Probability in the Engineering and Informational Sciences, **36**, 2022, 644–659.
doi:10.1017/S0269964820000716

ON EXTROPY PROPERTIES AND DISCRIMINATION INFORMATION OF DIFFERENT STRATIFIED SAMPLING SCHEMES

ABBAS EFTEKHARIAN
Department of Statistics, School of Basic Sciences, University of Hormozgan, Bandar Abbas, Iran
E-mail: eftekharian@hormozgan.ac.ir

GUOXIN QIU
Department of Business Administration, School of Business, Xinhua University of Anhui, Hefei, Anhui 230031, China
Department of Statistics and Finance, School of Management, University of Science and Technology of China, Hefei, Anhui 230026, China

Ranked set sampling (RSS) and some of its variants are sampling designs that are applied widely in different areas. When the underlying population contains different subpopulations, we can use stratified ranked set sampling (SRSS) which combines the advantages of stratification with RSS. In the present paper, we consider the information content of SRSS in terms of extropy measure. Some results using stochastic orders properties are obtained. The effect of imperfect ranking on discrimination information is analytically investigated. It is proved that discrimination information between the perfect SRSS and simple random sampling (SRS) data sets performs better than that of between the imperfect SRSS and SRS data sets.

Keywords: discrimination information, extropy, imperfect ranking, stratified ranked set sampling

1. INTRODUCTION

Ranked set sampling (RSS) as a new sampling scheme was originally proposed by McIntyre [9] to find a more efficient estimator of the mean pasture yields. RSS and some of its variants have been used widely in different applied areas such as economics, industrial application, biostatistics, environmental and ecological studies. According to the original form of RSS, r units (called a set of size r) are drawn randomly from the underlying population. The units of this set are ranked by visual inspection or judgment ranking without actual measurement. From the ranked set, the unit ranked lowest is taken for the actual measurement. Next, a second set of size r is drawn and the units of the set are ranked in the same way the first set. The unit with the second lowest rank is selected for quantifying actually. This process is followed until, from the rth set, the r ranked unit is quantified. The process can be repeated

© The Author(s), 2021. Published by Cambridge University Press 0269-9648/19 $25.00

https://doi.org/10.1017/S0269964820000716 Published online by Cambridge University Press

m times (or cycles) to get an RSS data set of size $n = mr$ from the underlying population. If judgment ranking is accurate, then the ranking is perfect and the selected units are actually a sample of r independent order statistics and its data set with m cycles is denoted by $\boldsymbol{X}_{\mathrm{RSS}}^{(n)} = \{X_{(i:r)j},\ i = 1, \ldots, r,\ j = 1, \ldots, m\}$, where $X_{(i:r)j}$ is the ith order statistic from a set of size r in the jth cycle. If judgment ranking has error, then we have imperfect ranking scenario such that the ith judgment order statistic, will no longer be the ith order statistic and corresponding data set is $\boldsymbol{X}_{\mathrm{RSS}}^{*(n)} = \{X_{[i:r]j},\ i = 1, \ldots, r,\ j = 1, \ldots, m\}$.

The authors developed the idea of RSS scheme by introducing some variants of the original RSS scheme. Stratified RSS (SRSS) is an extension of RSS which was proposed by Samawi [15]. SRSS is a sampling scheme in which a population is separated into L mutually distinct strata, with set size r_h of ranked set sample which is quantified within stratum h. In other words, this scheme can be considered as a set of L distinct ranked set samples. To explain SRSS, suppose that the population is divided into L mutually distinct strata and

$$(X_{h11}^*, X_{h12}^*, \ldots, X_{h1r_h}^*); (X_{h21}^*, X_{h22}^*, \ldots, X_{h2r_h}^*); \ldots; (X_{hr_h1}^*, X_{hr_h2}^*, \ldots, X_{hr_hr_h}^*),$$

be r_h independent random samples with size r_h selected from the hth stratum ($h = 1, 2, \ldots, L$). Hereafter, assume that X_{hji} stands for the quantitative measurement of the unit X_{hji}^*. Let $(X_{h1i}, X_{h2i}, \ldots, X_{hr_hi},\ i = 1, 2, \ldots, m_h)$ be the simple random sample with size $n_h = m_h r_h$ which is collected from the hth subpopulation characterized by $F_h(x)$ and $f_h(x)$ as the cumulative distribution function (cdf) and probability density function (pdf), respectively. It should be highlighted that the cdf of underlying population can be written as $F(x) = \sum_{h=1}^{L} W_h F_h(x)$, where W_h is the weight of the hth stratum and predetermined from the sampling scheme. To performing RSS scheme in the population with h subpopulation, we can consider the SRSS scheme as a two-phase sampling scheme. First, rank the units in each sample with respect to the variable of interest without actual measurement. In the hth stratum, the ordered sample corresponding to the jth sampling units is denoted by $(X_{hj(1:r_h)i}^*, X_{hj(2:r_h)i}^*, \ldots, X_{hj(r_h:r_h)i}^*,\ i = 1, 2, \ldots, m_h)$. Second, the ranked set sample with size $n_h = m_h r_h$ in the hth stratum is given by $(X_{h1(1:r_h)i}, X_{h2(2:r_h)i}, \ldots, X_{hr_h(r_h:r_h)i},\ i = 1, 2, \ldots, m_h)$, where $X_{hk(t:k)i}$ is the tth order statistic from the kth sample of size k in the ith cycle of the hth stratum. Also, suppose that $N_1, N_2, \ldots, N_L$ are the number of sampling units within each stratum such that $N = \sum_{h=1}^{L} N_h$ represents the total underlying population size. In addition, let $n_1, n_2, \ldots, n_L$ denote the number of sampling units measured within each stratum such that $n = \sum_{h=1}^{L} n_h$ represents the total sample size. From now on, without loss of generality, we assume that $m_h = 1$ for any $h = 1, \ldots, L$. So, $n_h = r_h$ and for simplicity of notation, we set $X_{hk(t:k)i} = X_{ht:k}$, where $X_{ht:k}$ is the tth order statistic from a sample of size k of the hth stratum for $k = 1, \ldots, n_h$. Furthermore, the notations $\boldsymbol{X}_{\mathrm{SSRS}}^{(n)}$ and $\boldsymbol{X}_{\mathrm{SRSS}}^{(n)}$ denote the data sets of stratified simple random sampling (SSRS) and SRSS of the same size n, respectively.

