))$ and $Var(\hat{R}^*(t))$.

Proposition 6 The covariances between $\hat{R}_1(t)$ and $\hat{R}_2(t)$, and $\hat{R}_1^*(t)$ and $\hat{R}_2^*(t)$ (under model M) are given by

$$m^2 n^2 \text{Cov} \left(\hat{R}_1(t), \hat{R}_2(t) \right) = m(m-1)n(n-1)R_1(t)R_2(t) + nm(m-1)R_1(t)\bar{F}(t) \\ + mn(n-1)R_1(t)\bar{G}(t) + mnR_1(t) - m^2 n^2 R_1(t)R_2(t) \quad (31)$$

and

$$m^2 n^2 \text{Cov} \left(\hat{R}_1^*(t), \hat{R}_2^*(t) \right) = \bar{F}_{[i']}(t)\bar{G}_{[j']}(t)P(X_{[i]} > Y_{[j]} > t) \\ + n \left[m^2 R_1(t)\bar{F}(t) - \sum_{i=1}^m P(X_{[i]} > Y > t)\bar{F}_{[i]}(t) \right] \\ + m \left[n^2 R_1(t)\bar{G}(t) - \sum_{j=1}^n P(X > Y_{[j]} > t)\bar{G}_{[j]}(t) \right] \\ + mnR_1(t) - m^2 n^2 R_1(t)R_2(t). \quad (32)$$

Proof It can be shown that

$$m^2 n^2 E \left(\hat{R}_1(t)\hat{R}_2(t) \right) = E(I_1 + I_2 + I_3 + I_4), \quad (33)$$

where

$$E(I_1) = E \left\{ \sum_{i \neq i'=1}^m \sum_{j \neq j'=1}^n I(X_i > Y_j > t) I(X_{i'} > t, Y_{j'} > t) \right\} \\ = m(m-1)n(n-1)R_1(t)R_2(t), \quad (34)$$

$$E(I_2) = E \left\{ \sum_{j=1}^n \sum_{i \neq i'=1}^m I(X_i > Y_j > t) I(X_{i'} > t, Y_j > t) \right\} \\ = nm(m-1)R_1(t)\bar{F}(t), \quad (35)$$

$$E(I_3) = E \left\{ \sum_{i=1}^m \sum_{j \neq j'=1}^n I(X_i > Y_j > t) I(X_i > t, Y_{j'} > t) \right\} \\ = mn(n-1)R_1(t)\bar{G}(t) \quad (36)$$

and

$$E(I_4) = E \left\{ \sum_{i=1}^m \sum_{j=1}^n I(X_i > Y_j > t) I(X_i > t, Y_j > t) \right\} = mnR_1(t). \quad (37)$$

From (33)–(37) and unbiasedness of $\hat{R}_1(t)$ and $\hat{R}_2(t)$, we get (31). Similarly,

$$m^2n^2E\left(\hat{R}_1^*(t)\hat{R}_2^*(t)\right) = E\left(J_1 + J_2 + J_3 + J_4\right), \tag{38}$$

where

$$\begin{aligned} E(J_1) &= E\left\{\sum_{i \neq i'=1}^m \sum_{j \neq j'=1}^n I\left(X_{[i]} > Y_{[j]} > t\right) I\left(X_{[i']} > t, Y_{[j']} > t\right)\right\} \\ &= \bar{F}_{[i']}(t)\bar{G}_{[j']}(t)P\left(X_{[i]} > Y_{[j]} > t\right), \end{aligned} \tag{39}$$

$$\begin{aligned} E(J_2) &= E\left\{\sum_{j=1}^n \sum_{i \neq i'=1}^m I\left(X_{[i]} > Y_{[j]} > t\right) I\left(X_{[i']} > t, Y_{[j]} > t\right)\right\} \\ &= \sum_{j=1}^n \sum_{i \neq i'=1}^m P\left(X_{[i]} > Y_{[j]} > t\right) \bar{F}_{[i']}(t) = n \sum_{i \neq i'=1}^m P\left(X_{[i]} > Y > t\right) \bar{F}_{[i']}(t) \\ &= n \left[\sum_{i=1}^m \sum_{i'=1}^m P\left(X_{[i]} > Y > t\right) \bar{F}_{[i']}(t) - \sum_{i=1}^m P\left(X_{[i]} > Y > t\right) \bar{F}_{[i]}(t) \right] \\ &= n \left[m^2 R_1(t) \bar{F}(t) - \sum_{i=1}^m P\left(X_{[i]} > Y > t\right) \bar{F}_{[i]}(t) \right], \end{aligned} \tag{40}$$

$$\begin{aligned} E(J_3) &= E\left\{\sum_{i=1}^m \sum_{j \neq j'=1}^n I\left(X_{[i]} > Y_{[j]} > t\right) I\left(X_{[i]} > t, Y_{[j']} > t\right)\right\} \\ &= \sum_{i=1}^m \sum_{j \neq j'=1}^n P\left(X_{[i]} > Y_{[j]} > t\right) \bar{G}_{[j']}(t) = m \sum_{j \neq j'=1}^n P\left(X > Y_{[j]} > t\right) \bar{G}_{[j']}(t) \\ &= m \left[\sum_{j=1}^n \sum_{j'=1}^n P\left(X > Y_{[j]} > t\right) \bar{G}_{[j']}(t) - \sum_{j=1}^n P\left(X > Y_{[j]} > t\right) \bar{G}_{[j]}(t) \right] \\ &= m \left[n^2 R_1(t) \bar{G}(t) - \sum_{j=1}^n P\left(X > Y_{[j]} > t\right) \bar{G}_{[j]}(t) \right] \end{aligned} \tag{41}$$

and

$$\begin{aligned} E(J_4) &= E\left\{\sum_{i=1}^m \sum_{j=1}^n I\left(X_{[i]} > Y_{[j]} > t\right) I\left(X_{[i]} > t, Y_{[j]} > t\right)\right\} \\ &= mnR_1(t). \end{aligned} \tag{42}$$

Now, (32) is immediate from (38)–(42) and unbiasedness of $\hat{R}_1^*(t)$ and $\hat{R}_2^*(t)$. \square

The next result shows asymptotic unbiasedness of $\hat{R}(t)$ and $\hat{R}^*(t)$, and provides approximations of their variances.