Fisher information has a key role in statistical inference and analysis of information. It is shown that the Fisher information content of an RSS data set is more than that of a simple random sampling (SRS) data set of the same size [3, p. 58, 5,10]. From an engineering point of view, the Shannon entropy or the Rènyi information might be more proper to be used as a measure to quantify the information content of a sample than the Fisher information. Researchers have paid considerable attention to these measures to quantify the uncertainty and information in the sense of the utility of using RSS data set instead of the commonly used SRS data set. For this line, we refer the reader to Jozani and Ahmadi [7], Eskandarzadeh *et al.* [4] and Tahmasebi *et al.* [17]. Recently, Lad *et al.* [8] introduced a new information measure termed by extropy. The extropy of a random variable X with

https://doi.org/10.1017/S0269964820000716 Published online by Cambridge University Press

pdf f is defined as

$$J(X) = -\frac{1}{2}\int_{-\infty}^{\infty} f^2(x)\,\mathrm{d}x = -\frac{1}{2}\int_0^1 f(F^{-1}(u))\,\mathrm{d}u,$$

where $F^{-1}(u) = \inf\{t : F(t) \geq u\}$ is the quantile function of cdf F. Lad *et al.* [8] presented some properties of extropy measure with their applications. Qiu [11] studied the monotonic properties of extropy based on order statistics and record values. Some results on the extropy properties of mixed systems were published by Qiu *et al.* [13]. The cumulative residual extropy and its properties were studied by Jahanshahi *et al.* [6]. Raqab and Qiu [14] investigated the properties of extropy on RSS data and suggested a discrimination information concept based on extropy as a discrimination tool between different data sets. Qiu and Eftekharian [12] developed some properties of extropy using stochastic orders on maximum and minimum RSS with unequal samples. In the present paper, we study the notions of the information content of SRSS data set under extropy information measure and monotonic properties using stochastic orders. Moreover, we investigate the effect of imperfect ranking on discrimination information and obtain discrimination information between RSS data in both perfect and imperfect ranking scenarios and its counterpart under SRS data and then extend the results to SRSS data.

The paper is organized as follows. Section 2 is provided the expressions of extropy of SRSS. Some comparison and bounds results are reported in Section 3. In Section 4, some results on the discrimination information between RSS data in both perfect and imperfect ranking scenarios and its counterpart under SRS data are obtained. Throughout the present paper, all random variables are implicitly assumed to be absolutely continuous. The term *increasing* is used for monotone non-decreasing and the term *decreasing* is used for monotone non-increasing.

2. EXTROPY BASED ON THE SRSS

Let $\boldsymbol{X}_{h\mathrm{SRS}}^{(n_h)} = (X_{h1}, \ldots, X_{hn_h})$ be a simple random sample of size n_h from X_h with pdf $f_h(x)$ based on the SRS method from the hth stratum. Thus, the extropy of the hth stratum is given by

$$\begin{aligned} J(\boldsymbol{X}_{h\mathrm{SRS}}^{(n_h)}) &= -\frac{1}{2}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f_h^2(x_1)\cdots f_h^2(x_{n_h})\,\mathrm{d}x_1\cdots\mathrm{d}x_{n_h} \\ &= -\frac{1}{2}\prod_{i=1}^{n_h}\int_{-\infty}^{\infty} f_h^2(x_i)\,\mathrm{d}x_i \\ &= -\frac{1}{2}[-2J(X_h)]^{n_h}, \end{aligned}$$

where $J(X_h) = -\frac{1}{2}\int_{-\infty}^{\infty} f_h^2(x)\,\mathrm{d}x = -\frac{1}{2}\int_0^1 f_h(F_h^{-1}(u))\,\mathrm{d}u$. Therefore, it can be shown that

$$J(\boldsymbol{X}_{\mathrm{SSRS}}^{(n)}) = -\frac{1}{2}\int\cdots\int\left(\sum_{h=1}^{L} W_h f_h(x_{h1})\cdots f_h(x_{hn_h})\right)^2 \mathrm{d}x_{11}\cdots\mathrm{d}x_{Ln_L}$$

https://doi.org/10.1017/S0269964820000716 Published online by Cambridge University Press

$$= -\frac{1}{2}\sum_{h=1}^{L} W_h^2[-2J(X_h)]^{n_h} - \sum_{h<t}\sum W_h W_t$$

$$= \sum_{h=1}^{L} W_h^2 J(\boldsymbol{X}_{h\mathrm{SRS}}^{(n_h)}) - \sum_{h<t}\sum W_h W_t.$$

Now, let $\boldsymbol{X}_{h\mathrm{RSS}}^{(n_h)} = (X_{h1:n_h}, X_{h2:n_h}, \ldots, X_{hn_h:n_h})$ be a sample of size n_h from the hth stratum based on the RSS scheme. Then, the extropy of the hth stratum is as follows:

$$J(\boldsymbol{X}_{h\mathrm{RSS}}^{(n_h)}) = -\frac{1}{2}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f_{h1:n_h}^2(x_1)\cdots f_{hn_h:n_h}^2(x_{n_h})\,\mathrm{d}x_1\cdots\mathrm{d}x_{n_h}$$

$$= -\frac{1}{2}\prod_{i=1}^{n_h}\int_{-\infty}^{\infty} f_{hi:n_h}^2(x_i)\,\mathrm{d}x_i = -\frac{1}{2}\prod_{i=1}^{n_h}[-2J(X_{hi:n_h})], \tag{1}$$

by considering

$$J(X_{hi:n_h}) = -\frac{1}{2}\int_{-\infty}^{\infty} f_{hi:n_h}^2(x)\,\mathrm{d}x = -\frac{1}{2}\int_0^1 f_{hi:n_h}(F_{hi:n_h}^{-1}(u))\,\mathrm{d}u,$$

where $f_{hi:n_h}(\cdot)$ and $F_{hi:n_h}^{-1}(\cdot)$ are the pdf and quantile function of the ith order statistic from the h stratum, respectively. Therefore, it is easy to see that

$$J(\boldsymbol{X}_{\mathrm{SRSS}}^{(n)}) = -\frac{1}{2}\int\cdots\int\left(\sum_{h=1}^{L} W_h f_{h1:n_h}(x_{h1})\cdots f_{hn_h:n_h}(x_{hn_h})\right)^2 \mathrm{d}x_{11}\cdots\mathrm{d}x_{Ln_L}$$

$$= -\frac{1}{2}\sum_{h=1}^{L} W_h^2\prod_{i=1}^{n_h}[-2J(X_{hi:n_h})] - \sum_{h<t}\sum W_h W_t$$

$$= \sum_{h=1}^{L} W_h^2 J(\boldsymbol{X}_{h\mathrm{RSS}}^{(n_h)}) - \sum_{h<t}\sum W_h W_t. \tag{2}$$

In special case, when $L = 1$, then $J(\boldsymbol{X}_{\mathrm{SRSS}}^{(n)}) = J(\boldsymbol{X}_{\mathrm{RSS}}^{(n)})$. In the following, we apply Eqs. (1) and (2) to obtain the following proposition.

PROPOSITION 1: *Suppose that $\boldsymbol{X}_{SRSS}^{(n)}$ is the SRSS data set of size n with f_h and F_h as the pdf and cdf of the hth stratum, for $h = 1, \ldots, L$, respectively. Then, its extropy is given by*

$$J(\boldsymbol{X}_{SRSS}^{(n)}) = -\frac{1}{2}\sum_{h=1}^{L} W_h^2 c_{n_h}\prod_{i=1}^{n_h}\int_0^1 f_h(F_h^{-1}(u))b_{2i-1,2n_h-1}(u)\,du - \sum_{h<t}\sum W_h W_t, \tag{3}$$

where

$$c_{n_h} = n_h^{n_h}\prod_{i=1}^{n_h} a_{i,n_h}, \quad a_{i,n_h} = \frac{\binom{2i-2}{i-1}\binom{2n_h-2i}{n_h-i}}{\binom{2n_h-1}{n_h-1}}, \tag{4}$$

and

$$b_{2i-1,2n_h-1}(u) = (2n_h-1)\binom{2n_h-2}{2i-2}u^{2i-2}(1-u)^{2n_h-2i}, \quad 0 < u < 1, \tag{5}$$

stands for the pdf of Beta$(2i-1, 2n_h-2i+1)$.

https://doi.org/10.1017/S0269964820000716 Published online by Cambridge University Press

PROOF: From [14], the extropy of RSS for the hth stratum with set size n_h is as follows:

$$J(\boldsymbol{X}_{h\mathrm{RSS}}^{(n_h)}) = -\frac{c_{n_h}}{2}\prod_{i=1}^{n_h}\int_0^1 f_h(F_h^{-1}(u))b_{2i-1,2n_h-1}(u)\,\mathrm{d}u, \tag{6}$$

where c_{n_h} and $b_{2i-1,2n_h-1}(u)$ are given by (4) and (5), respectively. Substituting $J(\boldsymbol{X}_{h\mathrm{RSS}}^{(n_h)})$ in the first term of Eq. (2) by the right-hand side of Eq. (6), we have Eq. (3). This completes the proof. ■

EXAMPLE 1: *Assume that Z_h is a random variable of exponential distribution with mean $1/\lambda_h$, for $h = 1, \ldots, L$. Thus, it is easy to check that*

$$J(\boldsymbol{Z}_{SSRS}^{(n)}) = -\frac{1}{2}\sum_{h=1}^{L} W_h^2(2\lambda_h)^{n_h} - \sum_{h<t}\sum W_h W_t,$$

$$J(\boldsymbol{Z}_{SRSS}^{(n)}) = -\frac{1}{2}\sum_{h=1}^{L} W_h^2 c_{n_h}\prod_{i=1}^{n_h}\left[\frac{2n_h-2i+1}{2n_h}\right]\lambda_h^{n_h} - \sum_{h<t}\sum W_h W_t.$$

In the next section, we provide some results on the extropy of $\boldsymbol{X}_{\mathrm{SRSS}}^{(n)}$ using stochastic orders.

3. COMPARISONS AND BOUNDS

In this section, we present some properties of the extropy of $\boldsymbol{X}_{\mathrm{SRSS}}^{(n)}$ on the basis of the some concepts of stochastic orders. For this purpose, we first express some definitions of aging properties and stochastic orders which exist between two random variables. For more details, the reader can refer to Shaked and Shanthikumar [16].

DEFINITION 1: *Let X and Y be two random variables with cdfs F and G, survival functions $\bar{F}$ and $\bar{G}$, and pdfs f and g, respectively.*

(1) The random variable X is called to have increasing (decreasing) failure rate [IFR (DFR)] *if $r_X(x)$ is increasing (decreasing) in $x \geq 0$, where $r_X = f/\bar{F}$ is the failure rate function of X;*

(2) The random variable X is said to have increasing (decreasing) failure rate in average [IFRA (DFRA)] *if $\int_0^x r_X(t)\,dt/x$ is increasing (decreasing) in $x \geq 0$;*

(3) The random variable X is called to be new better (worse) than used [NBU (NWU)] *if $\bar{F}(x+y) \leq (\geq)\bar{F}(x)\bar{F}(y)$ for all $x, y \geq 0$;*

(4) The random variable X is told to be smaller than Y in the convex transform order, denoted by $X \leq_c Y$, if $G^{-1}F(x)$ is convex in $x \geq 0$;

(5) The random variable X is called to be smaller than Y in the star order, denoted by $X \leq_* Y$, if $G^{-1}F(x)$ is star-shaped, equivalently, $G^{-1}F(x)/x$ is increasing in $x \geq 0$;

(6) The random variable X is said to be smaller than Y in the super-additive order, denoted by $X \leq_{su} Y$, if $G^{-1}F(x)$ is super-additive, equivalently, $G^{-1}F(x+y) \geq G^{-1}F(x) + G^{-1}F(y)$ for all $x, y \geq 0$;

(7) The random variable X is told to be smaller than Y in the dispersive order, denoted by $X \leq_{disp} Y$, if $G^{-1}F(x) - x$ is increasing in $x \geq 0$.

https://doi.org/10.1017/S0269964820000716 Published online by Cambridge University Press

From Definition 1, we first show that if the underlying distribution is increasing in the sense of dispersive order, then the extropy of SRSS data increases using the following theorem.

THEOREM 1: *If* $X \leq_{disp} Y$, *then* $J(\boldsymbol{X}^{(n)}_{SRSS}) \leq J(\boldsymbol{Y}^{(n)}_{SRSS})$.

PROOF: Using the definition of dispersive order, we can write $f(F^{-1}(u)) \geq g(G^{-1}(u))$, for all $0 < u < 1$. So, we get

$$J(\boldsymbol{Y}^{(n)}_{\mathrm{SRSS}}) - J(\boldsymbol{X}^{(n)}_{\mathrm{SRSS}}) = -\frac{1}{2}\sum_{h=1}^{L} W_h^2 c_h \prod_{i=1}^{n_h} \int_0^1 [g(G^{-1}(u)) - f(F^{-1}(u))] b_{2i-1:2n_h-1}(u)\,\mathrm{d}u \geq 0.$$

■

Suppose that X and Y are two non-negative random variables with pdfs f and g, respectively, provided that $f(0) \geq g(0) > 0$. If $X \leq_{\mathrm{su}} Y$ ($X \leq_* Y$ or $X \leq_{\mathrm{c}} Y$), then $X \leq_{\mathrm{disp}} Y$ [1]. Therefore, using Theorem 1, the following corollary is obtained.

COROLLARY 1: *Suppose that* X *and* Y *are two non-negative random variables with pdfs* f *and* g, *respectively, provided that* $f(0) \geq g(0) > 0$. *If* $X \leq_{\mathrm{su}} Y$ ($X \leq_* Y$ *or* $X \leq_{\mathrm{c}} Y$), *then* $J(\boldsymbol{X}^{(n)}_{SRSS}) \leq J(\boldsymbol{Y}^{(n)}_{SRSS})$.

Now, we verify that the extropy of SRSS data decreases when the sample size increases.

THEOREM 2: *Let* X *be DFR with* $f(F^{-1}(u)) \geq 1$ *for all* $0 < u < 1$. *Then,* $J(\boldsymbol{X}^{(n)}_{SRSS})$ *is decreasing in* $n \geq 1$.