Proposition 7 *The estimators $\hat{R}(t)$ and $\hat{R}^*(t)$ are asymptotically unbiased with approximate variances*

$$\begin{aligned} Var(\hat{R}(t)) \approx & \frac{1}{R_2^2(t)} Var(\hat{R}_1(t)) + \frac{R_1^2(t)}{R_2^4(t)} Var(\hat{R}_2(t)) \\ & - 2 \frac{R_1(t)}{R_2^3(t)} Cov(\hat{R}_1(t), \hat{R}_2(t)) \end{aligned} \tag{43}$$

and

$$\begin{aligned} Var(\hat{R}^*(t)) \approx & \frac{1}{R_2^2(t)} Var(\hat{R}_1^*(t)) + \frac{R_1^2(t)}{R_2^4(t)} Var(\hat{R}_2^*(t)) \\ & - 2 \frac{R_1(t)}{R_2^3(t)} Cov(\hat{R}_1^*(t), \hat{R}_2^*(t)). \end{aligned} \tag{44}$$

Proof Using continuous map proposition (Shao 2003, p. 59, theorem 1.10(i)) it can be concluded that $\hat{R}(t)$ and $\hat{R}^*(t)$ are strongly consistent. This and the dominated convergence theorem imply asymptotic unbiasedness of the two estimators.

Let T_1, \dots, T_k be random variables with means $\theta_1, \dots, \theta_k$, and define $\mathbf{T} = (T_1, \dots, T_k)$ and $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)$. Suppose there is a differentiable function $g(\mathbf{T})$ (an estimator of some parameter) for which we want an approximate estimate of variance. Using the first-order Taylor series expansion of g about $\boldsymbol{\theta}$, we get

$$Var(g(\mathbf{T})) \approx \sum_{i=1}^k [g'_i(\boldsymbol{\theta})]^2 Var(T_i) + 2 \sum_{i < j} g'_i(\boldsymbol{\theta})g'_j(\boldsymbol{\theta})Cov(T_i, T_j),$$

where

$$g'_i(\boldsymbol{\theta}) = \frac{\partial}{\partial t_i} g(\mathbf{t})|_{t_1=\theta_1, \dots, t_k=\theta_k}.$$

Now, (43) and (44) are concluded by taking $g(\mathbf{T}) = T_1/T_2$, where T_1 and T_2 are the numerator and the denominator of each estimator, respectively. The components needed for the two approximations are given in propositions 2, 4 and 6. □

In view of propositions 1, 2 and 4, it is expected that $\hat{R}^*(t)$ will be more efficient than $\hat{R}(t)$. It may be difficult to show this using the above proposition as we only have approximations for $Var(\hat{R}(t))$ and $Var(\hat{R}^*(t))$.

4 Perfect ranking setup

The judgment ranking process is of high importance in drawing a representative sample by RSS. In extreme cases, ranking may be so poor to yield a simple random sample. Hence, RSS is at least as efficient as SRS, intuitively. In the case of estimating $R_1(t)$

and $R_2(t)$, this was formally shown in propositions 3 and 5. The maximum efficiency is expected to arise in the absence of ranking errors, i.e. the perfect ranking scenario. This section aims to establish some properties in this respect. To this end, we need a few notions and results from matrix algebra, and a lemma which are set out here.

The L_1, L_∞ and L_2 norms for an $r \times c$ matrix $\mathbf{A} = [a_{ij}]$ are defined as

$$\|\mathbf{A}\|_1 = \max_{j=1, \dots, c} \sum_{i=1}^r a_{ij},$$

$$\|\mathbf{A}\|_\infty = \max_{i=1, \dots, r} \sum_{j=1}^c a_{ij}$$

and

$$\|\mathbf{A}\|_2 = \sqrt{\lambda_{\max}(\mathbf{A}'\mathbf{A})},$$

where $\lambda_{\max}(\mathbf{A}'\mathbf{A})$ is the largest eigenvalue of $\mathbf{A}'\mathbf{A}$ matrix. If the product of matrices \mathbf{A} and \mathbf{B} is defined, then

$$\|\mathbf{AB}\|_2 \leq \|\mathbf{A}\|_2 \|\mathbf{B}\|_2 \tag{45}$$

and

$$\|\mathbf{A}\|_2 \leq \|\mathbf{A}\|_1 \|\mathbf{A}\|_\infty. \tag{46}$$

See [Datta \(2010\)](#) for more details.

Lemma 1 *If $h = \sum_{j=1}^n \bar{F}(Y_{(j)}) I(Y_{(j)} > t)$ and $H = \sum_{j=1}^n \bar{F}(Y_{[j]}) I(Y_{[j]} > t)$, then $Var(h) \leq Var(H)$.*

Proof Using conditional variance formula, we have

$$\begin{aligned} Var(H) &= \sum_{j=1}^n Var(\bar{F}(Y_{[j]}) I(Y_{[j]} > t)) \\ &\geq \sum_{j=1}^n \left[\sum_{k=1}^n Var(\bar{F}(Y_{(k)}) I(Y_{(k)} > t)) q_{jk} \right] \\ &= \sum_{k=1}^n Var(\bar{F}(Y_{(k)}) I(Y_{(k)} > t)) = Var(h), \end{aligned}$$

as was asserted. □

Let $\tilde{R}_1^*(t)$ and $\tilde{R}_2^*(t)$ denote the estimators $\hat{R}_1^*(t)$ and $\hat{R}_2^*(t)$ under perfect ranking assumption, respectively. Then we have the following two propositions. It should be mentioned that the approach adopted in proofs is distinctly different from that of similar result in [Sengupta and Mukhuti \(2008\)](#).

Proposition 8 Under model M , $Var(\tilde{R}_1^*(t)) \leq Var(\hat{R}_1^*(t))$

Proof In view of (4), it is sufficient to show that

$$\begin{aligned}
 & E \left\{ m^2 \left[\sum_{j=1}^n \bar{F}(Y_{(j)}) I(Y_{(j)} > t) \right]^2 - \sum_{i=1}^m \left[\sum_{j=1}^n \bar{F}_{(i)}(Y_{(j)}) I(Y_{(j)} > t) \right]^2 \right\} \\
 & \leq E \left\{ m^2 \left[\sum_{j=1}^n \bar{F}(Y_{[j]}) I(Y_{[j]} > t) \right]^2 - \sum_{i=1}^m \left[\sum_{j=1}^n \bar{F}_{[i]}(Y_{[j]}) I(Y_{[j]} > t) \right]^2 \right\} \quad (47)
 \end{aligned}$$

and

$$E \left\{ \sum_{j=1}^n [G_{[j]}(X) - G_{[j]}(t)]^2 \right\} \leq E \left\{ \sum_{j=1}^n [G_{(j)}(X) - G_{(j)}(t)]^2 \right\}. \quad (48)$$

We proceed with proving the first inequality. Assume that $Z_{(i)} = \sum_{j=1}^n \bar{F}_{(i)}(Y_{[j]}) I(Y_{[j]} > t)$ and $Z_{[i]} = \sum_{j=1}^n \bar{F}_{[i]}(Y_{[j]}) I(Y_{[j]} > t)$. Then

$$\begin{aligned}
 Z_{[i]} = \sum_{j=1}^n \bar{F}_{[i]}(Y_{[j]}) I(Y_{[j]} > t) &= \sum_{j=1}^n \sum_{k=1}^m p_{ik} \bar{F}_{(k)}(Y_{[j]}) I(Y_{[j]} > t) = \sum_{k=1}^m p_{ik} Z_{(k)}.
 \end{aligned} \quad (49)$$

Let Ω_Y be the sample space on which Y is defined. If $\mathbf{P} = [p_{ir}]$ and $\mathbf{Z}' = (Z_{(1)}(\vartheta), \dots, Z_{(m)}(\vartheta))$ given a fixed $\vartheta \in \Omega_Y$, then using (45), (46) and (49) it follows that

$$\begin{aligned}
 \sum_{i=1}^m Z_{[i]}^2(\vartheta) &= \sum_{i=1}^m \left(\sum_{k=1}^m p_{ik} Z_{(k)}(\vartheta) \right)^2 = \|\mathbf{PZ}\|_2^2 \leq \|\mathbf{P}\|_2^2 \|\mathbf{Z}\|_2^2 \\
 &\leq \|\mathbf{P}\|_1 \|\mathbf{P}\|_\infty \sum_{i=1}^m Z_{(i)}^2(\vartheta) \\
 &= \sum_{i=1}^m Z_{(i)}^2(\vartheta).
 \end{aligned}$$

The last equality holds because $\sum_{i=1}^m p_{ik} = \sum_{k=1}^m p_{ik} = 1$. Accordingly,

$$\begin{aligned}
 & E \left\{ m^2 \left[\sum_{j=1}^n \bar{F}(Y_{[j]}) I(Y_{[j]} > t) \right]^2 - \sum_{i=1}^m \left[\sum_{j=1}^n \bar{F}_{(i)}(Y_{[j]}) I(Y_{[j]} > t) \right]^2 \right\} \\
 & \leq E \left\{ m^2 \left[\sum_{j=1}^n \bar{F}(Y_{[j]}) I(Y_{[j]} > t) \right]^2 - \sum_{i=1}^m \left[\sum_{j=1}^n \bar{F}_{[i]}(Y_{[j]}) I(Y_{[j]} > t) \right]^2 \right\}.
 \end{aligned}$$

Now, (47) is concluded if

$$\begin{aligned}
 & E \left\{ m^2 \left[\sum_{j=1}^n \bar{F}(Y_{(j)}) I(Y_{(j)} > t) \right]^2 - \sum_{i=1}^m \left[\sum_{j=1}^n \bar{F}_{(i)}(Y_{(j)}) I(Y_{(j)} > t) \right]^2 \right\} \\
 & \leq E \left\{ m^2 \left[\sum_{j=1}^n \bar{F}(Y_{[j]}) I(Y_{[j]} > t) \right]^2 - \sum_{i=1}^m \left[\sum_{j=1}^n \bar{F}_{(i)}(Y_{[j]}) I(Y_{[j]} > t) \right]^2 \right\}.
 \end{aligned}
 \tag{50}$$

For $i = 1, \dots, m$, suppose $h_{(i)} = \sum_{j=1}^n \bar{F}_{(i)}(Y_{(j)}) I(Y_{(j)} > t)$ and h be as in Lemma 1. We note that

$$\bar{F}_{(i)}(t) = \sum_{k=0}^{i-1} \binom{m}{k} [F(t)]^k [\bar{F}(t)]^{m-k}.$$

So, $\bar{F}_{(i)}(t)$ is an increasing function in i , and thereby $h_{(1)} < \dots < h_{(m)}$ are order statistics. Therefore,

$$\sum_{i=1}^m E(h_{(i)}^2) = \sum_{i=1}^m \int t^2 f_{h_{(i)}}(t) dt = m \int t^2 f_h(t) dt = mE(h^2),
 \tag{51}$$

where $f_{h_{(i)}}$ and f_h denote the density function of $h_{(i)}$ and h , respectively. Similarly, one can define $H_{(i)} = \sum_{j=1}^n \bar{F}_{(i)}(Y_{[j]}) I(Y_{[j]} > t)$ and H as in Lemma 1, and conclude that

$$\sum_{i=1}^m E(H_{(i)}^2) = mE(H^2).
 \tag{52}$$

From (51) and (52), (50) reduces to $E(h^2) \leq E(H^2)$. It is equivalent to $Var(h) \leq Var(H)$ as $E(h) = E(H)$. This holds by virtue of Lemma 1.

Let Ω_X be the sample space on which X is defined. If $\mathbf{Q} = [q_{js}]$ and $\mathbf{G}' = (G_{(1)}(X(\eta)) - G_{(1)}(t), \dots, G_{(n)}(X(\eta)) - G_{(n)}(t))$ for each fixed $\eta \in \Omega_X$, then (48) follows by applying (45) and (46) using \mathbf{Q} and \mathbf{G} . □

Proposition 9 Under model M , $Var(\tilde{R}_2^*(t)) \leq Var(\hat{R}_2^*(t))$.

Proof According to (19) it is sufficient to show that

$$\sum_{i=1}^m \bar{F}_{[i]}^2(t) \leq \sum_{i=1}^m \bar{F}_{(i)}^2(t)$$

and

$$\sum_{j=1}^n \bar{G}_{[j]}^2(t) \leq \sum_{j=1}^n \bar{G}_{(j)}^2(t).$$

Let \mathbf{P} and \mathbf{Q} be as in the the previous proposition, $\bar{\mathbf{F}}' = (\bar{F}_{(1)}(t), \dots, \bar{F}_{(m)}(t))$ and $\bar{\mathbf{G}}' = (\bar{G}_{(1)}(t), \dots, \bar{G}_{(n)}(t))$. Then using (45) and (46)

$$\begin{aligned} \sum_{i=1}^m \bar{F}_{[i]}^2(t) &= \sum_{i=1}^m \left(\sum_{k=1}^m p_{ik} \bar{F}_{(k)}(t) \right)^2 = \|\mathbf{P}\bar{\mathbf{F}}\|_2^2 \leq \|\mathbf{P}\|_2^2 \|\bar{\mathbf{F}}\|_2^2 \\ &\leq \|\mathbf{P}\|_1 \|\mathbf{P}\|_\infty \sum_{i=1}^m \bar{F}_{(i)}^2(t) \\ &= \sum_{i=1}^m \bar{F}_{(i)}^2(t). \end{aligned}$$

Proceeding as in the above with \mathbf{Q} and $\bar{\mathbf{G}}$, it is proved that $\sum_{j=1}^n \bar{G}_{[j]}^2(t) \leq \sum_{j=1}^n \bar{G}_{(j)}^2(t)$ □