PROOF: It is enough to show that $J(\boldsymbol{X}^{(n+L)}_{\mathrm{SRSS}}) \leq J(\boldsymbol{X}^{(n)}_{\mathrm{SRSS}})$. For this purpose, we set

$$\frac{J(\boldsymbol{X}^{(n+L)}_{\mathrm{SRSS}})}{J(\boldsymbol{X}^{(n)}_{\mathrm{SRSS}})} = \frac{-\frac{1}{2}\sum_{h=1}^{L} W_h^2 \prod_{i=1}^{n_h+1}[-2J(X_{hi:n_h+1})] - \sum_{h<t}\sum W_h W_t}{-\frac{1}{2}\sum_{h=1}^{L} W_h^2 \prod_{i=1}^{n_h}[-2J(X_{hi:n_h})] - \sum_{h<t}\sum W_h W_t}$$
$$= \frac{\sum_{h=1}^{L} W_h^2 \{\prod_{i=1}^{n_h}[-2J(X_{hi:n_h+1})][-2J(X_{hn_h+1:n_h+1})]\} + 2\sum_{h<t}\sum W_h W_t}{\sum_{h=1}^{L} W_h^2 \prod_{i=1}^{n_h}[-2J(X_{hi:n_h})] + 2\sum_{h<t}\sum W_h W_t},$$

where $X_{hi:n_h+1}$ is the ith order statistic from a sample of size $n_h + 1$. It is obvious that both the numerator and the denominator are positive in the last equality. From Raqab and Qiu [14] since $-2J(X_{hn_h+1:n_h+1}) \geq 1$, therefore

$$\frac{J(\boldsymbol{X}^{(n+L)}_{\mathrm{SRSS}})}{J(\boldsymbol{X}^{(n)}_{\mathrm{SRSS}})} \geq \frac{\sum_{h=1}^{L} W_h^2 \prod_{i=1}^{n_h}[-2J(X_{hi:n_h+1})] + 2\sum_{h<t}\sum W_h W_t}{\sum_{h=1}^{L} W_h^2 \prod_{i=1}^{n_h}[-2J(X_{hi:n_h})] + 2\sum_{h<t}\sum W_h W_t}.$$

However, we have $J(X_{hi:n_h+1}) \leq J(X_{hi:n_h})$ [14] so, we get $-2J(X_{hi:n_h+1}) \geq -2J(X_{hi:n_h})$. Hence,

$$\frac{J(\boldsymbol{X}^{(n+L)}_{\mathrm{SRSS}})}{J(\boldsymbol{X}^{(n)}_{\mathrm{SRSS}})} \geq \frac{\sum_{h=1}^{L} W_h^2 \prod_{i=1}^{n_h}[-2J(X_{hi:n_h})] + 2\sum_{h<t}\sum W_h W_t}{\sum_{h=1}^{L} W_h^2 \prod_{i=1}^{n_h}[-2J(X_{hi:n_h})] + 2\sum_{h<t}\sum W_h W_t} = 1,$$

and this completes the proof by noting that $J(\boldsymbol{X}^{(n)}_{\mathrm{SRSS}})$ is non-positive. ■

https://doi.org/10.1017/S0269964820000716 Published online by Cambridge University Press

In what follows, we extend the results of Raqab and Qiu [14] to the extropy of SRSS. To this end, some results without proof of the extropy of SRSS are provided which the main idea of the proofs is taken from the extropy of RSS for the hth stratum. For more details, one can refer to Raqab and Qiu [14].

Let X be a non-negative random variable. Shaked and Shanthikumar [16, p. 218] illustrated that $X \in \mathrm{IFR(IFRA, NBU)}$ if and only if $X \leq_{\mathrm{c}} (\leq_*, \leq_{\mathrm{su}})Z$, where Z is a random variable of exponential distribution with mean $1/\lambda$. Now, we can extend this result for all strata. In other words, let X_h be a non-negative random variable as well as Z_h be as considered in Example 1 for any $h = 1, \ldots, L$. Then, for the hth stratum $X_h \in \mathrm{IFR(IFRA, NBU)}$ if and only if $X_h \leq_{\mathrm{c}} (\leq_*, \leq_{\mathrm{su}})Z_h$. Now, we can conclude the following proposition using Example 1 and Corollary 1.

Proposition 2: *If $X_h \in \mathrm{IFR(IFRA, NBU)}$ and $f_h(0) \geq \lambda_h$, for any $h = 1, \ldots, L$, then*

$$J(\boldsymbol{X}^{(n)}_{SRSS}) \leq -\frac{1}{2}\sum_{h=1}^{L} W_h^2 c_{n_h} \lambda_h^{n_h} \prod_{i=1}^{n_h} \frac{2n_h - 2i + 1}{2n_h} - \sum_{h<t}\sum W_h W_t.$$

Proposition 3: *Suppose that X_h's are non-negative random variables with $r_{X_h}(x) \leq \lambda_h$, for all $x \geq 0$ and $h = 1, \ldots, L$. Then,*

$$J(\boldsymbol{X}^{(n)}_{SRSS}) \geq -\frac{1}{2}\sum_{h=1}^{L} W_h^2 \left(\frac{n_h \lambda_h}{2}\right)^{n_h} - \sum_{h<t}\sum W_h W_t.$$

Now, we make a bound for the extropy of SRSS data in comparing with SSRS data. To do this, let

$$P_{i,n_h}(u) = \binom{n_h}{i} u^i (1-u)^{n_h - i}, \quad 0 \leq u \leq 1,\ 1 \leq i \leq n_h, \tag{7}$$

be the probability function of $Bin(n_h, u)$ which is the binomial distribution. The following theorem is the key to constructing the bound for the extropy data.

Theorem 3: *If all $n_h \geq 1$, then*

$$J(\boldsymbol{X}^{(n)}_{SRSS}) \geq -\frac{1}{2}\sum_{h=1}^{L} W_h^2 \left[\int_0^1 f_h(F_h^{-1}(u))\, du\right]^{n_h} n_h^{2n_h} \prod_{i=2}^{n_h - 1} P^2_{i-1,n_h-1}(\eta_{i,n_h}),$$

where $\eta_{i,n_h} = (2i-2)/(2n_h - 2)$. In particular, if $n_1 = n_2 = \cdots = n_L = n_0$ and $n = n_0 L$, then

$$\frac{J(\boldsymbol{X}^{(n)}_{SRSS}) + \sum_{h<t}\sum W_h W_t}{J(\boldsymbol{X}^{(n)}_{SSRS})} \leq n_0^{2n_0} \prod_{i=2}^{n_0 - 1} P^2_{i-1,n_0-1}(\eta_{i,n_0}).$$

Proof: From (5) and (7), we can write

$$b_{2i-1,2n_h-1}(u) = (2n_h - 1)P_{2i-2,2n_h-2}(u), \quad 0 \leq u \leq 1,\ 1 \leq i \leq n_h.$$

Since $b_{2i-1,2n_h-1}(u)$ is a unimodal function, so it is easy to see that its maximum is actually attained at $\eta_{i,n_h} = (2i-2)/(2n_h - 2)$. Furthermore, for $i = 1$ and $i = n_h$, we get

$$P_{0,2n_h-2}(u) = (1-u)^{2n_h-2} \leq 1 = P_{0,2n_h-2}(0),$$

and

$$P_{2n_h-2,2n_h-2}(u) = u^{2n_h-2} \leq 1 = P_{2n_h-2,2n_h-2}(1),$$

respectively.

https://doi.org/10.1017/S0269964820000716 Published online by Cambridge University Press