5 Simulation study

We used Monte Carlo simulation to assess the performance of the proposed estimator. In doing so, it was assumed that both populations follow Weibull, normal or uniform distribution. Let X be distributed as Weibull with distribution function

$$F(x) = 1 - \exp \left\{ - \left(\frac{x}{\beta} \right)^\alpha \right\}, \quad x > 0; \alpha > 0, \beta > 0,$$

which is denoted by $X \sim W(\alpha, \beta)$. If $X \sim W(\alpha_1, \beta_1)$ and $Y \sim W(\alpha_2, \beta_2)$ are independent random variables, then it can be shown that

$$\begin{aligned} R(t) &= \frac{\alpha_2}{\beta_2^2} \exp \left\{ \left(\frac{t}{\beta_1} \right)^{\alpha_1} + \left(\frac{t}{\beta_2} \right)^{\alpha_2} \right\} \int_t^\infty x^{\alpha_2-1} \\ &\quad \times \exp \left\{ - \left(\frac{x}{\beta_1} \right)^{\alpha_1} - \left(\frac{x}{\beta_2} \right)^{\alpha_2} \right\} dx, \quad t > 0. \end{aligned}$$

Similarly, if X has a normal distribution with mean zero and variance σ^2 , $X \sim N(0, \sigma^2)$, and Y is a standard normal random variable, $Y \sim N(0, 1)$, then

$$\begin{aligned} R(t) &= \left[\int_t^\infty \Phi(x) \frac{1}{\sqrt{2\pi}\sigma} \exp \left\{ - \frac{x^2}{2\sigma^2} \right\} dx - \Phi(t) \left(1 - \Phi \left(\frac{t}{\sigma} \right) \right) \right] \\ &\quad \times \left[\left(1 - \Phi \left(\frac{t}{\sigma} \right) \right) (1 - \Phi(t)) \right]^{-1}, \quad t \in \mathbb{R}, \end{aligned}$$

where $\Phi(\cdot)$ is the distribution function of Y . Finally, suppose X and Y are uniformly distributed on intervals $(0, a)$ and $(0, b)$, respectively, where $a < b$. Then

$$R(t) = \frac{a-t}{2(b-t)}, \quad 0 < t < a.$$

We considered the following pairs of distributions:

- $X \sim W(2, 1)$ versus $Y \sim W(2, 2)$,
- $X \sim W(2, 1)$ versus $Y \sim W(1, 1)$,
- $X \sim W(1, 1)$ versus $Y \sim W(0.5, 3)$,
- $X \sim N(0, (1.25)^2)$ versus $Y \sim N(0, 1)$,
- $X \sim U(0, 1.5)$ versus $Y \sim U(0, 2)$.

Also, the selected two groups of sample sizes are $m, n = 2, 5, 10$ and $m, n = 20, 50, 100$.

Given a fixed t , the efficiency of $\tilde{R}^*(t) = \tilde{R}_1^*(t)/\tilde{R}_2^*(t)$ relative to $\hat{R}(t)$ is estimated as follows. For each combination of distributions and sample sizes, 10,000 pairs of samples were generated in SRS and RSS schemes. The two estimators were computed from each pair of samples in the corresponding designs, and their mean squared errors (MSEs) were determined. The relative efficiency (RE) is defined as the ratio of $\widehat{MSE}(\hat{R}(t))$ to $\widehat{MSE}(\tilde{R}^*(t))$. The RE values larger than one indicate that $\tilde{R}^*(t)$ is more efficient than $\hat{R}(t)$. Figures 1, 2, 3, 4, 5, 6, 7, 8, 9 and 10 in the appendix display the results.

It is observed that that RSS based estimator is more efficient than its SRS competitor as was shown in theory. Also for each pair of distributions and fixed m , the larger n , the higher RE, generally. This trend is sometimes violated, e.g. see Fig. 4, when $m = 50$ and t is small. The interesting thing to note is that how RE is affected by the time. Although RE (as a function of t) may show fluctuations at the beginning, it becomes monotone decreasing as time goes on. The point that RE is maximized depends on the choice of parent distributions. For example, when $X \sim W(2, 1)$ and $Y \sim W(1, 1)$, the maximum RE is nearly obtained at $t = 0$, while this is not the case with $X \sim W(2, 1)$ and $Y \sim W(2, 2)$.

To have better understanding of the estimators' behavior, we provide graphs for MSEs of the estimators as a function of t . To save space, three of the above mentioned pairs of distributions were only considered. Moreover, the sample sizes were limited to $(m, n) = (5, 5), (10, 10), (20, 20)$. The results are given in Figs. 11 and 12. It is seen that under both SRS and RSS designs, MSEs are decreasing in sample sizes, given a fixed t . Finally, Table 1 contains estimated biases of the estimators at some points, where the entries given in parenthesis are based on RSS. Analogs of Figs. 11 and 12 for bias (not provided here) suggest that biases in the two schemes are bounded above as a function of t .

6 Application

Although the use of stress-strength models was originally motivated by problems in physics and engineering, it is not limited to these contexts. It is worth mentioning

Table 1 Estimated biases under SRS and RSS when $m = n = 5, 10, 20$ for (a) $X \sim W(2, 1)$ and $Y \sim W(2, 2)$, (b) $X \sim N(0, (1.25)^2)$ and $Y \sim N(0, 1)$ and (c) $X \sim U(0, 1.5)$ and $Y \sim U(0, 2)$

(a)		t				
(m, n)		0.1	0.5	1	1.5	2
(5,5)		-0.0009 (0.0011)	0.0832 (0.0232)	0.4684 (0.4219)	0.7389 (0.7335)	0.7688 (0.7657)
(10,10)		-0.0001 (-0.0016)	0.0078 (0.0014)	0.2664 (0.1684)	0.6700 (0.6560)	0.7129 (0.7023)
(20,20)		-0.0015 (0.0001)	-0.0003 (0.0006)	0.0880 (0.0230)	0.5533 (0.5232)	0.5931 (0.5726)
(b)		t				
(m, n)		-2	-1	0	1	1.5
(5,5)		0.0147 (0.0169)	0.0435 (0.0190)	0.2924 (0.2347)	0.4095 (0.4051)	0.4348 (0.4357)
(10,10)		0.0139 (0.0142)	0.0118 (0.0126)	0.1315 (0.0485)	0.3160 (0.2705)	0.3663 (0.3532)
(20,20)		0.0141 (0.0138)	0.0146 (0.0145)	0.0308 (0.0107)	0.1751 (0.0762)	0.2160 (0.1270)
(c)		t				
(m, n)		0.1	0.3	0.5	0.7	1
(5,5)		-0.0162 (-0.0214)	-0.0186 (-0.0254)	-0.0065 (-0.0324)	0.0599 (-0.0311)	0.1625 (0.0513)
(10,10)		-0.0166 (-0.0171)	-0.0189 (-0.0222)	-0.0266 (-0.0286)	-0.0467 (-0.0599)	0.0120 (-0.0131)
(20,20)		-0.0173 (-0.0168)	-0.0219 (-0.0204)	-0.0273 (-0.0262)	-0.0584 (-0.0595)	-0.0055 (-0.0065)

that R provides a general measure of the difference between two populations, and has found applications in different fields such as economics, quality control, psychology, medicine and clinical trials. For instance, if Y is the response of a control group, and X is that of a treatment group, then R is a measure of the treatment effect.