Now, using (3), we have

$$\begin{aligned}
J(\boldsymbol{X}^{(n)}_{\mathrm{SRSS}}) &= -\frac{1}{2}\sum_{h=1}^{L} W_h^2 c_{n_h}(2n_h-1)^{n_h}\prod_{i=1}^{n_h}\int_0^1 f_h(F_h^{-1}(u))P_{2i-2,2n_h-2}(u)\,\mathrm{d}u \\
&\quad -\sum_{h<t}\sum W_hW_t \\
&= -\frac{1}{2}\sum_{h=1}^{L} W_h^2 (n_h(2n_h-1))^{n_h}\prod_{i=1}^{n_h}\left[a_{i,n_h}\int_0^1 f_h(F_h^{-1}(u))P_{2i-2,2n_h-2}(u)\,\mathrm{d}u\right] \\
&\quad -\sum_{h<t}\sum W_hW_t \\
&\geq -\frac{1}{2}\sum_{h=1}^{L} W_h^2 (n_h(2n_h-1))^{n_h}\left[\int_0^1 f_h(F_h^{-1}(u))\,\mathrm{d}u\right]^{n_h} a_{1,n_h}a_{n_h,n_h} \\
&\quad \times\prod_{i=2}^{n_h-1} a_{i,n_h}P_{2i-2,2n_h-2}(\eta_{i,n_h}) - \sum_{h<t}\sum W_hW_t,
\end{aligned}$$

where $a_{1,n_h} = a_{n_h,n_h} = n_h/(2n_h-1)$. Thus,

$$\begin{aligned}
J(\boldsymbol{X}^{(n)}_{\mathrm{SRSS}}) &\geq -\frac{1}{2}\sum_{h=1}^{L} W_h^2\left[\int_0^1 f_h(F_h^{-1}(u))\,\mathrm{d}u\right]^{n_h} n_h^{n_h+2}(2n_h-1)^{n_h-2} \\
&\quad \times\prod_{i=2}^{n_h-1} a_{i,n_h}P_{2i-2,2n_h-2}(\eta_{i,n_h}) - \sum_{h<t}\sum W_hW_t.
\end{aligned}$$

But, it can be shown that

$$(2n_h-1)a_{i,n_h}P_{2i-2,2n_h-2}(\eta_{i,n_h}) = n_hP^2_{i-1,n_h-1}(\eta_{i,n_h}).$$

Therefore, we obtain

$$\begin{aligned}
J(\boldsymbol{X}^{(n)}_{\mathrm{SRSS}}) &\geq -\frac{1}{2}\sum_{h=1}^{L} W_h^2\left[\int_0^1 f_h(F_h^{-1}(u))\,\mathrm{d}u\right]^{n_h} n_h^{2n_h}\prod_{i=2}^{n_h-1} P^2_{i-1,n_h-1}(\eta_{i,n_h}) \\
&\quad -\sum_{h<t}\sum W_hW_t.
\end{aligned}$$

Now, by considering $n_1 = n_2 = \cdots = n_L = n_0$ and $n = n_0L$ and setting in the last inequality we have

$$J(\boldsymbol{X}^{(n)}_{\mathrm{SRSS}}) \geq -\frac{1}{2}\sum_{h=1}^{L} W_h^2\left[\int_0^1 f_h(F_h^{-1}(u))\,\mathrm{d}u\right]^{n_0} n_0^{2n_0}\prod_{i=2}^{n_0-1} P^2_{i-1,n_0-1}(\eta_{i,n_0}) - \sum_{h<t}\sum W_hW_t.$$

On the other hand,

$$J(\boldsymbol{X}^{(n)}_{\mathrm{SSRS}}) = -\frac{1}{2}\sum_{h=1}^{L} W_h\left[\int_0^1 f_h(F_h^{-1}(u))\,\mathrm{d}u\right]^{n_h} - \sum_{h<t}\sum W_hW_t.$$

https://doi.org/10.1017/S0269964820000716 Published online by Cambridge University Press

So, we can write

$$J(\boldsymbol{X}^{(n)}_{\mathrm{SRSS}}) \geq -\frac{1}{2}\sum_{h=1}^{L} W_h^2 \left[\int_0^1 f_h(F_h^{-1}(u))\,\mathrm{d}u\right]^{n_0} n_0^{2n_0} \prod_{i=2}^{n_0-1} P^2_{i-1,n_0-1}(\eta_{i,n_0}) - \sum_{h<t}\sum W_h W_t$$

$$\geq \left\{-\frac{1}{2}\sum_{h=1}^{L} W_h^2 \left[\int_0^1 f_h(F_h^{-1}(u))\,\mathrm{d}u\right]^{n_0} - \sum_{h<t}\sum W_h W_t\right\} n_0^{2n_0}$$

$$\times \prod_{i=2}^{n_0-1} P^2_{i-1,n_0-1}(\eta_{i,n_0}) - \sum_{h<t}\sum W_h W_t.$$

Hence,

$$J(\boldsymbol{X}^{(n)}_{\mathrm{SRSS}}) \geq J(\boldsymbol{X}^{(n)}_{\mathrm{SSRS}}) n_0^{2n_0} \prod_{i=2}^{n_0-1} P^2_{i-1,n_0-1}(\eta_{i,n_0}) - \sum_{h<t}\sum W_h W_t,$$

and the proof is completed by noting that $J(\boldsymbol{X}^{(n)}_{\mathrm{SSRS}})$ is non-positive. ■

In the next section, we describe the discrimination information of SRSS data and investigate it based on both perfect and imperfect rankings.

4. EFFECT OF IMPERFECT RANKING ON DISCRIMINATION INFORMATION

In this section, we discuss the discrimination information between the underlying population distribution and distribution of SRSS in both perfect and imperfect ranking scenarios. To simplify notation, we first write n and f instead of n_h and f_h, respectively, and obtain some results based on SRS and RSS data sets and then extend the results obtained to SSRS and SRSS data sets. Different models have been suggested for imperfect ranking in the literature such as the fraction of random rankings model, the concomitant model and the fraction inverse rankings model (see [2,18]). In the present paper, we apply the fraction of random rankings model. Based on this model, one can consider the cdf of the ith judgment order statistic from sample size n as a convex combination of the cdf of the true ith order statistic with probability α and the cdf of the underlying population with probability $1-\alpha$, that is,

$$F_{[i:n]}(x) = \alpha F_{i:n}(x) + (1-\alpha)F(x), \quad 0 \leq \alpha \leq 1.$$

Assuming the underlying distribution is absolutely continuous, we can conclude that

$$f_{[i:n]}(x) = \alpha f_{i:n}(x) + (1-\alpha)f(x), \quad 0 \leq \alpha \leq 1. \tag{8}$$

Using (8), we can define the discrimination information between distribution of imperfect ranking RSS data and either distribution of perfect ranking RSS data or distribution of SRS data. Before making these, we recall the following definition from Raqab and Qiu [14].

Definition 2: *Let X and Y be two random variables with pdfs f and g, respectively. The discrimination information between X and Y denoted by $D(X,Y)$ is defined as*

$$D(X,Y) \triangleq D(f,g) = \frac{1}{2}\int_{-\infty}^{\infty} [f(x)-g(x)]f(x)\,dx.$$

https://doi.org/10.1017/S0269964820000716 Published online by Cambridge University Press

Based on Definition 2, the discrimination information between the ith order statistic of sample size n from RSS data and the underlying distribution f can be written as follows:

$$D_n(f_{i:n}, f) = \frac{1}{2}\int_{-\infty}^{\infty}[f_{i:n}(x) - f(x)]f_{i:n}(x)\,\mathrm{d}x.$$

By a similar argument, the discrimination information between the ith judgment order statistic from RSS data and the underlying distribution is derived by

$$D_n^*(f_{[i:n]}, f) = \frac{1}{2}\int_{-\infty}^{\infty}[f_{[i:n]}(x) - f(x)]f_{[i:n]}(x)\,\mathrm{d}x.$$

Remark 1: *Suppose that* $\boldsymbol{X}_{SRS}^{(n)}$ *and* $\boldsymbol{X}_{RSS}^{(n)}$ *are the SRS and perfect RSS data sets of the same size* n*, respectively. Raqab and Qiu [14] showed that*

$$\sum_{i=1}^{n} D_n(f_{i:n}, f) = nJ(X) - \sum_{i=1}^{n} J(X_{i:n}). \tag{9}$$

Moreover, they concluded that $\sum_{i=1}^{n} D_n(f_{i:n}, f) \geq 0$*[14], that is,*

$$\sum_{i=1}^{n} J(X_{i:n}) \leq nJ(X). \tag{10}$$

In the following, we obtain the discrimination information between the imperfect RSS and SRS data sets.