We now illustrate the proposed procedure using a data set collected by Murray et al. (2000). They conducted an experiment in which apple trees are sprayed with chemical containing fluorescent tracer, Tinopal CBS-X, at 2 % concentration level in water. Two nine-tree plots were chosen for spraying. One plot was sprayed at high volume, using coarse nozzles on the sprayer to give a large average droplet size. The other plot was sprayed at low volume, using fine nozzles to give a small average droplet size. Fifty sets of five leaves were identified from the central five trees of each plot, and used to draw 10 copies a ranked set sample of size five, from each plot. The variable of interest is the percentage of area covered by the spray on the surface of the leaves. The formal measurement entails chemical analysis of the solution collected from the surface of

Table 2 Ranked set sample data for the percentage area covered on the surface of the leaves of apple trees

Group	Copy	Rank 1	Rank 2	Rank 3	Rank 4	Rank 5
Control	1	0.003	0.028	0.244	0.057	0.143
	2	0.039	0.119	0.126	0.105	0.565
	3	0.034	0.118	0.130	0.218	0.296
	4	0.051	0.104	0.193	0.210	0.150
	5	0.032	0.141	0.130	0.250	0.229
	6	0.069	0.070	0.260	0.225	0.285
	7	0.100	0.091	0.244	0.130	0.347
	8	0.012	0.096	0.069	0.373	0.133
	9	0.046	0.117	0.126	0.223	0.273
	10	0.028	0.083	0.108	0.212	0.261
Treatment	1	0.036	0.137	0.183	0.270	0.487
	2	0.250	0.181	0.290	0.328	0.715
	3	0.089	0.032	0.269	0.419	0.315
	4	0.180	0.111	0.130	0.194	0.742
	5	0.100	0.009	0.184	0.277	0.122
	6	0.042	0.089	0.199	0.269	0.395
	7	0.044	0.083	0.227	0.177	0.742
	8	0.044	0.171	0.067	0.192	0.336
	9	0.009	0.017	0.217	0.438	0.544
	10	0.071	0.132	0.310	0.343	0.379

the leaves, and thereby is a time-consuming and expensive process. The judgment ranking within each set is based on the visual appearance of the spray deposits on the leaf surfaces when viewed under ultraviolet light. Clearly, the latter method is cheap, and fairly accurate if implemented by an expert observer.

The data are given in Table 2, where measurements obtained from the plot sprayed at high (low) volume constitute the control (treatment) group. Suppose the interest is in knowing whether the sprayer settings affect the percentage area coverage. Then, $\hat{R}^*(t)$ can serve as a measure of the treatment effect. Finally, threshold t in our proposed estimator may be interpreted as a lower bound on the response values which is easily available from previous studies or experts' opinions. Figure 13 shows $\hat{R}^*(t)$ as a function of t for the apple trees data. The horizontal line indicates the estimated value of the usual stress-strength reliability, $R = P(X > Y)$, which is equal to 0.6184. It is to be noted that how information about t could update the basic estimate of the treatment effect.

7 Conclusion

The challenge we have set ourselves in this work is estimation of a time-dependent reliability measure using RSS. Under an imperfect ranking model, components of the

suggested estimator are shown to be more efficient than their competitors in SRS. It is further established that their variances are minimized in the absence of ranking errors. The findings are supported by results of Monte Carlo experiment conducted to get insight of the behavior of the new measure.

The proposed estimator is the ratio of two estimators based on empirical distributions. A deficiency shared by such estimators is that they fail to capture smoothness of the corresponding attributes being estimated. Moreover, the empirical estimators may reveal large bias close to the boundaries which stems from the fact that they are unable to estimate beyond the largest observation. Thus, it would be interesting to overcome these shortcomings by adopting a proper approach. This can be done using kernel density estimation in line with Zardasht et al. (2012). The method merits investigation in the context of estimating the usual stress-strength reliability based on RSS, as well. These will be studied in subsequent works.

Acknowledgements The authors are indebted to the reviewer and the Associate Editor for helpful comments on the paper.

Appendix

See Figures 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 and 13.

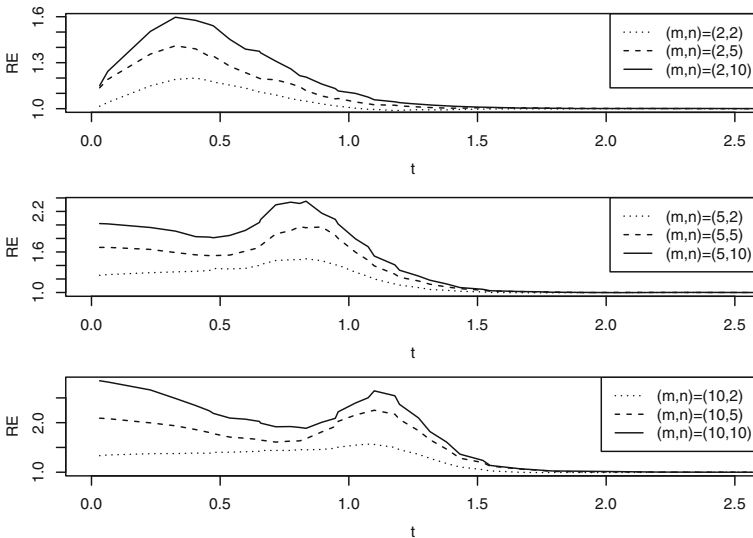


Fig. 1 Estimated REs for $X \sim W(2, 1)$ and $Y \sim W(2, 2)$ when $m, n = 2, 5, 10$