Theorem 4: *Let* $\boldsymbol{X}_{SRS}^{(n)}$ *and* $\boldsymbol{X}_{RSS}^{*(n)}$ *be the SRS and imperfect RSS data sets of the same size* n*, respectively. Then,*

$$\sum_{i=1}^{n} D_n^*(f_{[i:n]}, f) = nJ(X) - \sum_{i=1}^{n} J(X_{[i:n]}), \tag{11}$$

where

$$J(X_{[i:n]}) = -\frac{1}{2}\int_{-\infty}^{\infty} f_{[i:n]}^2(x)\,dx.$$

Proof: From Definition 2, we have

$$\begin{aligned}\sum_{i=1}^{n} D_n^*(f_{[i:n]}, f) &= \frac{1}{2}\sum_{i=1}^{n}\int_{-\infty}^{\infty}[f_{[i:n]}(x) - f(x)]f_{[i:n]}(x)\,\mathrm{d}x \\ &= \frac{1}{2}\sum_{i=1}^{n}\int_{-\infty}^{\infty} f_{[i:n]}^2(x)\,\mathrm{d}x - \frac{1}{2}\sum_{i=1}^{n}\int_{-\infty}^{\infty} f(x)f_{[i:n]}(x)\,\mathrm{d}x \\ &= nJ(X) - J(X_{[i:n]}),\end{aligned}$$

which is due to the fact that $f(x) = 1/n\sum_{i=1}^{n} f_{[i:n]}(x)$. ■

The following lemma will be useful to find the relation between (9) and (11).

https://doi.org/10.1017/S0269964820000716 Published online by Cambridge University Press

LEMMA 1: *Let f and $f_{i:n}$ be the pdfs of the underlying distribution and the ith order statistic from sample size n, respectively. Then,*

$$f^2(x) \le \frac{1}{n}\sum_{i=1}^{n} f_{i:n}^2(x).$$

PROOF: From identity $f(x) = 1/n\sum_{i=1}^{n} f_{i:n}(x)$, we can get

$$0 \le \sum_{i=1}^{n}[f_{i:n}(x) - f(x)]^2 = \sum_{i=1}^{n} f_{i:n}^2(x) - nf^2(x),$$

and this is precisely the assertion of the lemma. ■

Now, we apply Lemma 1 to establish a relation between the discrimination information based on the perfect and imperfect RSS data sets. The following theorem is the key to constructing this goal.

THEOREM 5: *For $\boldsymbol{X}_{RSS}^{(n)}$ and $\boldsymbol{X}_{RSS}^{*(n)}$ data sets, we have*

$$\sum_{i=1}^{n} D_n^*(f_{[i:n]}, f) \le \sum_{i=1}^{n} D_n(f_{i:n}, f).$$

PROOF: From (9) and (11), it suffices to prove that

$$\sum_{i=1}^{n}[J(X_{i:n}) - J(X_{[i:n]})] \le 0,$$

which is equivalent to

$$d := \frac{1}{2}\sum_{i=1}^{n}\int_{-\infty}^{\infty}[f_{[i:n]}^2(x) - f_{i:n}^2(x)]\,\mathrm{d}x \le 0. \tag{12}$$

Using (8) and replacing in the last inequality, we have

$$\begin{aligned} d &= \sum_{i=1}^{n}\int_{-\infty}^{\infty}[\alpha^2 f_{i:n}^2(x) + (1-\alpha)^2 f^2(x) + 2\alpha(1-\alpha)f(x)f_{i:n}(x) - f_{i:n}^2(x)]\,\mathrm{d}x \\ &= \alpha^2\sum_{i=1}^{n}\int_{-\infty}^{\infty} f_{i:n}^2(x)\,\mathrm{d}x + n(1-\alpha)^2\int_{-\infty}^{\infty} f^2(x)\,\mathrm{d}x + 2n\alpha(1-\alpha)\int_{-\infty}^{\infty} f^2(x)\,\mathrm{d}x \\ &\quad - \sum_{i=1}^{n}\int_{-\infty}^{\infty} f_{i:n}^2(x)\,\mathrm{d}x \\ &= -(1-\alpha^2)\left\{\int_{-\infty}^{\infty}\left[\sum_{i=1}^{n} f_{i:n}^2(x) - nf^2(x)\right]\mathrm{d}x\right\}. \end{aligned}$$

Note that $(1-\alpha^2)$ is non-negative. Therefore, using Lemma 1, we can conclude that the last equality is non-positive which establishes Eq. (12) and the proof is completed. ■

https://doi.org/10.1017/S0269964820000716 Published online by Cambridge University Press

Moreover, we can extend the discrimination information to express the relation between distribution of the ith order statistic and distribution of the ith judgment order statistic as

$$D_n^{**} = \frac{1}{2}\int_{-\infty}^{\infty} [f_{i:n}(x) - f_{[i:n]}(x)] f_{i:n}(x)\,\mathrm{d}x. \tag{13}$$

Using D_n^{**}, we get the discrimination information between the perfect and imperfect RSS data sets.

THEOREM 6: *Let $\boldsymbol{X}_{RSS}^{(n)}$ and $\boldsymbol{X}_{RSS}^{*(n)}$ be the perfect and imperfect RSS data sets, respectively. Then,*

$$\sum_{i=1}^{n} D_n^{**}(f_{i:n}, f_{[i:n]}) = (1-\alpha)\left[nJ(X) - \sum_{i=1}^{n} J(X_{i:n})\right], \tag{14}$$

where $\alpha \in [0,1]$.

PROOF: Using (8) and (13), we get

$$\begin{aligned}
\sum_{i=1}^{n} D_n^{**}(f_{i:n}, f_{[i:n]}) &= \frac{1}{2}\sum_{i=1}^{n}\int_{-\infty}^{\infty} f_{i:n}^2(x)\,\mathrm{d}x - \frac{1}{2}\sum_{i=1}^{n}\int_{-\infty}^{\infty} f_{i:n}(x) f_{[i:n]}(x)\,\mathrm{d}x \\
&= -\sum_{i=1}^{n} J(X_{i:n}) - \frac{1}{2}\sum_{i=1}^{n}\int_{-\infty}^{\infty} f_{i:n}(x) f_{[i:n]}(x)\,\mathrm{d}x \\
&= -\sum_{i=1}^{n} J(X_{i:n}) - \frac{1}{2}\sum_{i=1}^{n}\left[\alpha\int_{-\infty}^{\infty} f_{i:n}^2(x)\,\mathrm{d}x + (1-\alpha)\int_{-\infty}^{\infty} f_{i:n}(x) f(x)\,\mathrm{d}x\right] \\
&= (1-\alpha)\left[nJ(X) - \sum_{i=1}^{n} J(X_{i:n})\right].
\end{aligned}$$

■

COROLLARY 2: *Since $\alpha \in [0,1]$, by comparison (9) with (14), we conclude that*

$$\sum_{i=1}^{n} D_n^{**}(f_{i:n}, f_{[i:n]}) \le \sum_{i=1}^{n} D_n(f_{i:n}, f).$$

The following theorem provides a relation between Eqs. (11) and (14).