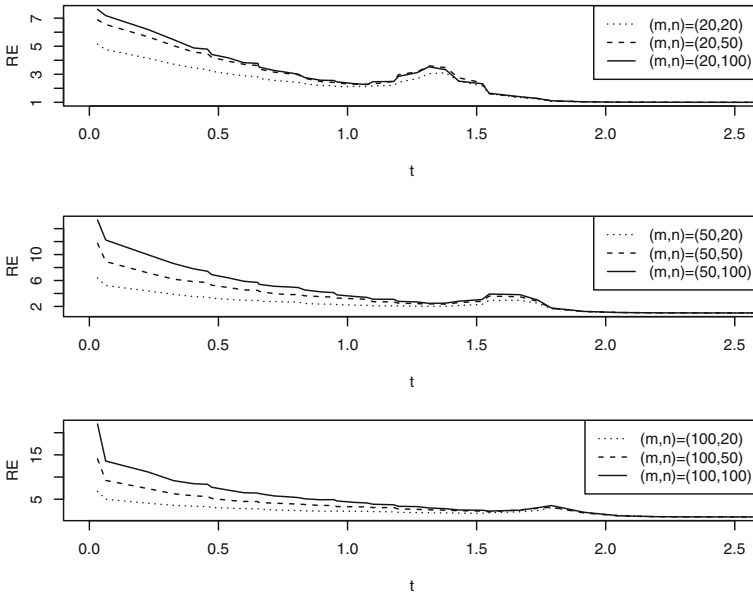


Fig. 2 Estimated REs for $X \sim W(2, 1)$ and $Y \sim W(2, 2)$ when $m, n = 20, 50, 100$

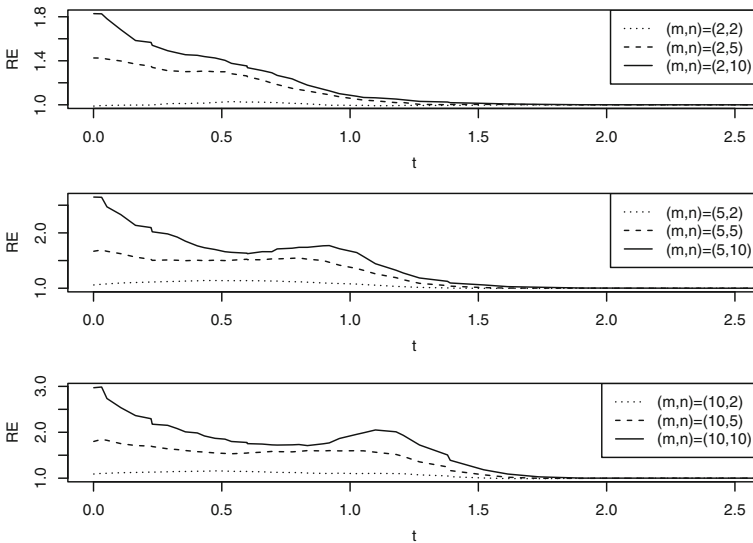


Fig. 3 Estimated REs for $X \sim W(2, 1)$ and $Y \sim W(1, 1)$ when $m, n = 2, 5, 10$

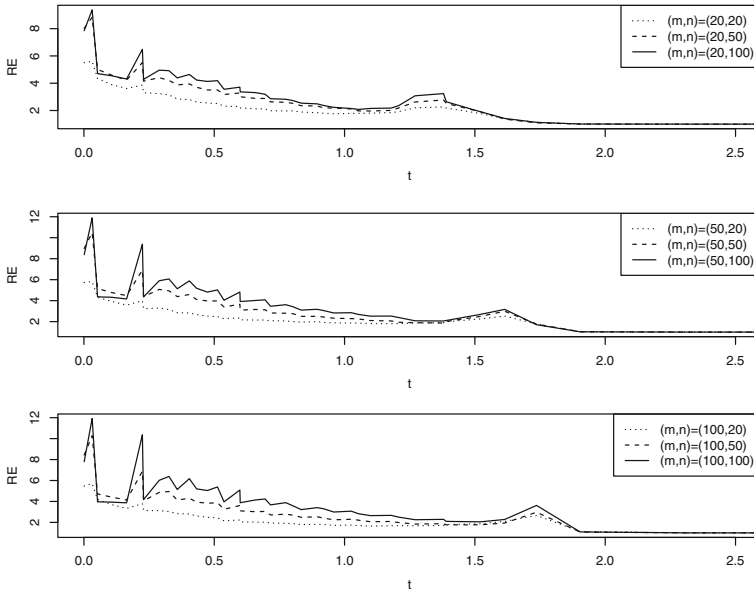


Fig. 4 Estimated REs for $X \sim W(2, 1)$ and $Y \sim W(1, 1)$ when $m, n = 20, 50, 100$

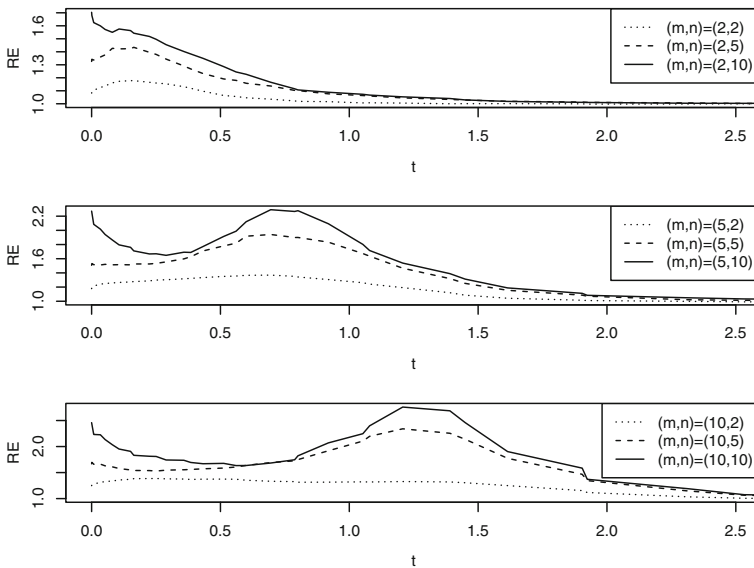


Fig. 5 Estimated REs for $X \sim W(1, 1)$ and $Y \sim W(0.5, 3)$ when $m, n = 2, 5, 10$

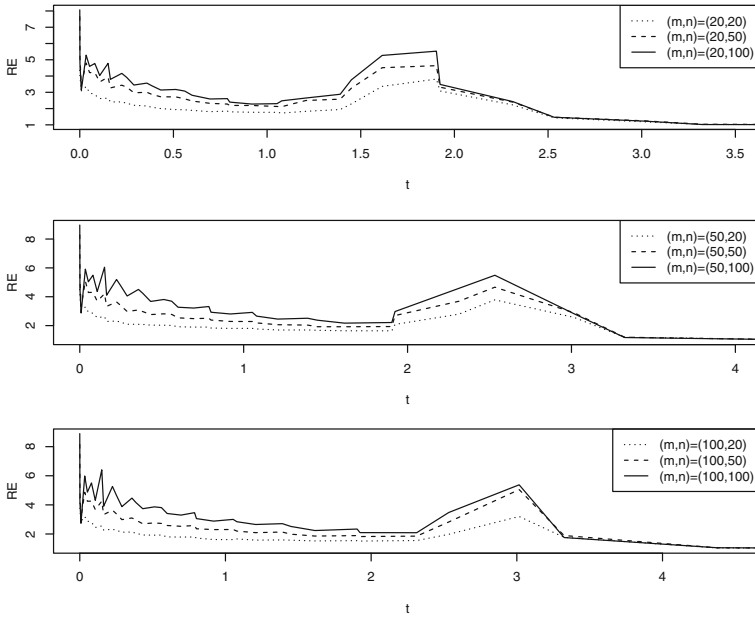


Fig. 6 Estimated REs for $X \sim W(1, 1)$ and $Y \sim W(0.5, 3)$ when $m, n = 20, 50, 100$