THEOREM 7: *For any $0 \le \alpha \le (\sqrt{5}-1)/2$, we have*

$$\sum_{i=1}^{n} D_n^{*}(f_{[i:n]}, f) \le \sum_{i=1}^{n} D_n^{**}(f_{i:n}, f_{[i:n]}).$$

https://doi.org/10.1017/S0269964820000716 Published online by Cambridge University Press

PROOF: It is sufficient to show that

$$\sum_{i=1}^{n} D_n^{**}(f_{i:n}, f_{[i:n]}) - \sum_{i=1}^{n} D_n^{*}(f_{[i:n]}, f) \geq 0.$$

To do this, from Eqs. (11) and (14), we have

$$\sum_{i=1}^{n} D_n^{**}(f_{i:n}, f_{[i:n]}) - \sum_{i=1}^{n} D_n^{*}(f_{[i:n]}, f) = \sum_{i=1}^{n} J(X_{[i:n]}) - (1-\alpha)\sum_{i=1}^{n} J(X_{i:n}) - n\alpha J(X).$$

But it is can be shown that

$$\sum_{i=1}^{n} J(X_{[i:n]}) = \alpha^2 \sum_{i=1}^{n} J(X_{i:n}) + n(1-\alpha^2)J(X).$$

Therefore,

$$\sum_{i=1}^{n} D_n^{**}(f_{i:n}, f_{[i:n]}) - \sum_{i=1}^{n} D_n^{*}(f_{[i:n]}, f) = (\alpha^2 + \alpha - 1)\left[\sum_{i=1}^{n} J(X_{i:n}) - nJ(X)\right]. \quad \textbf{(15)}$$

On the other hand, from (10), we have $\sum_{i=1}^{n} J(X_{i:n}) - nJ(X) \leq 0$; hence, the right-hand side of (15) is greater than or equal to zero when $\alpha^2 + \alpha - 1 \leq 0$, and this occurs for $0 \leq \alpha \leq (\sqrt{5}-1)/2$, which completes the proof. □ ■

COROLLARY 3: *From Corollary 2 and Theorem 7, for any* $0 \leq \alpha \leq (\sqrt{5}-1)/2$, *we conclude that*

$$\sum_{i=1}^{n} D_n^{*}(f_{[i:n]}, f) \leq \sum_{i=1}^{n} D_n^{**}(f_{i:n}, f_{[i:n]}) \leq \sum_{i=1}^{n} D_n(f_{i:n}, f).$$

Furthermore, Raqab and Qiu [14] proposed another criterion to measure the discrimination information as follows.

DEFINITION 3: *Let* X *and* Y *be two random variables with pdfs* f *and* g, *respectively. The discrimination information between* X *and* Y *denoted by* $R(X,Y)$ *is defined as follows:*

$$R(X,Y) \triangleq R(f,g) = \frac{1}{2}\int_{-\infty}^{\infty} \frac{f(x)}{g(x)} f(x)\, dx.$$

By Definition 3, the discrimination information between the ith order statistic of sample size n from RSS data and the underlying distribution f can be written as follows:

$$R_n(f_{i:n}, f) = \frac{1}{2}\int_{-\infty}^{\infty} \frac{f_{i:n}(x)}{f(x)} f_{i:n}(x)\, \mathrm{d}x. \quad \textbf{(16)}$$

Similar arguments can be applied to the discrimination information between the ith judgment order statistic from RSS data and the underlying distribution. To do this, we can define

$$R_n^{*}(f_{[i:n]}, f) = \frac{1}{2}\int_{-\infty}^{\infty} \frac{f_{[i:n]}(x)}{f(x)} f_{[i:n]}(x)\, \mathrm{d}x.$$

https://doi.org/10.1017/S0269964820000716 Published online by Cambridge University Press

REMARK 2: *Suppose that* $\boldsymbol{X}_{SRS}^{(n)}$ *and* $\boldsymbol{X}_{RSS}^{(n)}$ *are the SRS and perfect RSS data sets of the same size* n*, respectively. Raqab and Qiu [14] showed that*

$$\sum_{i=1}^{n} R_n(f_{i:n}, f) = \frac{n4^{n-1}}{2\binom{2n-1}{n-1}}. \tag{17}$$

Now, using (16), we obtain the discrimination information between the imperfect RSS and SRS data sets.

THEOREM 8: *Let* $\boldsymbol{X}_{SRS}^{(n)}$ *and* $\boldsymbol{X}_{RSS}^{*(n)}$ *be the SRS and imperfect RSS data sets of the same size* n*, respectively. Then,*

$$\sum_{i=1}^{n} R_n^*(f_{[i:n]}, f) = \frac{n}{2}\left[\frac{\alpha^2 4^{n-1}}{\binom{2n-1}{n-1}} - \alpha^2 + 1\right], \quad 0 \le \alpha \le 1. \tag{18}$$

PROOF: Using (8), we get

$$\begin{aligned}
&\sum_{i=1}^{n} R_n^*(f_{[i:n]}, f) \\
&= \frac{1}{2}\sum_{i=1}^{n}\int_{-\infty}^{\infty} \frac{[\alpha f_{i:n}(x) + (1-\alpha)f(x)]}{f(x)}[\alpha f_{i:n}(x) + (1-\alpha)f(x)]\,\mathrm{d}x \\
&= \frac{1}{2}\sum_{i=1}^{n}\int_{-\infty}^{\infty} [\alpha f_{i:n}(x) + (1-\alpha)f(x)]\left[\alpha i\binom{n}{i}F(x)^{i-1}\bar{F}(x)^{n-i} + (1-\alpha)\right]\mathrm{d}x \\
&= \alpha^2\sum_{i=1}^{n} R_n(f_{i:n}, f) + \frac{1}{2}\sum_{i=1}^{n}\int_0^1 \left[2\alpha(1-\alpha)i\binom{n}{i}u^{i-1}(1-u)^{n-i} + (1-\alpha)^2\right]\mathrm{d}u \\
&= \alpha^2\frac{n4^{n-1}}{2\binom{2n-1}{n-1}} + n\alpha(1-\alpha) + \frac{n}{2}(1-\alpha)^2 \\
&= \frac{n}{2}\left[\frac{\alpha^2 4^{n-1}}{\binom{2n-1}{n-1}} - \alpha^2 + 1\right].
\end{aligned}$$