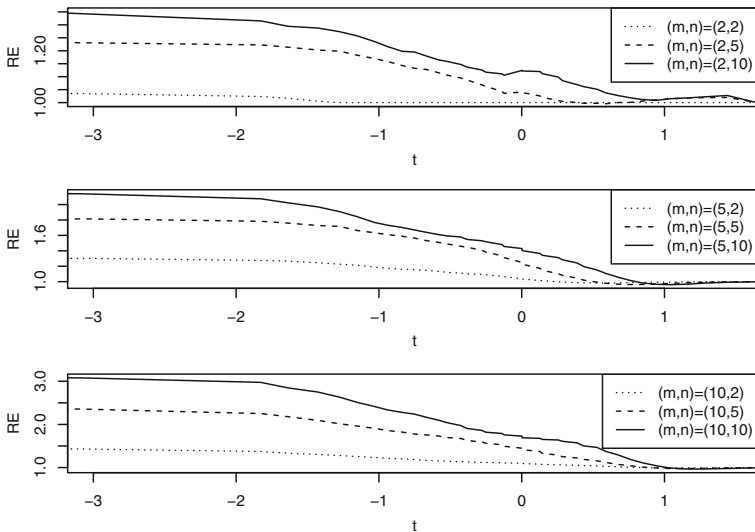


Fig. 7 Estimated REs for $X \sim N(0, (1.25)^2)$ and $Y \sim N(0, 1)$ when $m, n = 2, 5, 10$

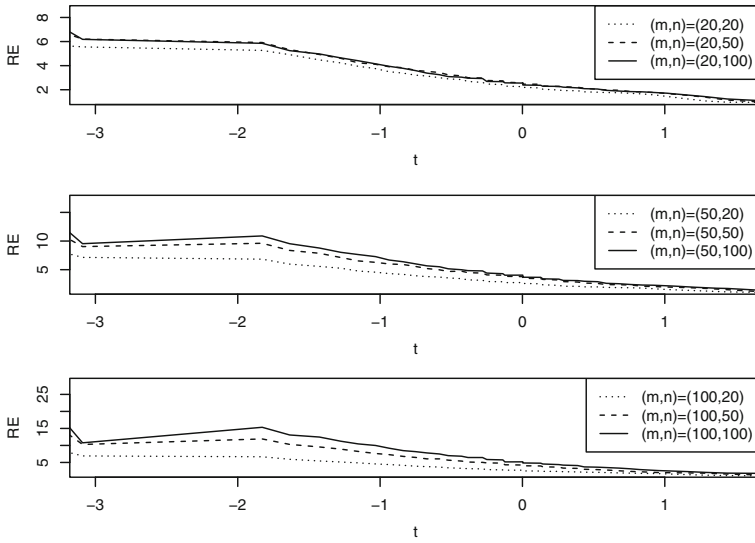


Fig. 8 Estimated REs for $X \sim N(0, (1.25)^2)$ and $Y \sim N(0, 1)$ when $m, n = 20, 50, 100$

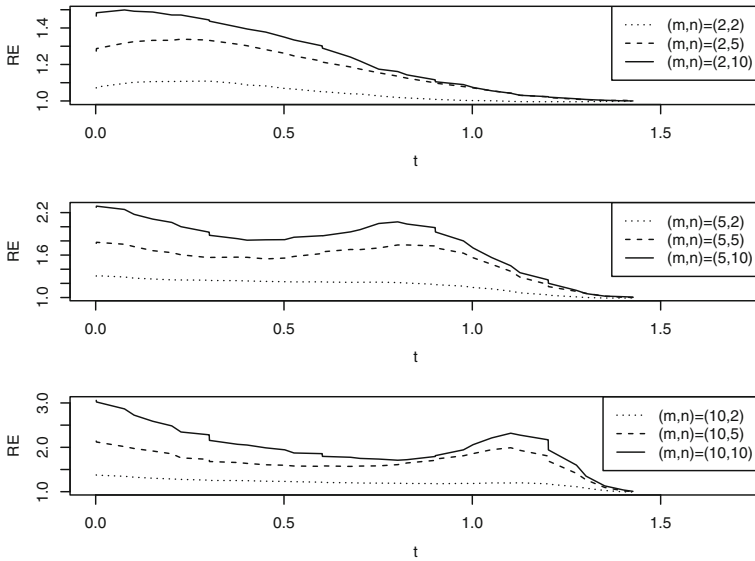


Fig. 9 Estimated REs for $X \sim U(0, 1.5)$ and $Y \sim U(0, 2)$ when $m, n = 2, 5, 10$

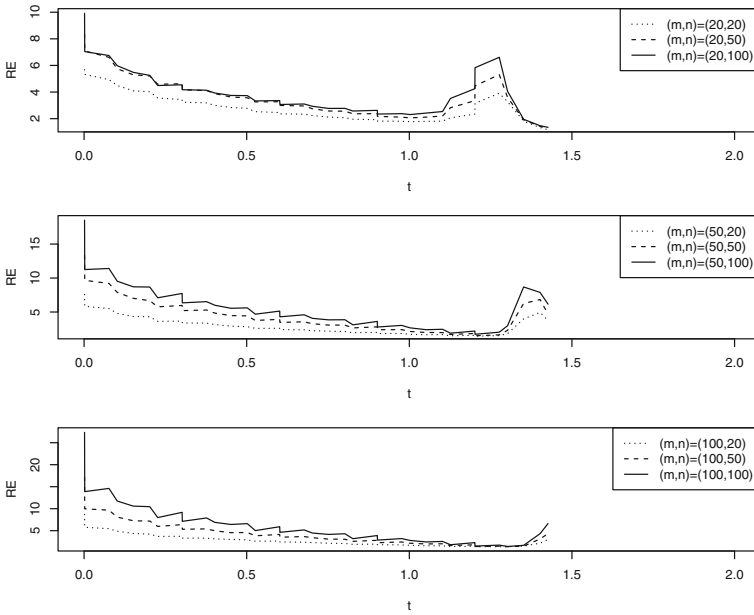


Fig. 10 Estimated REs for $X \sim U(0, 1.5)$ and $Y \sim U(0, 2)$ when $m, n = 20, 50, 100$

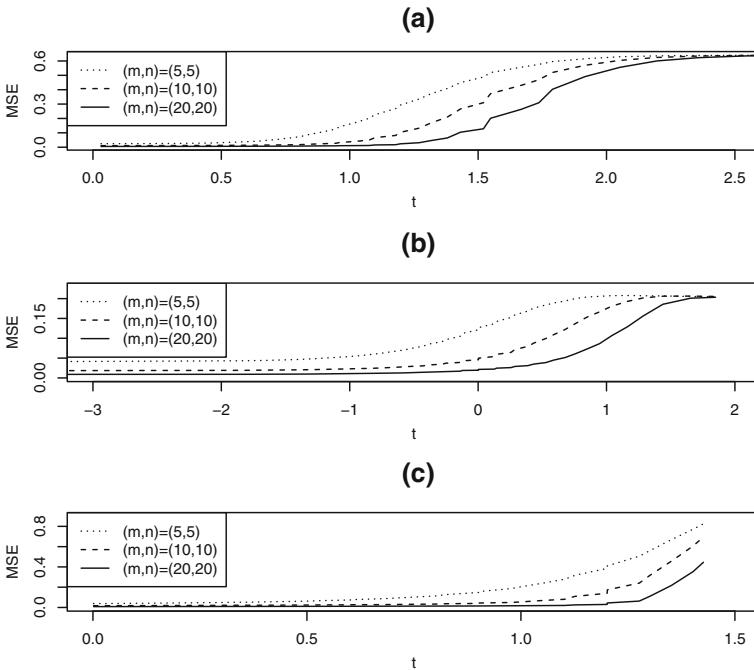


Fig. 11 Estimated MSEs under SRS when $m = n = 5, 10, 20$ for (a) $X \sim W(2, 1)$ and $Y \sim W(2, 2)$, (b) $X \sim N(0, (1.25)^2)$ and $Y \sim N(0, 1)$ and (c) $X \sim U(0, 1.5)$ and $Y \sim U(0, 2)$

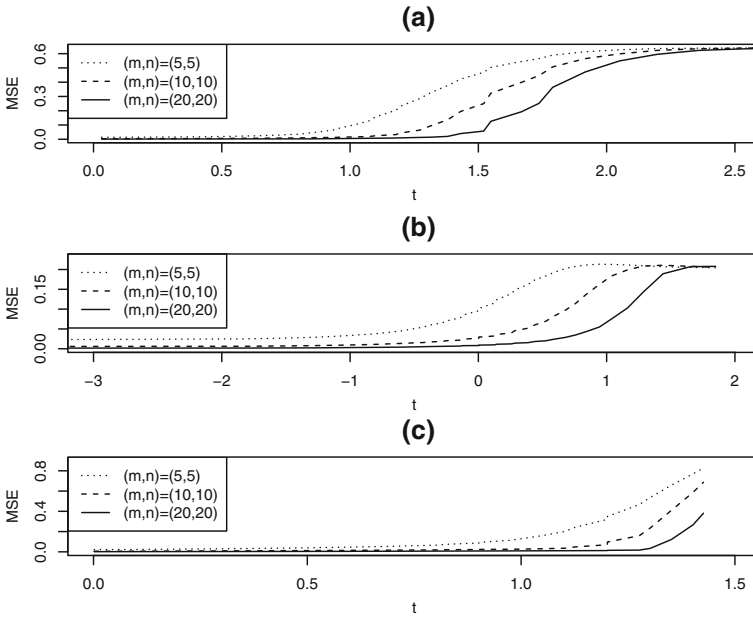


Fig. 12 Estimated MSEs under RSS when $m = n = 5, 10, 20$ for **(a)** $X \sim W(2, 1)$ and $Y \sim W(2, 2)$, **(b)** $X \sim N(0, (1.25)^2)$ and $Y \sim N(0, 1)$ and **(c)** $X \sim U(0, 1.5)$ and $Y \sim U(0, 2)$

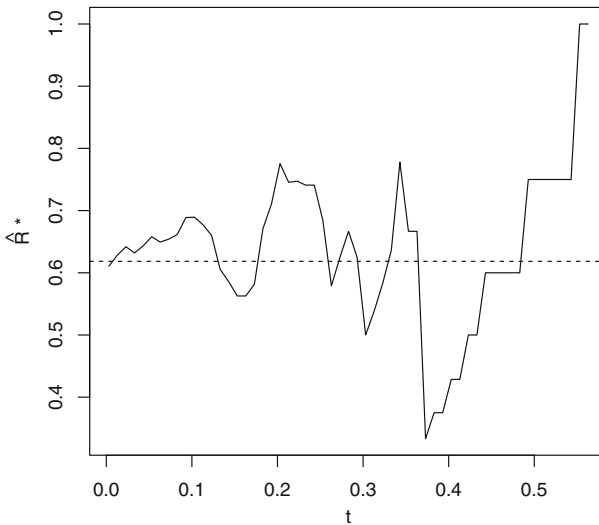


Fig. 13 Estimated $R(t)$ as a function of t for the apple trees data

References

Bohn LL, Wolfe DA (1994) The effect of imperfect judgment rankings on properties of procedures based on the ranked-set samples analog of the Mann-Whitney-Wilcoxon statistic. *J Am Stat Assoc* 89:168–176

- Chen Z, Bai Z, Sinha BK (2004) Ranked set sampling: theory and applications. Springer, New York
- Datta BN (2010) Numerical linear algebra and applications, 2nd edn. SIAM, Philadelphia
- Díaz-Francés E, Montoya JA (2013) The simplicity of likelihood based inferences for $P(X < Y)$ and for the ratio of means in the exponential model. *Stat Pap* 54:499–522
- Kotz S, Lumelskii Y, Pensky M (2003) The stress-strength model and its generalizations. Theory and applications. World Scientific, Singapore
- Kvam PH, Samaniego FJ (1994) Nonparametric maximum likelihood estimation based on ranked set samples. *J Am Stat Assoc* 89:526–537
- McIntyre GA (1952) A method of unbiased selective sampling using ranked sets. *Aust J Agric Res* 3:385–390
- Murray RA, Ridout MS, Cross JV (2000) The use of ranked set sampling in spray deposit assessment. *Asp Appl Biol* 57:141–146
- Sengupta S, Mukhuti S (2008) Unbiased estimation of $P(X > Y)$ using ranked set sample data. *Statistics* 42:223–230
- Shao J (2003) Mathematical statistic, 2nd edn. Springer, New York
- Zardasht V, Asadi M (2010) Evaluation of $P(X_t > Y_t)$ when both X_t and Y_t are residual lifetimes of two systems. *Stat Neerl* 64:460–481
- Zardasht V, Zeepongsekul P, Asadi M (2012) On nonparametric estimation of a reliability function. *Commun Stat* 41:983–999