■

COROLLARY 4: *For* $\alpha \in [0,1]$*, and by comparison (17) and (18), we conclude that*

$$\sum_{i=1}^{n} R_n^*(f_{[i:n]}, f) \le \sum_{i=1}^{n} R_n(f_{i:n}, f).$$

PROOF: It is easily seen that

$$\sum_{i=1}^{n} R_n(f_{i:n}, f) - \sum_{i=1}^{n} R_n^*(f_{[i:n]}, f) = \frac{n}{2}(1-\alpha^2)\left[\frac{4^{n-1}}{\binom{2n-1}{n-1}} - 1\right].$$

Since $\alpha \in [0,1]$, it is enough to show that $(4^{n-1}/\binom{2n-1}{n-1}) - 1 \ge 0$. For this purpose, we computed numerically this term for different values of n in Table 1. Observe that this term is always greater than zero. ■

https://doi.org/10.1017/S0269964820000716 Published online by Cambridge University Press

TABLE 1. The values of $(4^{n-1}/\binom{2n-1}{n-1}) - 1$ for different values of n

n	2	5	10	20	50	75	100	500
$(4^{n-1}/\binom{2n-1}{n-1}) - 1$	0.33	1.03	1.84	2.99	5.28	6.69	7.87	18.82

Now, we extend the results obtained to SRSS. To this end, let $\boldsymbol{X}^{(n)}_{\text{SSRS}}$, $\boldsymbol{X}^{(n)}_{\text{SRSS}}$ and $\boldsymbol{X}^{*(n)}_{\text{SRSS}}$ be the SSRS, perfect SRSS and imperfect SRSS data sets of the same size n, respectively.

COROLLARY 5: *For $\boldsymbol{X}^{(n)}_{SSRS}$, $\boldsymbol{X}^{(n)}_{SRSS}$ and $\boldsymbol{X}^{*(n)}_{SRSS}$ data sets and according to Corollaries 3 and 4, the following outcomes can be deduced:*

(i) For any $0 \le \alpha \le (\sqrt{5}-1)/2$, we have

$$\sum_{h=1}^{L}\sum_{i=1}^{n_h} W_h D^{*}_{n_h}(f_{h[i:n_h]}, f_h) \le \sum_{h=1}^{L}\sum_{i=1}^{n_h} W_h D^{**}_{n_h}(f_{hi:n_h}, f_{h[i:n_h]}) \le \sum_{h=1}^{L}\sum_{i=1}^{n_h} W_h D_{n_h}(f_{hi:n_h}, f_h).$$

(ii) For $\alpha \in [0,1]$, we get

$$\sum_{h=1}^{L}\sum_{i=1}^{n_h} W_h R^{*}_{n_h}(f_{h[i:n_h]}, f_h) \le \sum_{h=1}^{L}\sum_{i=1}^{n_h} W_h R_{n_h}(f_{hi:n_h}, f_h).$$

According to Corollary 5, we conclude that the discrimination information between perfect SRSS data ($\boldsymbol{X}^{(n)}_{\text{SRSS}}$) and SSRS data ($\boldsymbol{X}^{(n)}_{\text{SSRS}}$) is more than its counterpart between imperfect SRSS data ($\boldsymbol{X}^{*(n)}_{\text{SRSS}}$) (under the fraction of random rankings model) and SSRS data ($\boldsymbol{X}^{(n)}_{\text{SSRS}}$).

5. CONCLUSIONS

The information properties of the SRSS data in terms of extropy were explored in the present paper. Firstly, we provided the expression for the extropy of SRSS data. Then, the bounds of the extropy of SRSS data under the different conditions were obtained. Finally, we investigated the effect of imperfect ranking on discrimination information. The discrimination information between the perfect SRSS and SRS data sets was shown to be better than that of between the imperfect SRSS and SRS data sets.

Acknowledgments

The authors express their sincere thanks to the anonymous referees and the associate editor for their useful comments and constructive criticisms on the original version of this manuscript, which led to this considerably improved version.

The second author was supported by the National Nature Science Foundation of China (71871208, 71971204) and the Anhui Provincial Natural Science Foundation (1908085MG236, gxbjZD54).

https://doi.org/10.1017/S0269964820000716 Published online by Cambridge University Press

References

1. Ahmed, A., Alzaid, A., Bartoszewicz, J., & Kochar, S. (1986). Dispersive and superadditive ordering. *Advances in Applied Probability* 18(4): 1019–1022.
2. Bohn, L.L. & Wolfe, D.A. (1994). The effect of imperfect judgment rankings on properties of procedures based on the ranked-set samples analog of the Mann-Whitney-Wilcoxon statistic. *Journal of the American Statistical Association* 89(425): 168–176.
3. Chen, Z., Bai, Z., & Sinha, B (2003). *Ranked set sampling: theory and applications*. Springer Science & Business Media.
4. Eskandarzadeh, M., Crescenzo, A.D., & Tahmasebi, S. (2018). Measures of information for maximum ranked set sampling with unequal samples. *Communications in Statistics – Theory and Methods* 47(19): 4692–4709.
5. Frey, J. (2014). A note on Fisher information and imperfect ranked-set sampling. *Communications in Statistics – Theory and Methods* 43(13): 2726–2733.
6. Jahanshahi, S., Zarei, H., & Khammar, A. (2019). On cumulative residual extropy. *Probability in the Engineering and Informational Sciences*. doi:10.1017/S0269964819000196.
7. Jozani, M.J. & Ahmadi, J. (2014). On uncertainty and information properties of ranked set samples. *Information Sciences* 264: 291–301.
8. Lad, F., Sanfilippo, G., Agro, G. (2015). Extropy: complementary dual of entropy. *Statistical Science* 30(1): 40–58.
9. McIntyre, G. (1952). A method for unbiased selective sampling, using ranked sets. *Crop and Pasture Science* 3(4): 385–390.
10. Park, S. & Lim, J. (2012). On the effect of imperfect ranking on the amount of Fisher information in ranked set samples. *Communications in Statistics – Theory and Methods* 41(19): 3608–3620.
11. Qiu, G. (2017). The extropy of order statistics and record values. *Statistics & Probability Letters* 120: 52–60.
12. Qiu, G. & Eftekharian, A. (2020). Extropy information of maximum and minimum ranked set sampling with unequal samples. *Communications in Statistics – Theory and Methods*. doi: 10.1080/03610926.2019.1678640.
13. Qiu, G., Wang, L., & Wang, X. (2019). On extropy properties of mixed systems. *Probability in the Engineering and Informational Sciences* 33(3): 471–486.
14. Raqab, M.Z. & Qiu, G. (2019). On extropy properties of ranked set sampling. *Statistics* 53(1): 210–226.
15. Samawi, H.M. (1996). Stratified ranked set sample. *Pakistan Journal of Statistics – All Series* 12: 9–16.
16. Shaked, M. & Shanthikumar, J.G (2007). *Stochastic orders*. Springer Science & Business Media.
17. Tahmasebi, S., Longobardi, M., Kazemi, M., & Alizadeh, M. (2020). Cumulative Tsallis entropy for maximum ranked set sampling with unequal samples. *Physica A: Statistical Mechanics and its Applications*. doi: 10.1016/j.physa.2020.124763.
18. Vock, M. & Balakrishnan, N. (2011). A Jonckheere–Terpstra-type test for perfect ranking in balanced ranked set sampling. *Journal of Statistical Planning and Inference* 141(2): 624–630.

https://doi.org/10.1017/S0269964820000716 Published online by Cambridge University Press